BAYESIAN BENEFITS FOR AUDITING:

A Proposal to Innovate Audit Methodology

KOEN DERKS

Bayesian Benefits for Auditing: A Proposal to Innovate Audit Methodology

Koen Derks

ISBN 978-90-8980-982-7
Published by Nyenrode Business Universiteit
This publication is typeset in LaTeX using the Memoir class
Printed by Ipskamp Printing B.V., Enschede
Cover design and illustration by Viktor Beekman, @viktordepictor

NYENRODE BUSINESS UNIVERSITEIT

Bayesian Benefits for Auditing: A Proposal to Innovate Audit Methodology

Thesis

with regard to the doctorate/PhD degree

at Nyenrode Business Universiteit

on authority of the

Rector Magnificus, prof. dr. K. M. Becking

in accordance with the Doctorate Committee.

The public defense takes place on

Friday March 17, 2023

at exactly four o“clock

by

Koen Peter Derks

born on April 12, 1995

in Blaricum, The Netherlands.

Examination Committee

Supervisor:	Prof. dr. R. M. Wetzels	Nyenrode Business University
Co-supervisor:	Prof. dr. J. J. B. de Swart MBA	Nyenrode Business University
Co-supervisor:	Prof. dr. E. M. Wagenmakers	University of Amsterdam
Other members:	Prof. dr. S. K. Asare	University of Florida
	Prof. dr. J. P. van Buuren RA	Nyenrode Business University
	Prof. dr. M. D. Lee	University of California Irvine
	Prof. dr. G. C. M. Majoor RA	Nyenrode Business University

Voor mijn vader Laurens
mijn moeder Maaike
mijn zusje Laura
mijn liefste Sophie

Contents

Chapter 1

Introduction

Abstract

This project aimed to promote, develop, and make available techniques for Bayesian statistical inference in the auditing profession. Auditors, audit firms, and audit standard setters have long shied away from developing innovative statistical methods. To a large extent, they have ignored the Bayesian revolution in statistics, which means that many auditors analyze their data using the familiar frequentist statistical framework that has been used for decades. This is a troubling state of affairs: Auditors perform important work in today's society only to analyze their data using statistical methods that are suboptimal for answering the questions at hand. For this reason, this thesis proposes that audit practitioners can innovate greatly by adding Bayesian methods to their statistical toolbox. However, before Bayesian statistics can be fully adopted by auditors in practice, three conditions must be fulfilled. First, there must be a clear argument for using Bayesian statistics in a modern auditing context; second, auditors must have an easy-to-use toolbox to perform Bayesian inference; and third, Bayesian techniques must be easily accessible to all audit practitioners via user-friendly open-source software.

1.1 A brief historical preamble

This thesis advocates the use of Bayesian statistics in auditing theory and practice. It aims to address a number of academic and practical needs, particularly in the area of statistical sampling. To understand how these needs came to be, it is important to first have a basic understanding of how statistical sampling has evolved in this area. Let's begin at the beginning.

The word "audit" comes from the Latin verb audire, which means "to hear". Perhaps it would not surprise you to learn that the earliest known auditors were spies for the ancient Persian king Darius, who served as the "King's ears" by monitoring the conduct of his subordinates. The first written checking activities, which are also thought to be an early form of auditing, date back a few hundred years later to ancient China, Egypt, and Greece. Throughout recorded history, the function of an audit has been to provide groups or individuals with information or reassurance about the conduct or performance of others (Flint, 1988). Because of their function, audits have historically been crucial to maintaining social stability; after all:

> "Without audit, no control; and if there is no control, where is the seat of power?" (Mackenzie, as cited in Normanton, 1966, p. 7)

Nowadays, auditors serve the public interest rather than the interests of kings and pharaohs, but the function of an audit remains the same. The auditing profession as we know it today can be traced back to 1844, when the British Parliament passed the Joint Stock Companies Act (Lee and Ali, 2008). The law required directors of companies to report to shareholders using an audited financial statement. Furthermore, the act made auditors responsible for ensuring that these financial statements were presented fairly.

By the late nineteenth century, most companies wishing to trade on the stock exchange were required to provide annual audited financial statements and balance sheets (Leung et al., 2007). During this time, the demand for auditing increased, and so did the size and complexity of companies, resulting in a large number of transactions on the balance sheets. As a result, it became impractical for auditors to confirm every transaction. This popularized the use of sampling (i.e., selecting and auditing a subset of all transactions on the balance sheet) in auditing research and practice. Audit researchers were quick to advocate that statistical sampling could help practitioners be more efficient while maintaining audit quality:

> "The purposes of the external audit will be better accomplished with the use of [statistical] sampling than would be the case if the same dollar-cost was expended upon more detailed vouching." (Kirkham and Gaa, 1939, p. 144)

At the turn of the twentieth century, the body of scientific literature on statistical audit sampling was full of examples of frequentist sampling methodologies being used (Kraft Jr, 1968). However, auditors who conducted statistical sampling in practice quickly discovered that large samples were required to achieve acceptable statistical results at high levels of assurance, and that this was very expensive. In response, auditors increased their reliance on internal control systems in their audit procedures: When auditors determined that a company's internal control systems were effective, they reduced the number of samples that followed.

In the mid-twentieth century, audit researchers began to think critically about how to take into account pre-existing information to reduce the amount of required audit effort. At that time, Bayesian statistics was already established in the field of mathematics and statistics (van de Schoot et al., 2017) thanks to the pioneering work of Jeffreys (1939), and it was being advocated for use in many business decision-making situations (Kraft Jr, 1968). This alternative type of statistical inference piqued the interest of auditing researchers because it enabled them to formally incorporate pre-existing information into the statistical analysis, such as information about the auditee's internal control systems. Researchers were quick to promote Bayesian inference for use in auditing:

> "[T]he value of the Bayesian approach to the accountant is not that it stresses the existence of prior information, but rather that it suggests a methodology for incorporating prior information into the analysis." (Birnberg, 1964, p. 111)

In the late twentieth century, when computing power had increased and Bayesian computations had become tractable for auditors in practice, Bayesian statistics was argued to be an appropriate and theoretically correct tool to be used in audit sampling. The academic literature from that time reflects agreement among audit researchers that widespread adoption of Bayesian statistics in practice was imminent:

> "In the past, the theoretically correct model has been virtually impossible to use in practice. Now with the increasing use of computerized audit decision aids, it is becoming increasingly more feasible. It is not too difficult to predict that within the foreseeable future such complex models will be an integral part of the audit." (Leslie, 1984, pp. 104–105)

> "Field workers now have the computing power necessary to do sophisticated audit planning and evidence integration. We scholars can no longer avoid these issues by using the excuse of computational impracticality." (Kinney, 1984, pp. 131–132)

However, at the dawn of the twenty-first century, nearly three decades later, the anticipated adoption of Bayesian statistics in auditing did not occur. Frequentist inference was still the foundation of most audit textbooks (e.g., Touw and Hoogduin, 2012), it remained what auditors were taught and were using in practice, and it was still being promoted by audit researchers (e.g., Edmonds et al., 2019) and (guidance on) auditing standards (e.g., Stewart, 2012; American Institute of Certified Public Accountants (AICPA), 2019). In his thesis—which contains some of the most recent fundamental texts on Bayesian inference in auditing—Stewart (2013) provides his account on why Bayesian statistics was never fully able to find its place:

> "In my opinion, one reason for this lack of progress is a general decline in interest in the application of quantitative, probabilistic and statistical methods in auditing. Accounting firms reduced basic audit research in this area as they focused on the development of new services. Academic research into and teaching of such methods declined as a consequence, leading to a further decline in practitioner awareness and proficiency. The methods were also perceived as difficult to apply and meld with professional judgment, and the cost-benefit of training auditors to be proficient in them was questioned. Interests, proficiency, and academic engagement declined." (Stewart, 2013, p. 23)

Unfortunately, his account still summarizes the current state of affairs. To this day, statistical sampling remains one of the primary methods auditors use to obtain reasonable assurance about the misstatement in a population (Christensen et al., 2015; Glover et al., 2015). Hence, the advantages that made Bayesian statistics initially appealing to audit researchers and practitioners are still relevant, but a lack of guidance, proficiency and easy-to-use statistical tools is holding off the adoption of Bayesian methods in practice. On top of that, due to advances in computational capabilities and the changes in the audit environment because of those advances (Salijeni et al., 2019), the body of fundamental literature on Bayesian auditing has grown outdated and is therefore difficult to relate to for auditors. As a result, the fundamental arguments for Bayesian inference in auditing have largely faded from collective memory.

In my opinion, the adoption of Bayesian statistics in auditing requires three conditions to be fulfilled. First, there must be a clear argument for using Bayesian statistics in a modern auditing context; second, auditors must have an easy-to-use toolbox to perform Bayesian inference in practice; and third, Bayesian techniques must be freely available to all auditors via user-friendly open-source software. This thesis outlines my efforts to meet these three requirements.

1.2 Frequentist versus Bayesian probability

The concept of probability will be covered regularly in this thesis. Therefore, I believe it is useful to provide a bird's-eye view of the two lenses through which probability can be approached: the frequentist approach and the Bayesian approach. The fundamental difference between these two approaches lies in how they interpret probability and uncertainty.

According to the frequentist school of thought, the probability of an event is the long-term relative frequency of that event (Wagenmakers et al., 2008). For example, the probability of obtaining heads when flipping a fair coin in the long run is $\theta = \frac{1}{2}$. The outcome of each flip (i.e., the data) is determined by chance: Every flip the coin lands on either heads or tails, each of which occurs with a probability of $\frac{1}{2}$. This means that if you flip the coin many times, you would anticipate it to land on heads half of the time. However, if you only flip the coin a few times, it is reasonable to expect that, simply by chance, not exactly half of the flips will result in a head. For instance, it is not difficult to imagine a scenario in which you would flip the coin three times and obtain three heads in a row. Because in a frequentist view the data is the outcome of a probabilistic process, it can make a probabilistic statement about the data given an assumption about θ. For example, what is the the probability of obtaining three heads in a row given that $\theta = 0.5$?

According to the Bayesian school of thought, the probability of an event is a degree of belief in the occurrence of that event rather than its long-term relative frequency (Wagenmakers et al., 2008). In contrast to how the frequentist approach is commonly applied, Bayesian statistics assumes that there is uncertainty about θ, however, the knowledge about θ can be expressed in the form of a distribution that assigns a relative probability to each of its possible values: a prior distribution. Such a probability distribution can be used to capture your current state of knowledge about θ, even before any data is observed. For example, your beliefs about the probability of the coin landing heads may be different from mine if, say, you have information that the coin is slightly bent and I do not. Then, it is reasonable to incorporate this pre-existing information into the prior distribution for θ. Using a mathematical rule called Bayes' theorem, the prior information can be combined with the information in the data to determine the post-data, or posterior, knowledge about θ.

$$\text{prior} + \text{data} \rightarrow \text{posterior}$$

Because a Bayesian approach assigns probabilities to values of θ, it can make a probabilistic statement about θ given the data (and the prior information). For example, what is the probability that θ is (higher or lower than) $\frac{1}{2}$ given three heads in a row?

1.2.1 Example

Let's consider an example in an audit context to briefly illustrate the practical differences between the frequentist approach and the Bayesian approach. Suppose an auditor is tasked with obtaining reasonable assurance about a population of accounts receivable not being materially misstated. A population not being materially misstated in an audit context means that the population does not contain misstatements (i.e., errors) larger than a set maximum: the performance materiality. For illustrative purposes, let's say the performance materiality for the population is $\theta_{max} = \frac{1}{2}$, but note that typically this value is set much lower. The auditor has decided to perform audit sampling, taking a random sample of ten items from this population. After inspecting the $n = 10$ items in the sample, they discover that $k = 0$ items contain a misstatement. Using these data, the auditor wants to form an opinion about whether the population is materially misstated or not.

In a frequentist approach, the auditor typically defines the probability of misstatement in the population θ as the performance materiality θ_{max}, but keep in mind that this is an assumption. The left panel in Figure 1.1 shows the probability of observing a particular number of misstatements k in the sample of ten items given the auditor's assumption about θ. The figure illustrates that, even if the probability of misstatement in the population is $\frac{1}{2}$, it is still possible to obtain a sample of ten items and discover no misstatements. However, there is a very low probability of obtaining this sample ($\frac{1}{2}^{10} = 0.001$). This may lead the auditor to conclude that the population is likely not materially misstated. Note that this commonly used frequentist conclusion is based on the probability of the data given the auditor's assumption about θ.

As a Bayesian, the auditor starts by specifying their current knowledge about the probability of misstatement in the population via the prior distribution. For illustrative purposes, suppose that the auditor does not know which values of the probability of misstatement are more likely than others. Then, a prior distribution that assumes all possible values of θ to be equally likely can reflect their current state of knowledge. This uniform prior distribution is shown as a dashed line in the

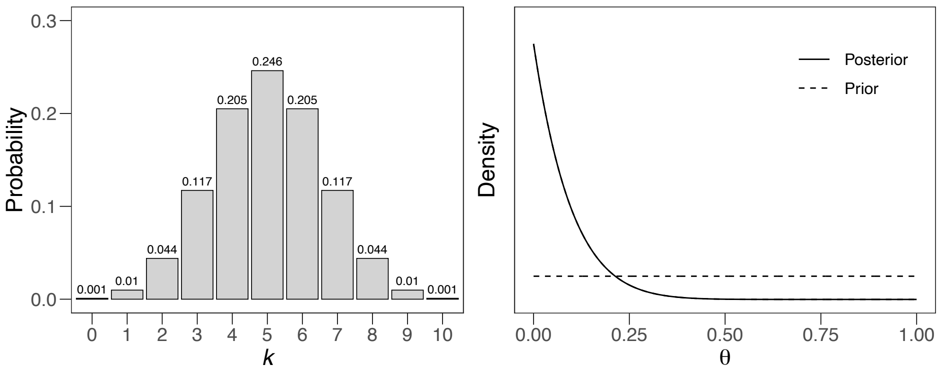


Figure 1.1: The left panel shows the probability of observing zero to ten misstatements k in a sample of $n = 10$ items if the true probability of misstatement θ is $\frac{1}{2}$. The right panel shows a uniform prior distribution and the posterior distribution for the probability of misstatement θ after observing a sample of $n = 10$ items containing $k = 0$ misstatements.

right panel of Figure 1.1. Next, the data from the sample can be used to update the prior distribution to a posterior distribution, which is displayed as a solid line in the right panel of Figure 1.1. The panel shows that the posterior distribution assigns a relatively high probability to low values of θ (e.g., the probability that θ is smaller than $\frac{1}{2}$ is 0.999). This may lead the auditor to conclude that the population is likely not materially misstated. Note how, in comparison to the frequentist approach described in the preceding paragraph, a Bayesian conclusion is based on the auditor's knowledge about θ given the data (and the prior information).

In statistical audit sampling, both the frequentist and the Bayesian approaches are reasonable and can be applied within the framework prescribed by international auditing standards (International Auditing and Assurance Standards Board (IAASB), 2018; American Institute of Certified Public Accountants (AICPA), 2021; Public Company Accounting Oversight Boards (PCAOB), 2020). However, the Bayesian approach to audit sampling comes with considerable advantages for auditors. In the following sections, I will go into greater detail about the advantages of Bayesian inference for auditing, as well as describe the academic and practical needs that this thesis addresses.

1.3 Academic relevance

The main argument for the Bayesian approach in auditing is that it is consistent with the philosophy of the audit. For instance, to comply with auditing standards, auditors must reduce the audit risk to an acceptable level (e.g., 5 percent) and obtain a reasonable level of assurance (e.g., 95 percent) that the total amount of misstatement does not exceed performance materiality (American Institute of Certified Public Accountants (AICPA), 2019). The way that frequentist statistics is typically applied in audit sampling does not easily allow for such conclusions.

As shown in the previous section, a commonly used conclusion of this approach concerns the probability of data given an assumption about the performance materiality (for example, there is a 5 percent probability of observing these data given that the population contains misstatements that exceed the performance materiality). This is typically not what auditors are hoping to conclude from their samples because it is not a probabilistic statement about the misstatement in the population. Bayesian statistics makes it easier for auditors to draw the conclusions they want. A Bayesian conclusion is a statement about the misstatement in the population given the data (for example, given the data there is less than 5 percent probability that the misstatement in the population exceeds the performance materiality), which is in line with what auditors are hoping to conclude. Furthermore, Bayesian statistics makes it easier for auditors to align their approach to audit sampling with the situation in practice. In auditing, pre-existing information is often available (such as the risk of material misstatement), which auditors typically want to take into account when conducting a statistical analysis. Bayesian statistics enables auditors to aggregate many sources of information throughout the audit in a statistically sound manner and base their opinion on the aggregated information. In sum, the objectives of the auditor are closely aligned with the characteristics of Bayesian inference because the audit process is essentially Bayesian in nature (Stewart, 2013, pp. 22–24). Hence, Bayesian statistical models are ideally suited to apply in this context.

Despite its natural alignment with the audit, there is currently little academic interest in describing the Bayesian statistical framework in the context of auditing. Moreover, the few innovative statistical methods that are being presented are challenging for auditors to apply in practice (e.g., Meeden, 2003; Martel-Escobar et al., 2018). Both of these issues largely contribute to the sparse academic literature developing Bayesian techniques in an auditing context, even though these techniques are continually being applied and described in other scientific fields (e.g., Brown and Prescott, 2015; Wagenmakers et al., 2018b). This, in my opinion, indicates the need for improving the current state of the Bayesian auditing literature. This thesis makes three relevant contributions to the academic auditing literature in response to this need.

First, the thesis restates and expands the fundamental arguments in favor of Bayesian inference in an auditing context. During the late-twentieth century, many scholars like Kinney (1975), Felix (1976), Leslie (1984), Stringer and Stewart (1986), Steele (1992), van Batenburg and Kriens (1989) and Johnstone (1994), have advocated the Bayesian approach to audit sampling. However, the audit environment has changed since then (Salijeni et al., 2019), and the methods that were previously described as being optimal are no longer tailored to the current auditing environment. For instance, the use of relatively simple prior distributions and statistical models is emphasized in previous literature, but this is no longer necessary due to an increase in computing power and the increased availability of external data sources. Thanks to availability of probabilistic programming languages like winBUGS (Lunn et al., 2000) and Stan (Carpenter et al., 2017)), applying complex Bayesian models in practice is currently relatively easy for auditors. In the last two decades, some articles (e.g., Meeden, 2003; Higgins and Nandram, 2009) proposing innovations in statistical auditing methodology have been published, but many of these articles fail to acknowledge the existence of external sources of information. Furthermore, during this time period, only a small number of articles have been published in this field that employ Bayesian inference (e.g., Laws and O'Hagan, 2000; Martel-Escobar et al., 2018; Laws and O'Hagan, 2002; Johnstone, 2018). These articles do, of course, apply Bayesian statistics, but they often do not go into detail about its fundamental theory or its benefits for auditors in a modern auditing environment. Unfortunately, this means that the current body of literature describing the theory and practice of Bayesian statistics in an auditing context is out of step with the kinds of audit environments that today's auditors are faced with. For this reason, this thesis reiterates and modernizes the arguments in favor of Bayesian inference made by scholars during the late twentieth century.

Second, the thesis introduces an innovative Bayesian approach to audit sampling where the auditor develops a statistical model for the data and uses the Bayes factor as a yardstick for audit evidence. The academic literature developing statistical auditing methodology is sparse. This can be partly explained by the fact that there is a lot of pressure during an audit, the stakes are high, and innovations should have the support of both standard-setters and auditors. Sadly, this means that statistical methodology does not advance as quickly in our field as it does in other scientific disciplines. As a result, the statistical techniques that are described in the current auditing literature are frequently suboptimal to address the questions that auditors have. For example, there is an increasing amount of external data that is available in the audit, and much research has been done on how to obtain this data (e.g., Appelbaum et al., 2018), but little research has been done on how to incorporate this data into later stages of the audit (e.g., audit sampling). Although other scientific disciplines have widely adopted statistical techniques that achieve this (Gelman et al., 2013), statistical models used in auditing are often simple and inefficient. The thesis seeks to fill this methodological gap in the academic auditing literature. To achieve this, it introduces Bayesian (generalized linear) modeling as a framework to incorporate multiple sources of information into the statistical analysis. Additionally, it introduces a measure for statistical audit evidence that has not been previously discussed in the audit sampling literature: the Bayes factor (Kass and Raftery, 1995; Johnstone, 2018, pp. 33–34). Simply put, the Bayes factor indicates how much more likely the sample data is to occur in the presence of material misstatement than in the absence of material misstatement, or vice versa. This makes it an intuitive measure of audit evidence that closely matches the auditor's reasoning. By providing a full Bayesian framework for estimation and hypothesis testing, the thesis aims to lay the foundation for the development of more advanced statistical methods in this field.

Third, the thesis captures the output of academic research into open-source software such that it can be used by all researchers and practitioners. Statistical methodology described by auditing researchers is not often made available to students, researchers and practitioners, despite this practice being common in some other scientific fields (e.g., Foster and Deardorff, 2017; Love et al., 2019). Hence, most available software tools implementing statistical auditing methodology are specialized commercial programs like IDEA (CaseWare Analytics, 2022) and ACL (Dilligent, 2022), which are both costly and nontransparent because they do not provide access to their source code. Open-source software is software whose source code is publicly available without restriction and without cost. Most importantly, open–source software is developed collaboratively in a community with software developers and users. Hence, the development of open-source software is sometimes compared to a crowded Bazaar and that of closed-source software to a Cathedral (Raymond, 1999). This has a number of advantages, for example that the end user always comes first and development of the software is normally faster than with closed-source (AlMarzouq et al., 2005; Morgan and Finnegan, 2007). Especially in the field of auditing, where trust and transparency are essential, open-source software has the potential to contribute. In addition to being more transparent and, most importantly, free, open-source software can serve as a link between auditors in practice and innovative statistical methodology that results from academic research. This thesis implements the basic (Bayesian) methods for statistical audit sampling into the open-source software JASP (JASP Team, 2022), thereby providing a tool for all academic practitioners to use and expand upon. With that, it addresses the absence of openly accessible statistical software and provides a platform for collaboration between audit researchers and practitioners.

In sum, this thesis aims to develop an innovative approach to statistical auditing that elevates the field to the Bayesian methodological standards currently upheld in other scientific disciplines (van de Schoot et al., 2021).

1.4 Practical relevance

Although the Bayesian approach to auditing has many advantages, these advantages are often unknown to auditors or difficult to apply in practice. This, in my opinion, indicates the need to make Bayesian methods more accessible to auditors. This thesis makes three relevant contributions to the auditing practice in response to this need.

First, the thesis aims to promote Bayesian inference in the auditing profession by offering a thorough overview and discussion of the practical advantages of Bayesian statistics in a modern auditing context. In short, four major benefits of Bayesian statistics are put into focus. First, Bayesian statistics offers auditors a straightforward interpretation of statistical results. While frequentist results are often prone to misinterpretation (Hoekstra et al., 2014; Nickerson, 2000, pp. 246–263), a Bayesian conclusion is in line with what auditors want to learn from their samples. Second, Bayesian statistics enables statistical conclusions to be easily extended to any level of complexity. In a Bayesian analysis, both the prior distribution and the statistical model can take into account pre-existing information. This means that auditors can make optimal use of their pre-existing information and, as a result, provide a fine-grained audit opinion because it is specifically tailored to the audit and the auditee. Third, since auditors must explicitly state their assumptions in a Bayesian approach via the prior distribution, this can increase transparency towards stakeholders of the audit. Information incorporated into the analysis should be properly justified, providing stakeholders with complete transparency into how the statistical results were obtained and where improvements in the organization lie (de Swart et al., 2013). Last but not least, Bayesian inference can help the auditor work more efficiently. The use of pre-existing information means that more information is available at the start of the analysis, resulting in a smaller required sample size or an increase in statistical precision. To aid auditors with the adoption of these methods in practice, many of the chapters in this thesis discuss and offer practical recommendations on how to weigh the benefits and drawbacks of the Bayesian approach.

Second, the thesis aims to develop easy-to-use and easy-to-justify Bayesian statistical methods for parameter estimation and hypothesis testing that are specifically tailored to audit sampling. Hence, it equips auditors with the tools to employ Bayesian inference in real-world situations. For example, it explains how to justify the use of pre-existing information in the prior distribution and the statistical model and proposes an intuitive measure of audit evidence: the Bayes factor. For didactic purposes, each chapter includes practical examples, data sets, and R code, making it easy for any auditor to follow along, or to put these methods into practice.

Third, the thesis aims to make Bayesian techniques freely available to all audit practitioners via user-friendly open-source software. It describes the development of JASP for Audit, an implementation of Bayesian (and frequentist) audit sampling techniques in the open-source statistical software program JASP (url: `https://jasp-stats.org`). By programming the fundamental Bayesian techniques for audit sampling into openly accessible software, it aims to ensure that any auditor is able to use Bayesian inference at all times.

Therefore, the thesis has practical value for a large audience, not purely auditors who conduct audit sampling.

1.5 Chapter outline

1.5.1 Part I. Bayesian Parameter Estimation

The first part of the thesis centers around Bayesian parameter estimation in audit sampling, with a particular emphasis on the use of pre-existing information.

Chapter 2 elaborates on the basic principles underlying Bayesian parameter estimation, and discusses how information can be incorporated into the prior distribution. The key idea that this chapter introduces is that the prior distribution enables auditors to statistically build upon pre-existing information about the (probability of) misstatement in their sampling procedures. In this chapter, I discuss how incorporating pre-existing information into the prior distribution comes with several practical advantages, such as a potential reduction in sample size and an increase in transparency. Furthermore, the chapter outlines five methods to construct a prior distribution in an audit sampling context. These prior distributions are constructed using pre-existing audit information, thus, they allow for a sensible justification by the auditor. The chapter concludes with a discussion on the considerations that apply to constructing a prior distribution.

Chapter 3 builds on the basic principles of Bayesian parameter estimation from Chapter 2 and discusses how pre-existing information can be incorporated into the statistical model. The key idea that this chapter introduces is that the statistical model enables auditors to statistically build upon multiple sources of data about the population in their sampling procedures. In this chapter, I discuss how incorporating additional data into the statistical model comes with two advantages for the auditor. First, it helps them form a more fine-grained opinion about the population because they can clearly explain how the incorporated data affects the misstatement; second, it makes it easier for them to identify misstatements because they can more precisely distinguish between items in the population. The chapter continues to demonstrate that a Bayesian generalized linear modeling approach can help the auditor to construct a statistical model for audit sampling that aligns with the situation in practice, and that this comes with a potential increase in efficiency. The chapter concludes with practical recommendations for the use of this approach in an audit context.

1.5.2 Part II. Bayesian Hypothesis Testing

The second part of the thesis centers around Bayesian hypothesis testing, with a focus on the quantification of statistical audit evidence.

Chapter 4 discusses the concept of statistical audit evidence, and introduces Bayesian hypothesis testing as a means to quantify statistical evidence from audit samples. The key idea this chapter introduces is that statistical evidence can be quantified using the Bayes factor; a measure of relative evidence comparing two hypotheses being tested. The chapter shows that, when it comes to effectiveness and efficiency, Bayesian hypothesis testing using the Bayes factor addresses several practical disadvantages that come with frequentist hypothesis testing. It goes on to demonstrate various scenarios from a modern auditing context in which the Bayes factor can be used. The chapter concludes with practical implications of the use of the Bayes factor in practice.

Chapter 5 outlines the development of a default Bayesian hypothesis test for audit sampling. The Bayes factor is sensitive to the specification of the prior distribution, however, the sensitivity of the Bayes factor to the choice of prior has not been previously studied in the context of audit sampling. Unfortunately, prior distributions that are tempting to use in an audit sampling context because they are easy to justify (such as a uniform prior) can yield Bayes factors that quantify evidence in the other direction than the data point to. This chapter introduces an impartial prior distribution as a means to solve this problem. The statistical motivation for the impartial prior distribution is that the resulting Bayes factor is consistent, that is, the Bayes factor from an impartial prior distribution will always quantify evidence for the hypothesis best supported by the data. The key idea this chapter introduces is that an impartial Bayesian hypothesis test is appropriate for many situations, while also being simple to use and easy to justify. The chapter concludes with a comparison of (impartial) Bayes factors and frequentist p-values, and discusses what this implies for auditors in practice.

1.5.3 Part III. Software Implementation

The third and final part of this thesis centers around a software implementation of the Bayesian statistical framework for audit sampling proposed in the foregoing chapters.

Chapter 6 introduces JASP for Audit, an open-source statistical software program with a graphical user-interface that implements both Bayesian and frequentist statistical techniques for audit sampling. As part of this project, JASP for Audit was developed as an add-on module for the existing statistical software JASP (JASP Team, 2022), with the goal to help the auditor in the statistical aspects of an audit. It does this by offering (among other analyses) a guided workflow that follows the familiar four-step audit sampling process, makes the correct statistical decisions under the hood, and automatically creates an audit report containing the statistical results and the interpretation of these results. This chapter discusses the advantages of JASP for Audit for the auditing practice and offers a detailed walkthrough of three real-world examples. The chapter concludes with recommendations for the use of JASP for Audit in practice. Furthermore, Appendix 6.B discusses the 'jfa' package (Derks, 2022), an R based implementation of the functionality provided by JASP for Audit.

The thesis is concluded with a discussion on future research directions.

Part I

Bayesian Parameter Estimation

Chapter 2

Incorporating Audit Information into the Prior Distribution

Abstract

Auditors often have prior information about the auditee before starting the substantive testing phase. We show that applying Bayesian statistics in substantive testing allows for integration of this information into the statistical analysis through the prior distribution. For example, an auditor might have performed an audit last year, they might have information on certain controls in place, or they might have performed analytical procedures in an earlier stage of the audit. Incorporating this information directly in the statistical procedure enables auditors to tailor their sampling plan to the auditee, thereby increasing audit transparency and efficiency. However, defining a suitable prior distribution can be difficult because what constitutes a suitable prior depends on the specifics of the audit and the auditee. To help the auditor in constructing a prior distribution we introduce five methodologies, consider their pros and cons, and give examples of how to apply them in practice.

Keywords: Audit, Bayesian statistics, financial statements, prior distribution.

2.1 Introduction

A financial audit is an inspection of an organization's financial statements before they are released to the public. In the audit report, the auditor presents their opinion on the fairness of these statements to inform stakeholders of the organization about its current financial situation. An organization's financial statements

This chapter is published as Derks, K., de Swart, J., van Batenburg, P., Wagenmakers, E.–J., & Wetzels, R. (2021). Priors in a Bayesian audit: How integration of existing information into the prior distribution can improve audit transparency and efficiency. *International Journal of Auditing*, 25(3), 621–636. doi: `https://doi.org/10.1111/ijau.12240`

are derived from several large populations of transactions (Neter and Loebbecke, 1975, 1977; Ramage et al., 1979; Titera, 2013) that, until a few decades ago, were all assessed in detail (Power, 1992). Nowadays, technological advancements (e.g., big data and artificial intelligence) theoretically allow auditors to inspect these large (digital) populations completely. However, in most businesses, the data and business process quality would make a complete inspection result in a high number of seemingly irrelevant audit findings (Brown-Liburd et al., 2015). Moreover, Earley (2015) and Gepp et al. (2018) argue that these techniques are only slowly progressing in the audit practice. Recently, Yoon and Pearce (2021) argue that such analytical procedures and statistical sampling both have unique benefits, and that they are best used to complement each other. Statistical sampling therefore remains a much-used technique due to its efficiency in larger populations (Christensen et al., 2015; Hitzig, 1995; American Institute of Certified Public Accountants (AICPA), 2019; Maingot and Quon, 2009; Srivenkataramana, 2018; Van Der Nest et al., 2015). After all, statistical sampling enables the auditor to infer a conclusion about a certain characteristic of the population based on only a small subset of this population.

The challenge that auditors routinely face is to tailor their statistical procedures to the specific situation of the audit and the auditee (Brivot et al., 2018; Coram et al., 2003; Lombardi et al., 2014). Due to variability in organizations' administrative systems, controls, or possible malicious intent, no two organizations share the same quality of financial reporting and, thus, can be audited in the exact same manner (Stewart, 2013; Kachelmeier et al., 2014). In this chapter, we show that Bayesian statistics allows for this tailoring, arguably resulting in a more transparent and more efficient audit.

The International Standards on Auditing (ISA; International Auditing and Assurance Standards Board (IAASB), 2018) prescribe that the auditor can use statistical sampling to quantify the risk of an incorrect judgment resulting from their substantive testing procedures. To conduct statistical sampling, ISA 530 mandates "the use of probability theory to evaluate sample results" (International Auditing and Assurance Standards Board (IAASB), 2018, paragraph 5g). While in probability theory there are two main schools of thought, Bayesianism and frequentism (Dienes, 2011; Wagenmakers et al., 2008), auditing firms have mostly relied on frequentist methods to design their statistical sampling procedures and substantiate their conclusions (Christensen et al., 2015). In addition, frequentist estimation and testing have a crucial role in the accounting curriculum of most universities (e.g., Lee et al., 2018; Touw and Hoogduin, 2012). Although the standards' requirements for statistical sampling cannot be found exclusively in frequentist methods, it is clear that these methods currently dominate audit theory and practice. However, Bayesian methods that use probability theory to evaluate sample results may legitimately be used to make an inference in substantive testing.

The auditing standards prescribe that auditors are allowed to reduce the quantity of evidence that is required from substantive testing if they have existing information on an auditee that indicates a low risk profile (ISA 530, International Auditing and Assurance Standards Board (IAASB), 2018). For the statistical analysis, this usually translates into a reduced sample size for auditees with a lower risk profile. However, within the frequentist framework, it is not possible to incorporate this existing information into the statistical methodology, except on an ad-hoc basis (Wagenmakers et al., 2008). Frequentist methods allow the auditor to modify their audit procedures on the basis of existing information, but they do not allow the auditor to coherently incorporate this information into the analysis. Because it is unclear in a frequentist analysis what the mathematical relationship is between the existing information and the sample information it is said to be incoherent (Lindley, 2004). As a result of this incoherency, frequentist substantive testing procedures often expose the auditor to an audit risk that is greater than is desired (Jiambalvo and Waller, 1984; Kinney, 1983; Stewart, 2013).

In contrast to frequentist statistics, Bayesian statistics has been continuously advocated as a means to integrate existing information into substantive testing in a coherent manner (Beck et al., 1985; Corless, 1972; Crosby, 1980; Felix, 1976; Godfrey and Neter, 1984; Laws and O'Hagan, 2002; Meeden, 2003; Sahu and Smith, 2006; Tsui et al., 1985). More concretely, in the Bayesian approach, the auditor quantifies their existing information in a prior probability distribution such that it captures the information available to them. Since the prior distribution contains information that is specific to the auditee and the population, it allows for an informed and tailor-made point of departure for substantive testing. Bayesian statistics uses the rules of probability theory to revise the information in the prior distribution in light of the observed sample. This approach to revising information highlights the cumulative character of a Bayesian analysis, one on which many scholars and practitioners have agreed is appropriate in an audit context (Kinney, 1983; Stewart, 2013) because an audit itself is a continuous process (Leslie, 1984). Another argument why Bayesian methods are appropriate for auditors is that the audit standards describe an audit along the lines of that same philosophy. For example, ISA 330 states that "an audit of financial statements is a cumulative and iterative process" (International Auditing and Assurance Standards Board (IAASB), 2018, paragraph A60). Thus, one could argue that the Bayesian approach fits well with the audit standards and the goal of the auditor.

Despite these advantages, the use of Bayesian methods in the audit is scarce. A potential reason for this scarcity is that auditors need to translate existing audit information into a prior probability distribution, which is not necessarily part of the expertise of an auditor (Corless, 1972; Felix, 1976; Martel-Escobar et al., 2018; Stewart, 2013). First, determining the type of audit information to incorporate into the prior distribution can be difficult because what information might be incorporated depends on the situation at hand. Second, the translation from relevant audit information into a prior probability distribution is perceived as difficult (Abdolmohammadi, 1985, 1987; Corless, 1972). Nevertheless, by overcoming these hurdles the auditor can build upon existing information in a coherent manner, resulting in concrete advantages—such as a more accurate estimation of the population misstatement (Knoblett, 1970) and audit risk (Stewart, 2013), a potential reduction in sample size, and formalized predictions—that can increase efficiency and transparency in audit sampling. Furthermore, it has been shown in earlier studies based on the well-known audit populations from Neter and Loebbecke (1975) that Bayesian methods result in upper bounds that achieve nominal coverage (Chan and Smieliauskas, 1990; Swinamer et al., 2004). To increase the feasibility of Bayesian methods in the audit practice and to make these tangible advantages of Bayesian statistics more easily available for auditors, we will introduce five methods for incorporating existing audit information into a prior probability distribution. Please note that we do not see these Bayesian methods as a replacement for frequentist methods, but rather as an efficient addition to the auditor's statistical toolbox if needed.

The structure of this chapter is as follows. In Section 2.2 the three building blocks of Bayesian inference are introduced: the prior distribution, the likelihood and the posterior distribution. Next, Section 2.3 discusses the pros and cons of using a prior distribution in an audit context. In Section 2.4, five methods of constructing a prior distribution are discussed. The last section presents our concluding comments.

2.2 The Bayesian approach to audit sampling

In audit sampling, the goal of the auditor is to make a statement about a certain characteristic, θ, of the population. Generally, the auditor does not inspect the entire population but only a sample, y, from this population. As a consequence, the information about θ from the sample is extrapolated to the population, introducing uncertainty and a probabilistic statement about θ.

The Bayesian way of making a probability statement about the characteristic θ, given the sample y, is through the posterior density $p(\theta \mid y)$. The posterior density is defined through Bayes' theorem as the product of two densities, the prior density $p(\theta)$ and the likelihood function $l(y \mid \theta)$, conditioned on the value of the sample y. Because the marginal probability of the data $p(y)$ is not dependent on θ and with a fixed sample it is a constant, the computation of the posterior density is often shown as follows:

$$\underbrace{p(\theta \mid y)}_{\text{Posterior}} \propto \underbrace{l(y \mid \theta)}_{\text{Likelihood}} \times \underbrace{p(\theta)}_{\text{Prior}}. \tag{2.2.1}$$

As Equation 2.2.1 illustrates, Bayes' theorem implies that the information in the prior distribution is combined with the information in the sample to form the posterior distribution. In the following subsections, we will further elaborate on the prior distribution, the likelihood, and the posterior distribution.

2.2.1 The prior distribution

The prior distribution $p(\theta)$ reflects the auditor's existing information about θ before seeing any information from a sample. An adequate prior distribution assigns a relative plausibility of occurrence to every possible value of θ such that the probability across all possible values of θ is equal to one. Which values of θ are possible depends on the audit question at hand, and the auditor must specify the family of the prior distribution accordingly.

For example, in monetary unit sampling (MUS) the goal is to estimate the amount of misstatement in the population, and monetary units (e.g., individual dollars) are selected and evaluated as individual units. The possible values of the total misstatement amount θ for a population of M monetary units therefore lie in the interval $[0; M]$ and the data are generally assumed to be generated by a Poisson distribution (Stewart, 2012). A gamma prior distribution is a common choice for MUS (Stewart, 2013; Stewart et al., 2007) since it remains a gamma distribution when updated by the information in the data. However, Broeze (2006, pp. 68–71) shows that using a beta distribution as a prior distribution for MUS is also appropriate when using the proportional errors (i.e., taints) of the transactions when evaluating the sample. In the remainder of this chapter, we will demonstrate the five methods for constructing a prior distribution using a beta distribution because it is easy to explain. However, because in some cases proportional errors are unlikely to occur, we included a discussion of the gamma distribution in Appendix 2.A.

Because the total probability that a prior distribution assigns to all possible values of θ is equal to one the prior distribution is a probability distribution, and the auditor can summarize their prior knowledge about θ by calculating location measures such as the mean, median, and mode of the distribution. For example, the prior mode is the most likely value of θ before the sample is analyzed, and the prior mean is the expected value of θ before the sample is analyzed. Similarly, the auditor can summarize the spread of the prior distribution with an x-percent interval that ranges from the $(100 - x)/2^{\text{th}}$ percentile of the prior distribution to the $(100+x)/2^{\text{th}}$ percentile. For example, the 90 percent interval $[\theta_{.05}; \theta_{.95}]$ implies that, with a probability of 90 percent, the population characteristic θ lies between the values $\theta_{.05}$ and $\theta_{.95}$. Similarly, the prior distribution can also be interpreted with respect to the value $\theta_{.95}$ as stating that with 95 percent probability, the value of θ is lower than $\theta_{.95}$. Therefore the value $\theta_{.95}$ can be interpreted as a 95 percent upper bound since 95 percent of the probability mass lies below this value.

In the case of a beta prior distribution, a useful interpretation is that the auditor can interpret the information in the prior distribution as equivalent in information to that of an earlier sample of n transactions containing k errors (Crosby, 1981; Steele, 1992). Since the auditor usually assumes these earlier samples to be correct, they may be deducted from the current intended sample size, increasing audit efficiency (Stewart, 2013; Touw and Hoogduin, 2012).

To put the beta prior distribution in a more concrete context, consider an example from a standard financial audit in which the characteristic of interest θ is generally the misstatement in the population. Suppose an auditor specifies a Beta(1, 1) prior distribution on θ (Figure 2.1), implying that every value of the (proportion of) misstatement is equally likely to occur a priori. Because this prior distribution is flat, the interpretation of this prior in terms of a prior sample is that the information contained in this prior is the information from a sample of size zero.

2.2.2 The likelihood

The likelihood function $l(y \mid \theta)$ reflects the information that the observed data contain about the population characteristic θ. It quantifies the likelihood of the sample outcomes occurring under specific values of θ (Etz, 2018).

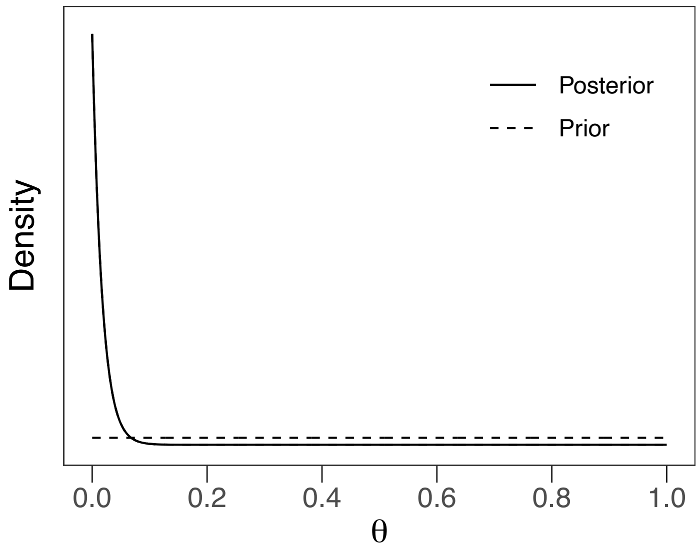


Figure 2.1: Example of a Beta(1, 1) prior distribution on the population misstatement proportion θ. The information in the prior distribution is weighed with the information in the sample, the Binomial($k = 0 \mid n = 58,\ \theta$) likelihood, to form the Beta(1, 59) posterior distribution.

Returning to the previous example with the Beta(1, 1) prior distribution, the likelihood function of observing k misstatements in a sample of size n, with an underlying error rate parameter $\theta \in [0,\ 1]$ is Binomial($k \mid n,\ \theta$). Suppose that the auditor analyzes a sample of $n = 58$ observations, in which they find $k = 0$ misstatements. This implies that the likelihood function of this sample is Binomial($k = 0 \mid n = 58,\ \theta$).

2.2.3 The posterior distribution

The posterior distribution $p(\theta \mid y)$ reflects the auditor's updated knowledge about the characteristic θ after having inspected a sample y from the population. The posterior distribution follows from the fact that each candidate value of θ from the prior distribution induces a prediction about the observed data. Comparing the quality of these predictions from the prior distribution with the actual observed data induces a prediction error. Given the inspected sample y, some values of θ have a lower prediction error than other values. It is this prediction error that holds information about the extent to which the data ought to adjust the plausibility of different values for θ, since the prediction error measures the discrepancy between the auditor's prior information and the information from the sample. Bayes' rule stipulates that values of θ that predict the sample relatively well receive a boost in plausibility, whereas values of θ that predict the sample relatively poorly suffer a decline. By assessing predictive accuracy for the candidate values of θ the prior distribution is updated to the posterior distribution, which consequently contains all information about θ that is available to the auditor.

Because the posterior distribution is a probability distribution the updated knowledge about θ can be assessed by location measures such as the mean, median, and mode of the posterior distribution. For example, the auditor can make a statement about the most likely error in the population by looking at the mode of the posterior distribution. Similarly, percentiles of the posterior distribution can be interpreted in terms of probability (Hoekstra et al., 2014; Kruschke and Liddell, 2018). For instance, a Bayesian auditor will typically say that, with 95 percent certainty, the misstatement in the population is lower than the 95$^{\text{th}}$ percentile of the posterior distribution, $\theta_{.95}$, since 95 percent of the probability mass lies below this value of θ.

In Figure 2.1 the mode of the posterior distribution is $\frac{0}{58} = 0$, implying that the most likely misstatement in the population is zero percent. The posterior mean is 0.017, implying that the expected misstatement in the population is 1.7 percent. The 95$^{\text{th}}$ percentile of the posterior distribution lies at 0.049 meaning that, with 95 percent probability, the misstatement in the population is lower than 4.9 percent.

2.3 Pros and cons of the prior distribution

Using the prior distribution as a point of departure for substantive testing increases transparency and efficiency for both the auditee and the auditor. However, this comes at the cost of justifying the prior distribution.

2.3.1 Pro 1: The prior distribution increases transparency for the auditor

Currently, when the auditor wants to incorporate existing information into their frequentist analysis, they must do so in a manner that is not fully transparent. For instance, the auditor can "use the Audit Risk Model (ARM) to subjectively multiply the risk of material misstatement by the risk of incorrect acceptance of an hypothesis to arrive at an incoherent hybrid overall audit risk" (Stewart, 2013, p. 24).

Bayes' theorem, which explains how existing information is revised by sample evidence, can coherently incorporate this existing information in the prior distribution and update it using the observed sample outcomes. The use of Bayes' rule to intuitively show how the information in the prior distribution is updated by the sample to the information in the posterior distribution has the potential to improve transparency towards auditors.

2.3.2 Pro 2: The prior distribution increases transparency towards stakeholders

Using the prior distribution as a point of departure for substantive testing increases transparency towards the auditee because it requires auditors to be explicit about what information is incorporated into their analysis, and under which assumptions this information is translated into a probability distribution. We believe that this results in two concrete benefits for the auditee.

Firstly, the prior distribution encourages the auditor to quantify their assurance at the highest possible level of the organization as much as possible (de Swart et al., 2013). Business processes at the top of the auditee's organization often provide assurance over processes lower in the organization. Therefore, by spending more effort on obtaining assurance from top-level processes the auditor can achieve a substantial increase in efficiency in a later stage of the audit, potentially lowering audit fees for the auditee. This is also beneficial for the auditee because it places the auditor in a position where they are able to suggest concrete improvements to the auditee at the highest levels of the organization, for example to improve efficiency in later audits. Moreover, the auditee may be inspired to conduct audit activities to incorporate information obtained by the auditor in the audit themselves next year.

Secondly, the prior distribution enables the auditor to optimize their mix of internal controls, analytical procedures, and substantive testing every year. Importantly, the auditee is able to review, verify, or scrutinize the information and controls used by the auditor to construct the prior distribution. For example, the auditee can point out relevant controls to the auditor that have not been included in the prior distribution, or they can provide additional assurance to the auditor by performing analytical procedures themselves. From the perspective of the auditee, the use of the prior distribution by the auditor provides insight into the services paid for and provided.

In the following section, we give several examples of how to efficiently incorporate existing information and make the underlying assumptions explicit.

2.3.3 Pro 3: The prior distribution allows for improved estimation of the misstatement (and can reduce sample size)

If the auditor can justify incorporating appropriate prior information, then this will result in an increasingly precise estimate of the misstatement due to the extra knowledge that is used in the estimation procedure. Moreover, if this prior information implies a reduction in risk profile, incorporating this information into the prior distribution will result in a reduction of the sample size needed to get to the desired assurance about θ.

To illustrate, if a prior distribution is applied that contains almost no information about values of θ (e.g., the Beta(1, 1) prior distribution in Figure 2.1), almost all information required to arrive at a reasonable assurance about θ comes from the information in the sample. Alternatively, if a prior distribution is applied that contains appropriate, risk-reducing, information about θ, less information is needed from the sample to arrive at a reasonable assurance about θ. The top panel in Figure 2.2 illustrates this interchangeability between the incorporated prior information and the information from substantive testing.

As an example, the bottom left panel displays the noninformative prior distribution and the posterior distribution from the previous section. After seeing an error-free sample of 58 transactions, the 95 percent posterior upper bound in this case is 4.9 percent. The bottom right panel shows a scenario in which the auditor has access to appropriate audit information, through which they constructed an informative Beta(1, 19) prior distribution. In the scenario in which the informative prior is constructed, the auditor only needs to inspect a sample of 40 transactions, which amounts to a reduction of 18 transactions as compared to the scenario in which the noninformative prior was constructed. Note that in the informative scenario, combining the informative Beta(1, 19) prior with the Binomial($k = 0 \,|\, n = 40,\, \theta$) likelihood of the sample, the posterior distribution is again a Beta(1, 59) distribution whose 95 percent upper bound lies at 4.9 percent.

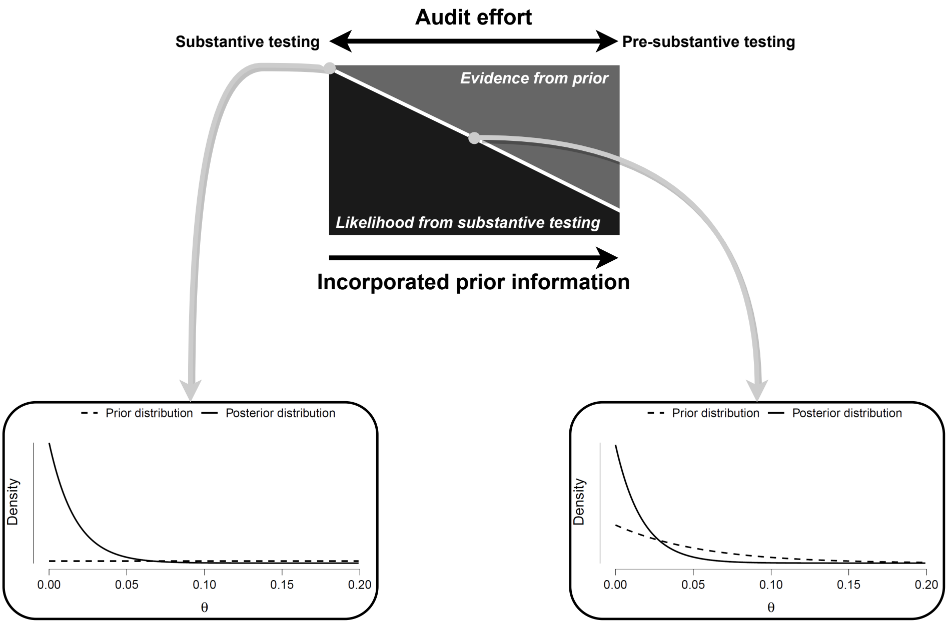


Figure 2.2: The top panel illustrates the balance between the amount of incorporated prior information and the required information (likelihood) from substantive testing. The Beta(1, 1) prior distribution in the bottom left panel contains little prior information and combined with a Binomial($k = 0 \,|\, n = 58,\, \theta$) likelihood, forms a Beta(1, 59) posterior distribution. The Beta(1, 19) prior distribution in the right panel contains more information and therefore requires less information from a sample to reach the same assurance. Combined with a Binomial($k = 0 \,|\, n = 40,$ θ) likelihood, the posterior distribution is again a Beta(1, 59) distribution.

The latter scenario demonstrates the added value of the prior information since the auditor requires less information from substantive testing to arrive at the same amount of assurance about θ. Incorporating appropriate information into the prior distribution can, therefore, increase efficiency and transparency in the audit by coherently reducing the required sample size in substantive testing. The assumptions behind a noninformative prior distribution are quick to justify, but the absence of information contained therein results in a larger sample size than may be necessary.

2.3.4 Con 1: Justification of the prior distribution takes time and effort

An informative prior distribution is efficient in the sense that it allows for improved estimation of the misstatement and a potential reduction in sample size. However, the auditor must realize that the information that is incorporated into the prior distribution needs to be justified. Hence, the more information that is incorporated into the prior, the more work is required to substantiate why the incorporated information is appropriate for the population, and how it is incorporated into the prior distribution. Attention should be given to the amount of time and effort it takes to substantiate the information in the prior distribution for a substantial reduction in audit work versus the time and effort it takes to perform the reduction in audit work. Performing the Bayesian analysis makes most sense if the profit of the Bayesian analysis—the time and effort that is saved by reducing in sample size—outweighs the costs—the time and effort it takes to justify the Bayesian analysis.

Whether the pros of constructing a prior distribution outweigh its cons depends on the time and effort it takes to select and audit an extra sample. If the costs of selecting an extra sample exceed the time and effort to set up an informative prior distribution, a Bayesian analysis can be a profitable alternative to a frequentist analysis. For example, suppose that the auditor is performing an on-site indoor air quality audit (e.g., Asadi et al., 2013) for the auditee's office buildings around the world. They have the choice of either taking a large sample or performing an analytical procedure to construct the prior distribution. The analytical procedure in question correlates a building's air quality with its energy consumption, data which is available in digital format to the auditor. Of course, constructing the prior distribution on the basis of this correlation is not a trivial task and might take substantial time and effort. However, because the auditee has offices all around the world, travelling to—and inspecting—each building takes a significant amount of time and money. In these cases, it is likely that the possible reduction in sample size achieved by the information from the correlation analysis outweighs the time and effort that goes into specifying the prior distribution. Even if the auditor has easy access to the auditee and therefore decides that constructing the prior distribution is not worth the time and effort, they can fall back to a prior distribution that incorporates no existing information. The advantage of the Bayesian approach is that it provides the auditor with the flexibility to choose either of these options.

Finally, we are aware that the prior distribution is often perceived as difficult to construct and that this can be seen as a con as well. However, in the next section we attempt to resolve this by showing five methods of constructing a prior distribution. The specification of the prior distribution is complex because the available information must be represented by a probability distribution. Earlier work on prior distributions in an audit context has mainly focused on eliciting one directly from auditors' professional judgment (Abdolmohammadi, 1985, 1987). For instance, Chesley (1978) asked auditors to assign probabilities to specific values of θ (i.e., by eliciting a cumulative probability function). A reversed method was examined by Crosby (1981), who asked auditors to assign specific values of θ to probabilities (i.e., by eliciting values for the first, second, and third quartile of a probability density function). Such methods allow for coherent integration of expert knowledge about θ into the analysis, but require that the auditor has an advanced understanding of statistical concepts (e.g., probability distributions), something that is not necessarily part of an auditor's core expertise.

In the following section, we provide alternatives to such expert elicitation methods by discussing how various other sources of audit information can be formally incorporated into the prior distribution. The proposed methods provide a logical translation from audit information to a prior distribution, thereby making the specification of the prior distribution less daunting and complex for the auditor. Moreover, we have implemented these methods in statistical software so that they are available to auditors who are not experts on statistics but are willing to use the software approved by their technical office. This way, constructing a prior distribution on the basis of their already existing information requires no additional time and effort from the auditor. We show that these methods have the potential to increase audit efficiency and transparency.

2.4 Integrating information into the prior distribution

We discuss five methods for integrating existing information into the prior distribution. The first method is not informative about the characteristic θ, whereas the other four methods integrate four types of information about θ into corresponding prior distributions: information regarding the prior probability of (in)tolerable misstatement occurring in the population, information from earlier (implicit) samples, historical information from last year's audit, and information from analytical procedures. For ease of explanation, the following methods assume that the auditor does not expect to find any misstatements in the sample. Note that the calculations are very similar when misstatements are expected (the resulting sample sizes would of course be different).

2.4.1 Method 1: No explicit information

The auditor can refrain from expressing an explicit opinion about θ by incorporating as little information as possible in the prior distribution, so that the resulting posterior distribution relies solely on the information from the sample (Blocher, 1981; Martel-Escobar et al., 2018). Consider the scenario in which the auditor has no access to any existing information about which values of θ are more plausible a priori. Then, a suitable prior distribution is the previously introduced Beta(1, 1) distribution as it assigns equal prior probability mass to all values of θ, and the posterior mode (most likely misstatement) is $\frac{k}{n}$, which is equal to that of a frequentist approach (Albert, 2003; Tuyl et al., 2008).

Although the Beta(1, 1) distribution is often a default choice due to its connection to frequentist methodology and its ease of application, it implies a conservative prior opinion about θ with respect to the upper bound of tolerable misstatement in the population. More specifically, the Beta(1, 1) distribution expresses the prior opinion that intolerable misstatement is highly likely to occur in the population. To illustrate, for an upper bound of tolerable misstatement of 5 percent, the Beta(1, 1) prior distribution assumes that, with 95 percent probability, the total misstatement in the population is larger than the upper bound of tolerable misstatement. If this reflects the auditor's true existing information about the misstatement in the population, a full inspection might arguably be a better choice.

By choosing a flat distribution no information is incorporated into the prior distribution. Therefore, the Beta(1, 1) prior distribution is not advised when the auditor has access to information about the characteristic θ. However, it can be useful in situations where the auditor, for conservative reasons, wants to specifically refrain from expressing an opinion about θ, wants to retain some of the properties of a frequentist analysis, or when the auditor wants to have a benchmark analysis for a more informed Bayesian analysis.

2.4.1.1 Example to detect overstatements

Suppose the auditor must audit a new client and wants to determine the misstatement θ in a population. Because they are auditing a new client and do not have access to existing information about this population, they want to state a noninformed opinion with respect to θ. Since the prior distribution contains no information about θ, there is no need for the auditor to justify the information in the prior distribution.

Prior to calculating a sample size, the auditor makes an assessment of the expected (tolerable) errors in the sample. Assuming a Beta(1, 1) prior distribution, and applying Bayes' theorem to find the minimum sample size $n_{\text{Method 1}}$ such that if no errors are detected in the sample, the posterior distribution has a 95 percent upper bound of 5 percent, the auditor finds $n_{\text{Method 1}} = 58$. Using the Beta(1, 1) prior is slightly more efficient than a frequentist analysis, in which the auditor must inspect $n_{\text{freq}} = 60$ observations to reduce the sampling risk sufficiently (American Institute of Certified Public Accountants (AICPA), 2019). However, the posterior most likely error resulting from a Beta(1, 1) prior distribution is equal to that of a frequentist analysis, and is zero after seeing this error-free sample.

2.4.2 Method 2: Information about the probability of (in)tolerable misstatement

The auditor can incorporate existing information regarding the prior probability of (in)tolerable misstatement occurring in the population in the prior distribution on θ. In contrast to the expected misstatement, the probability of misstatement represents the auditor's assessment about how likely it is that the population contains material misstatement. Auditors generally state their opinion on the population misstatement using the posterior most likely error (ISA 450, ISA 530; International Auditing and Assurance Standards Board (IAASB), 2018) after they have assessed that their work was sufficient for such statement by comparing the upper bound of the posterior distribution to the performance materiality θ_{max} (ISA 320; International Auditing and Assurance Standards Board (IAASB), 2018). We propose to use the value of θ_{max} as an anchoring point for the prior distribution. Anchoring the prior distribution on the performance materiality allows the auditor to exploit the areas under the prior distribution above and below θ_{max}. These areas respectively express the prior information about the probability of the population misstatement being tolerable or not before a sample is selected and analyzed.

When using a beta distribution, the area under the prior distribution that lies below the performance materiality $P_{-} = \int_0^{\theta_{max}} \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} d\theta$ quantifies the prior probability of tolerable misstatement occurring in the population, while the area that lies above the performance materiality $P_{+} = \int_{\theta_{max}}^1 \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} d\theta$ quantifies the prior probability of intolerable misstatement occurring in the population. Given the auditor's existing information about the probability of (in)tolerable misstatement occurring in the population, and given the expected misstatement in the sample, the α and β parameters of the beta prior distribution are defined.

To illustrate how anchoring works in this context, suppose the auditor sets the performance materiality at 5 percent ($\theta_{max} = 0.05$). Furthermore, they expect to find no misstatements in the sample (implying that the prior parameter $\alpha = 1$) and assume a priori that tolerable misstatement is equally plausible to occur in the population as intolerable misstatement (implying that $P_{-} = P_{+} = 0.5$). Since the prior probability of (in)tolerable misstatement is specified by the area under the prior distribution, equal prior probabilities can be achieved by setting the median of the beta distribution to the performance materiality θ_{max}. When $\alpha = 1$, the median of the beta distribution can be expressed as $\theta_{max} = \theta_{.50} = 1 - 2^{-\frac{1}{\beta}}$ (Kerman, 2011). Consequently, the β parameter of the prior distribution can be determined as $\beta = \frac{\ln(P_{+})}{\ln(1-\theta_{max})} = \frac{\ln(0.5)}{\ln(1-0.05)} = 13.51$. Therefore, the auditor's existing information regarding the prior probability of (in)tolerable misstatement occurring in the population is incorporated into a Beta(1, 13.51) distribution.

2.4.2.1 Example to detect overstatements

We will now apply this method to a practical example from Harrison et al. (2002). The auditee has an internal control where software verifies the general ledger for transactions created by employees in the field. The auditor has information that on days that show a high sales activity across the country, one of the employees can disable the control system to get orders processed faster. To gain assuring evidence that the general ledger contains no material misstatement, the auditor might perform substantive testing of transactions that were recorded on days where the control was disabled. Suppose the auditor finds that this control is a critical check for 60 percent of all sales transactions, and so 60 percent of sales transactions made on the days that the system was disabled was not checked.

The prior distribution can be justified by the existing information that 60 percent of the transactions in the population might contain misstatement. Thus, the auditor wants to incorporate the existing information that the prior probability of intolerable misstatement is 60 percent. The auditor specifies $P_{+} = 0.6$ and plans for zero tolerable deviations, thereby assuming a Beta(1, 9.96) prior distribution.

Applying Bayes' theorem to find the sample size $n_{\text{Method 2}}$ such that if no errors are detected in the sample, the posterior distribution has a 95 percent upper bound of 5 percent, the auditor finds $n_{\text{Method 2}} = 49$. Compared to the benchmark Bayesian analysis (method 1) this method results in a reduction in the sample size of $n = 9$. Compared to the standard frequentist method where no prior information is incorporated the reduction in sample size amounts to $n = 11$.

2.4.3 Method 3: Information from earlier samples

The auditor can incorporate information from a prescribed reduction in the required sample size in the prior distribution on θ. Consider for example the scenario where the auditor can reduce their required sample size based on the risk assessment of the population, as determined by their audit guide. Since the information in the prior distribution can be interpreted as equivalent to that of an earlier sample, the auditor can use the reduction in sample size to construct their prior distribution (de Swart et al., 2013; Steele, 1992).

An application for this method can be found in the auditing standards, which state that the auditor may act on their collected audit information through the audit risk model (ARM) by reducing their required sample size in substantive testing (International Auditing and Assurance Standards Board (IAASB), 2018). The ARM provides an association between the specified audit risk and the assessed risk of material misstatement; see Equation 2.4.1.

$$\text{Audit risk} = \text{Inherent risk} \times \text{Control risk} \times \text{Detection risk} \tag{2.4.1}$$

According to the ARM, the audit risk is divided into three constituents: inherent risk, control risk and detection risk. Inherent risk is the risk of a material misstatement due to an error in a financial statement before consideration of any related controls. Control risk is the risk of a material misstatement not prevented or detected by the internal control systems of the auditee (e.g., computer-managed databases). Detection risk is the risk that an auditor will fail to find material misstatements that exist in the auditee's financial statements. According to the ARM, the auditor is only involved in the detection risk and may adjust this accordingly to accommodate the other two risks so that the acceptable level of audit risk is retained. Therefore, a lower assessed risk of material misstatement may allow a higher tolerable detection risk and, in turn, requires less persuasive audit evidence (International Auditing and Assurance Standards Board (IAASB), 2018). ISA 330 and ISA 530 prescribe that a lower quantity of evidence from substantive testing applies in these situations.

Currently, it is unclear how risk assessments can reduce the required sample size. Sampling manuals from audit firms and other institutions that perform substantive testing show a large variation in the extent to which reliance on risk assessments reduces the required sample size. Table 2.1 displays examples of the sample size reduction factor R (Touw and Hoogduin, 2012). To calculate R we need—for a given upper bound of tolerable misstatement—n_+: the largest sample size in the manual, and n_-: the smallest. The largest sample size n_+ is often the same number, based on 95 percent confidence and zero expected misstatements in the sample. The smallest sample size n_- is based on the assumption of a low risk of material misstatement and a system of internal controls that functions well. Of course, the words 'low' and 'well' are subjective, and therefore, the number n_- varies when different audit sampling manuals are considered (see Table 2.1). By incorporating their risk assessments through the prior distribution, the auditor is able to make the relationship between these risk assessments and the required sample size explicit.

Table 2.1: The sample size reduction factor $R = {}^{n_+}/_{n_-}$ in various audit manuals.

	R
Big 4 office	15
NL public body	2.7
EU public body	3
NL tax and customs	8
NL Tier2 office	3

Table 2.1 shows that a reduction in sample size is common practice, and prescribed by the ISA. As explained at the beginning of this section, this reduction $\Delta n = n_+ - n_-$ can be incorporated in the prior distribution (assuming that the unseen samples contain no misstatements). More specifically, if the auditor expects no misstatements in the sample then $\alpha = 1$. Next, Δn can be incorporated into the beta prior distribution by setting the β parameter of the prior distribution to $1 + \Delta n$. For example, if the auditor, based on their assessments of inherent risk and control risk, has reduced the required sample by 20 transactions, this information is incorporated in a Beta(1, 21) prior distribution.

Vice versa, instead of constructing the prior distribution based on a reduction in the sample size, the prior distribution can also be constructed on the basis of a sample that is already analyzed. Suppose that the auditor has taken the required sample containing 58 transactions, but found one deviation and therefore cannot conclude with 95 percent certainty that the misstatement is lower than the performance materiality of 5 percent. The posterior distribution in this scenario is a Beta(2, 59) distribution, which has its 95$^{\text{th}}$ percentile at 7.6 percent. However, instead of judging that the population contains material misstatement, the auditor wants to perform additional testing on the population. By using the posterior distribution as a prior distribution for a second sample, the auditor can extend their analyzed sample without any penalty (Dienes, 2011; Rouder, 2014; Wagenmakers et al., 2008). This is different than in a frequentist approach, where the auditor is unable to coherently extend testing of their analyzed sample while preserving the intended audit risk. Using the Beta(2, 59) posterior distribution as a prior distribution, the auditor can maintain the required assurance on θ by inspecting 33 extra error-free samples from the population. These 33 samples together with the 58 original samples coherently add up to $n = 91$, the sample size that is required when using a Beta(1, 1) distribution and planning for one expected misstatement in the sample.

2.4.3.1 Example to detect overstatements

Suppose an auditor is testing for an upper bound of tolerable misstatement of 5 percent and is planning for zero expected misstatements in the sample. Furthermore, suppose that their audit guide prescribes a sample of 58 transactions when both inherent risk and control risk are high. However, the auditor has collected evidence that both inherent risk and control risk can be judged as medium instead of high. When both risks are judged as medium, their audit guide postulates that the required sample size for zero expected misstatements is 32.

The prior distribution can be justified by the reduction in required sample size. Using the reduction in the required sample size, the auditor specifies $\beta = 1 + \Delta n = 27$, and arrives at a Beta(1, 27) prior distribution.

Using the Beta(1, 27) prior distribution, and applying Bayes' theorem to find the sample size $n_{\text{Method 3}}$ such that if no errors are detected in the sample, the posterior distribution has a 95 percent upper bound of 5 percent, the auditor again finds $n_{\text{Method 3}} = 32$. Compared to the benchmark Bayesian analysis (method 1), the reduction in sample size is $n = 27$. Compared to the standard frequentist method where no prior information is incorporated the reduction in sample size amounts to $n = 29$.

2.4.4 Method 4: Information from last year's audit

The auditor can incorporate information from last year's results in the prior distribution on θ. A method for letting this historical information inform the current audit is suggested by van Batenburg and Kriens (1989), who consider last year's posterior distribution as a point of departure. If the result of last year's audit was positive, this posterior distribution generally has its 95 percent upper bound below the performance materiality. Assuming the auditor wants to exploit the information collected in year $t-1$ in the prior distribution for year t, it is reasonable to state a prior distribution on θ for year t that is equal to the posterior distribution of year $t-1$.

Using the posterior distribution from year $t-1$ directly as a prior distribution for year t implies that the estimated maximum misstatement in the population is already lower than performance materiality being the required maximum tolerable misstatement. The justification for such a prior distribution is grounded in the case where the population from year t is the same as that of year $t-1$. No additional samples would be required to achieve the required assurance about θ. Of course, the auditor shall have to decide the extent to which last year's population is compatible with that of this year. The assumption that the two are completely equivalent usually does not hold, as the population in the current year will only be comparable to that of the previous year to a certain extent. However, customer, product, quantity, and price of certain transactions might be equivalent, and the auditor can incorporate the extent of this equivalence in the prior distribution on θ.

To quantify the auditor's knowledge on the equivalence of the populations in year t and year $t-1$ a weighting factor f can be introduced to the beta prior distribution (van Batenburg et al., 1994). If the auditor finds that the two populations are comparable, they can assign a higher weight to the results of last year's audit and the value for f should fall close to one. For example, if 70 percent of the transactions in the general ledger consists of transactions that are equivalent (with respect to the customer, product, quantity, and price) to those that were audited in last year's audit, the auditor can specify $f = 0.7$. However, if they think that the populations are incomparable, the value for f should fall close to zero. For example, suppose major errors were found in year $t-1$ after which the auditee implemented controls that make it very unlikely that these errors occur in year t. Since the major errors in the population of year $t-1$ are unlikely to occur in the population of year t, the information from year $t-1$ should be taken into account limitedly. This is reflected in a low prior probability for this information.

To illustrate this method, suppose the auditor has access to last year's posterior distribution Beta(α_{t-1}, β_{t-1}), which has its 95$^{\text{th}}$ percentile below the performance materiality. Given a value for f, the corresponding prior distribution can be determined as Beta($1 + (\alpha_{t-1} - 1) \times f$, $1 + (\beta_{t-1} - 1) \times f$).

2.4.4.1 Example to detect overstatements

Suppose last year the auditor performed substantive testing without incorporating any existing information. They inspected a full sample of 58 records and found that zero transactions were misstated. The posterior distribution from year $t-1$ is a Beta(1, 59) distribution, which has its 95$^{\text{th}}$ percentile at 0.049. Consequently, the auditor inferred that, with 95 percent confidence, the misstatement in the population was below the upper bound of tolerable misstatement of 5 percent.

The prior distribution can be justified by the equivalence between the populations of last year and those of the current year. This year, the auditor has collected information that the populations of last year and this year are comparable to the extent of 60 percent ($f = 0.6$) since 60 percent of customers, products, quantity, and price are equivalent. New customers in the population (40 percent) have not been subject to the validated controls of last year. Therefore, the auditor specifies $f = 0.6$ and determines $\beta = 35.8$.

Using a Beta(1, 35.8) prior distribution, and applying Bayes' theorem to find the sample size $n_{\text{Method 4}}$ such that if no misstatements are detected in the sample, the posterior distribution has a 95 percent upper bound of 5 percent, the auditor finds $n_{\text{Method 4}} = 23$. Compared to the benchmark Bayesian analysis (method 1) this method results in a reduction in the sample size of $n = 35$. Compared to the standard frequentist method where no prior information is incorporated the reduction in sample size amounts to $n = 37$.

2.4.5 Method 5: Information from analytical procedures

The auditor can incorporate information from analytical procedures (e.g., a benchmark analysis) in the prior distribution on θ. In the previously discussed methods of constructing a prior distribution based on existing knowledge, we have described the procedure by which this knowledge can be translated into the prior distribution. However, there is no such clear procedure for information acquired from analytical procedures, since these procedures can vary strongly depending on the type of information that is incorporated in the prior. The following approach thus necessitates a stronger substantiation of its data and assumptions, and how these assumptions are incorporated into the prior distribution. However, given the increasing availability of big data from other sources than the general ledger, such as sensor data or financial data from external vendors, we foresee that the possibility to apply a wide range of analytical procedures will grow rapidly in the near future. Bayesian methods offer a way to stack these procedures and determine how much additional comfort is needed from substantive testing to come to a reasonable conclusion about θ.

2.4.5.1 Example to detect overstatements

For example, Stringer and Stewart (1986) introduced statistical techniques to assess potential misstatements based on, for instance, regression models. A concrete example of such a model is benchmarking the relationship between sales and costs of sales within the auditee's industry sector. This relationship can be modelled by a linear equation:

$$C = \beta_0 + \beta_1 \cdot S + \epsilon, \tag{2.4.2}$$

where C and S denote the auditee's Cost of Sales and Sales, respectively. In practice, this relationship is often more complex than is presented in Equation 2.4.2, and the auditor must carefully construct and evaluate the applied regression model. However, for ease of understanding we will continue our example with this simplified model.

After estimating the parameters β_0 and β_1 from a data set consisting of values of C and S for peer companies, the auditor must check the assumptions underlying their linear regression. If these assumptions hold, the prediction for the costs of sales of the auditee, given the actual sales of the auditee, can be derived in the form of a normal probability distribution. In a typical analytical procedure, this prediction of the cost of sales is summarized in the form of a 90 percent two-sided interval. Next, the auditor verifies whether the booked cost of sales by the auditee falls within the predicted interval. If it does, then less evidence from substantive testing is required than if the booked cost of sales falls outside of the predicted interval.

The procedure described above is a frequentist one. However, similar procedures within the Bayesian philosophy have been proposed before. For instance, Deakin and Granof (1974) and Kinney and Bailey (1976) noted that regression analysis can be used to revise the auditor's prior probabilities of the population being misstated or not by testing whether the auditee's booked cost of sales is different from the predicted cost of sales.

The prior distribution can be justified by the data and the auditee's numerical prediction of the cost of sales. In this analytical procedure, our proposal for the prior distribution on θ is to use the relative error distribution from the linear regression. The relative error distribution is the normal distribution (Normal(μ, σ)) that captures the uncertainty of the prediction of the cost of sales by means of the linear regression, scaled to be a percentage of the total cost of sales. The mean μ of the prior distribution on θ is the relative deviation of the auditee's booked cost of sales when compared to the predicted cost of sales according to the benchmark data $\frac{C-\hat{C}}{C}$. The standard deviation of the prior distribution on θ, induced by the benchmark data, is expressed by the standard deviation of the distribution of ϵ. We propose to use this distribution as the prior distribution on θ to quantify the amount of substantive testing needed on top of this analytical procedure to conclude that the audit risk is sufficiently small enough.

Suppose that the auditor is assessing the risk of including fraudulent bribery payments at the auditee's organization. To perform an audit of the payments, the auditor needs to investigate both the auditee's costs of sales and their actual sales. The sum of the booked costs of sales is \$223,994,405 and the sum of the sales is \$298,112,312, respectively. The allocated performance materiality has been set to \$112,500, or 5 percent of the booked cost of sales. The existing information is a benchmark of audited figures on the cost of sales and actual sales of peer companies ($n = 100$) in the auditee's industry group. These benchmark data are plotted in Figure 2.3.

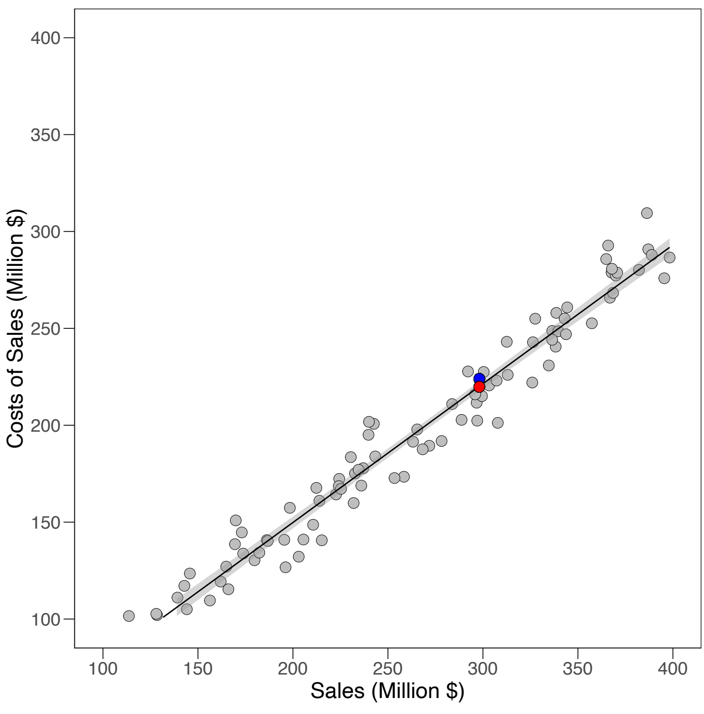


Figure 2.3: Scatter plot of the cost of sales C (million \$) versus the actual sales S (million \$) for peer companies of the auditee. The blue and red dot indicate the auditee's booked and expected costs of sales based on the benchmark, respectively.

The auditor estimates the parameters in Equation 2.4.2 using linear regression leading to $\beta_0 = 241,300$ and $\beta_1 = 0.7366$. This gives the following estimate $\hat{C}$ of the auditee's costs of sales:

$$\hat{C} = \$241,300 + 0.7366S = \$219,817,866. \tag{2.4.3}$$

The auditor confirms that ϵ is normally distributed having a standard deviation of $\sigma_\epsilon = \$11,090,408$. Assuming that the benchmark data is representative for the auditee, the auditor can incorporate the information from this benchmark in the Normal(μ, σ) prior distribution on θ. The mean μ of the prior distribution on θ is the relative deviation of the auditee's booked costs of sales when compared to the predicted cost of sales according to the benchmark data $\frac{C-\hat{C}}{C} = \frac{223,994,405-219,817,866}{223,994,405}$, and equals 1.9 percent. The standard deviation σ of the prior distribution on θ can be determined as $\frac{\sigma_\epsilon}{C} = \frac{11,090,408}{223,994,405}$, and is 5 percent. Since the auditor is focusing on overstatements only, the prior distribution is truncated to the interval [0; 1]. Finally, the auditor has arrived at the truncated Normal(0.019, 0.05) distribution as the prior distribution on θ.

Applying Bayes' theorem to find the sample size $n_{\text{Method 5}}$ such that, if no errors are detected in the sample, the posterior distribution has a 95 percent upper bound of 5 percent, the auditor finds $n_{\text{Method 5}} = 50$. Compared to the benchmark Bayesian analysis (method 1) this method results in a reduction in the sample size of $n = 8$. Compared to the standard frequentist method where no prior information is incorporated the reduction in sample size amounts to $n = 10$.

Keeping in mind that there are in principle infinite possibilities to use all sorts of data and statistical learning methods to create prior distributions, this approach is much more generic than those in the previous sections. The only restriction on the statistical learning method used in the analytical procedure is that this method not only delivers a most likely value for the audited figure but also a probability distribution of the method's error. The commonly used linear regression certainly meets this requirement. However, by using validation samples next to training samples, even nonparametric methods can be equipped with a way to derive a probability distribution for the error, so we do not see this restriction as a severe one.

2.4.6 Comparison of posterior distributions

To illustrate the effect of incorporating the existing information from the previous subsections on the auditor's final judgment about θ, Figure 2.4 shows each of the prior distributions (left panel) and their corresponding posterior distributions (right panel) after inspecting a sample of 30 transactions, of which zero transactions were misstated. The figure illustrates that prior distributions that incorporate increasingly stronger audit information (i.e., assign more prior probability mass to lower values of θ) result in posterior distributions that assign more probability mass to lower values of θ.

Table 2.2 shows the posterior mode θ_{mle}, the corresponding 95 percent upper bounds of the posterior distributions $\theta_{.95}$, and the margin of error of the estimate

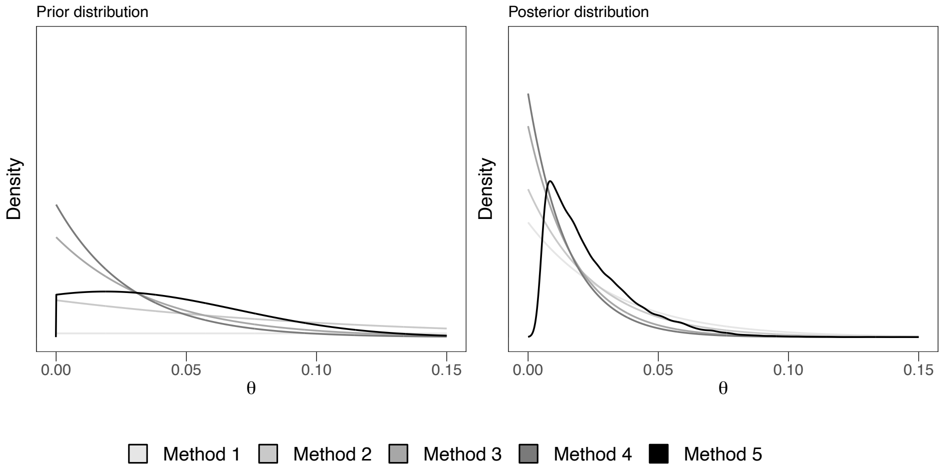


Figure 2.4: Prior distributions (left panel) and posterior distributions (right panel) for each of the five discussed methods after inspecting a sample of $n = 30$ transactions, of which $k = 0$ contained a misstatement. As can be seen from the figure, prior distributions that initially assign more probability mass to lower values of θ also result in posterior distributions that assign more probability mass to lower values of θ.

Table 2.2: Description of prior distributions and the corresponding posterior distributions after taking a sample of $n = 30$, $k = 0$. Each of the Beta(α, β) prior distributions is combined with a Binomial($k = 0 \,|\, n = 30$, θ) likelihood to form a Beta($\alpha + k$, $\beta + n - k$) posterior distribution. The truncated Normal(μ, σ) prior distribution has no analytical solution for the combination with a Binomial($k = 0 \,|\, n = 30$, θ) likelihood and the posterior distribution is therefore determined via Hamiltonian Monte Carlo (HMC) sampling. The mean of the HMC samples is given as the posterior mode θ_{mle}.

Method	Prior	Posterior	θ_{mle}	$\theta_{.95}$	$\theta_{.95} - \theta_{mle}$
1	Beta(1, 1)	Beta(1, 31)	0	0.0921	0.0921
2	Beta(1, 9.96)	Beta(1, 39.96)	0	0.0722	0.0722
3	Beta(1, 27)	Beta(1, 57)	0	0.0512	0.0512
4	Beta(1, 35.8)	Beta(1, 65.8)	0	0.0445	0.0445
5	Truncated Normal(0.019, 0.05)		0.0243	0.0660	0.0417

$\theta_{.95} - \theta_{mle}$ (i.e., the difference between the upper bound and the mode of the posterior distribution). Compared to the benchmark Bayesian analysis in method 1, the other prior distributions result in a more precise estimate of the population misstatement. Moreover, as can be seen from the 95 percent upper bounds, only the posterior distribution described in method four has its 95$^{\text{th}}$ percentile below 5 percent, and thus supports the statement that the population misstatement is, with 95 percent probability, lower than the performance materiality of 5 percent. The prior distribution from this example incorporates relatively strong audit information since it assumes that 60 percent of all transactions in the population were equivalent to those of last year with respect to customer, product, quantity, and price. Table 2.2 also shows that the most likely error is often estimated to be zero, except for the most likely error resulting from the prior constructed in method five. That is because, even though the prior distribution incorporated assuring audit information, it also assumed a prior most likely error of 1.9 percent.

2.5 Concluding comments

Over the years, there has been a need for efficient use of expert knowledge and existing information in audit data analytics (Appelbaum et al., 2017; Chesley, 1975). We have shown that applying Bayesian statistics allows for incorporation of this knowledge and information into the statistical analysis. Next, we have outlined five methodologies that auditors can use to incorporate this information into a prior distribution. We are convinced the auditor can only face the growing challenges of today's auditing field, ensuring maximum audit quality while maintaining high efficiency, by tailoring their audit specifically to the situation of the auditee. Bayesian statistics allows for coherent incorporation of many sources of acquired audit information into the statistical sampling procedure, allowing auditors to build on existing information, and offering them the flexibility to control how they arrange their activities to aggregate audit evidence over the audit as a whole. Especially in today's critically examined audits, Bayesian statistics provides a transparent and efficient manner of auditing.

For convenience, we have demonstrated the five methods for constructing a prior distribution using a beta distribution. However, the methods we have discussed are generic. When applied to a different situation (e.g., a gamma distribution), the specifics of the equations are different, but the incorporated prior information remains the same. A comparison of the beta and gamma distributions for MUS is beyond the scope of this chapter, and we refer to the work of Stewart (2013, pp. 62–70) for further reading. However, we have included illustrations of the discussed methods using a gamma distribution in Appendix 2.A.

When drawing inferences in the Bayesian framework, the auditor has to consider that substantive test work, including sample sizes, can be reduced via a prior distribution. If a prior is weakly informative, the potential reduction in sample size is small. On the other hand, if a prior is very informative, the potential reduction in sample size is large. However, this large potential reduction in sample size comes at a cost, as the auditor needs to show, firstly, that the existing information is valid and relevant for the audit, and secondly that the translation of this information into the prior distribution is appropriate. We have discussed the tools to translate this existing information into the prior distribution.

However, in practice the existing information is often prone to measurement error. To the extent that this measurement error exists, it is important that the auditor takes the quality of the information into account when constructing a prior distribution. It can be useful for the auditor to work together with a statistician to determine if, and how, the existing information should be translated into a probability distribution. As more complex information is incorporated into the analysis, the role of the statistician becomes increasingly important to assure that the information is incorporated accurately into the prior distribution.

In our approach, the auditor constructs the prior distribution on the basis of audit evidence (Methods 2–5), which means that there is information to justify the prior distribution. In the case where this is not possible, the auditor is able to fall back to the trivial prior distribution which reflects no existing information (Method 1). Because in both these scenarios it is transparent how the prior distribution is constructed and what information it incorporates, it can be scrutinized by other stakeholders, (internal and external) reviewers, or team members. For example, transparency towards a regulator can be given in the scenario where the regulator is trying to trace the steps of the auditor, or wants to scrutinize the auditor's judgment.

However, to mitigate critique that the auditor can reason to a foregone conclusion based on the specific choice of prior distribution they can assess the robustness of their outcomes to the choice of the prior distribution. This can be done via a sensitivity analysis (Hoijtink et al., 2019; Liu and Aitkin, 2008), where the auditor considers the amount of achieved assurance or estimated maximum misstatement for different prior specifications. Sensitivity analyses have been proposed before in auditing (Martel-Escobar et al., 2005), as well as in other scientific fields, to see if evidence for a particular scenario is relatively stable across a range of prior beliefs, suggesting that the statistical analysis yields conclusive results in multiple scenarios. The auditor can assess this sensitivity to the prior distribution by changing its parameters, or by changing the family of the prior distribution. For example, the auditor can compare their informative Beta(α, β) prior distribution with the benchmark Beta(1, 1) prior distribution, or calculate the results using a gamma prior distribution.

We are aware that, despite its broad use in the audit practice, strong arguments against the use of the Audit Risk Model (ARM) as shown in method 3 have been formulated. The main objection is that the constituents of the audit risk influence each other, and should therefore not be treated as independent probabilities (Cushing and Loebbecke, 1983; Jiambalvo and Waller, 1984; Kinney, 1983; Leslie, 1984). Nonetheless, the ARM is widely used and accepted in practice. We would like to highlight the fact that we do not attempt to criticize the validity of the ARM. In fact, in Method 3 we build on the ARM by transforming assessments of inherent risk and control risk into a prior distribution. We view it as a successful model demonstrating a methodological approach to structuring audit risk (Arzhenovskiy et al., 2019).

Finally, we have attempted to eliminate the difficulty of auditors determining the proper distributional family, specifying the prior, conducting sensitivity analyses, and documenting these choices, by implementing these methods in the R package 'jfa' (Derks, 2022) and JASP for Audit (Derks et al., 2021b), a module for the freely available and open-source statistical software program JASP (JASP Team, 2022) that is designed to facilitate Bayesian statistical auditing. We hope that, by implementing these methods in a digital environment for which no extensive programming knowledge is required, they can be of aid to researchers and auditors interested in using Bayesian methods in audit sampling.

2.A Illustrations using the gamma distribution

The Gamma(α, β) distribution has a shape parameter α and a rate (inverse scale) parameter β that determine its shape. The gamma prior distribution is a common choice for monetary unit sampling, in which a sample of n transactions containing a total of k proportional errors (taints) is selected from a population of M monetary units. The Gamma(α, β) distribution is combined with the Poisson($\lambda = k\frac{M}{n} \mid \theta$) likelihood to form a Gamma($\alpha + k$, $\beta + \frac{1}{M/n}$) posterior distribution. The most likely error (mode) of the distribution, given $\alpha \geq 1$, is $\frac{\alpha-1}{\beta}$.

2.A.1 Method 1: No explicit information

The posterior distribution resulting from a Gamma(1, 0) prior distribution depends solely on the information in the sample since its posterior most likely error is $k\frac{M}{n}$. This specific prior distribution is improper, however, meaning that gives an infinite amount of probability mass to positive values of θ.

2.A.2 Method 2: Information about the probability of (in)tolerable misstatement

For the gamma distribution, the areas corresponding to tolerable and intolerable misstatement, respectively, are given below.

$$P_{-} = \int_{0}^{\theta_{max}} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \qquad P_{+} = \int_{\theta_{max}}^{M} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta. \tag{2.A.1}$$

Given an assessment of the probability of tolerable misstatement $P_{+} = 1 - P_{-}$ and planning for zero expected deviations ($\alpha = 1$), the β parameter of the gamma prior distribution can be determined as $\beta = -\frac{\ln(P_{+})}{\theta_{max}}$. The prior distribution is therefore a Gamma(1, $-\frac{\ln(P_{+})}{\theta_{max}}$) distribution.

2.A.3 Method 3: Information from earlier samples

Considering the reduction in required sample size $\Delta n = n_{+} - n_{-}$, the β parameter of the gamma distribution can be set to $\beta = \frac{1}{M/\Delta n}$. The prior distribution is therefore a Gamma(1, $\frac{1}{M/\Delta n}$) distribution.

2.A.4 Method 4: Information from last year's audit

Suppose the auditor has access to last year's posterior Gamma(α_{t-1}, β_{t-1}) distribution, which has its 95 percent upper bound below the performance materiality. Given a value for f, the prior distribution is a Gamma($1 + (\alpha_{t-1} - 1) \times f$, $\beta_{t-1} \times f$) distribution.

2.B R code to reproduce method 5

```r
## For this method it is required that "rstan" is installed
## Use install.packages("rstan") to install the package in R
## For more information about installation of the "rstan" package,
## see https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started

# Generate the benchmark data
S <- 298112312    # Sales of auditee
C <- 223994405    # Cost of sales of auditee

n         <- 100  # Number of data points
minSales <- 1e8
maxSales <- 4e8
ratio    <- 0.74
noise    <- 1.15e7

set.seed(123)
benchmark <- data.frame("sales" = runif(n, min = minSales, max = maxSales))
benchmark[["costofsales"]] <- ratio * benchmark$sales +
                               noise * rnorm(n, mean = 0, sd = 1)

# Create a figure of the benchmark data
xBreaks <- pretty(c(1e8, 4e8), min.n = 4)
yBreaks <- pretty(c(1e8, 4e8), min.n = 4)

# Estimate a linear regression model using the benchmark data
benchmarkRegression <- lm(costofsales ~ sales, data = benchmark)

# Derive the prediction for the auditee's cost of sales
predictedC <- predict(benchmarkRegression,
                      newdata = data.frame(sales = S),
                      interval = "prediction", level = 0.90)

# Derive the relative deviation of the auditee's booked costs of sales
# when compared to the predicted cost of sales
mu <- (C - predictedC[1]) / C

# Derive the standard deviation of the distribution of the residuals
stdev <- sd(benchmarkRegression$residuals) / C

# Load the rstan library
library(rstan)

# Compile the stan model
model <- "
data {
    int<lower=0> n;      //# Number of items in sample
    int<lower=0> k;      //# Number of misstatements in sample
    real mu;             //# Mean of the prior distribution
```

```
    real sigma;         //# Standard deviation of the prior distribution
}
parameters {
    real<lower=0,upper=1> theta;
}
model {
    theta ~ normal(mu, sigma)T[0, 1];
    k ~ binomial(n, theta);
}
"
stanModel <- stan_model(model_code = model)

# The required sample size is the minimum integer n that brings the
# 95th percentile of the posterior distribution,
# given that zero errors are found in the sample, below the performance
# materiality (5 percent).

n <- 50 # Required sample size
k <- 0

# Perform sampling
stanFit <- sampling(stanModel,
                    data = list(n = n, k = k, mu = mu, sigma = stdev),
                    iter = 40000, warmup = 5000,
                    chains = 1, cores = 1)

# Extract samples for theta
samples <- extract(stanFit)
theta <- samples[["theta"]]

# Extract the 95th percentile of the posterior distribution for theta
quantile(theta, probs = 0.95) # 0.04958

# By changing the code "n <- 50" into "n <- 49" and running the code below
# it again you can see the results for n = 49.
# For n = 49, the upper bound equals 5.01 percent and so this sample
# size is not sufficient.
```

Chapter 3

Incorporating Audit Information into the Statistical Model

Abstract

The audit environment of today offers a wealth of information in the form of data. Consequently, data about the auditee is expected to guide and improve auditors' approach to tests of details. However, to be able to make optimal use of this data, auditors must have tools that facilitate the effective and efficient use of quantitative information throughout an audit. In this chapter, we introduce Bayesian generalized linear modeling as a statistical framework to incorporate this information into tests of details, thereby enabling auditors to deliver a fine-grained and specifically tailored audit opinion to stakeholders. We begin with an introduction of Bayesian inference in audit sampling, then explain the main concepts underpinning Bayesian generalized linear modeling and show how this approach allows auditors to bridge the gap between analytics on integrally available data and analytics on data that is available on a sample basis, making optimal use of their information.

Keywords: Analytical procedures, audit sampling, Bayesian, statistical model, stratification.

3.1 Introduction

Today's auditors stand at a precipice. The impact of the data revolution (Kitchin, 2014) is gradually permeating audit theory and practice, and as a result, most auditors have some understanding of the possibilities of using (big) data analytics

Parts of this chapter are published (in dutch) as Derks, K, de Swart, J. & Wetzels, R. (2022). Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor. *Maandblad voor Accountancy en Bedrijfseconomie*, *96*(1/2), 37–46. `https://doi.org/10.5117/mab.96.78836` and under review for publication as Derks, K., de Swart, J., Wagenmakers, E.–J., & Wetzels, R. (2022). How to Incorporate Audit Information into the Statistical Model: Bayesian Generalized Linear Modeling for Audit Sampling.

in their audit activities. While methods of data analysis are slowly being adopted in audit practice (Gepp et al., 2018), the heightened interest in these methods bears the anticipation that they will increase auditors' knowledge of the auditee and, consequently, improve their ability to collect sufficient and appropriate audit evidence (Yoon et al., 2015). For this reason, audit researchers (e.g., Cao et al., 2015), audit firms (e.g., Deloitte, 2021), and audit standard setters (e.g., Public Company Accounting Oversight Board (PCAOB), 2017; Koninklijke Nederlandse Beroepsorganisatie van Accountants (NBA), 2017) have expressed a desire to use methods of data analysis as a means to guide and inform tests of details. However, to do so optimally, auditors must have tools that allow them to use quantitative information throughout an audit. Here, we introduce Bayesian generalized linear modeling as a statistical framework for incorporating such information into tests of details. The aim of this chapter is to demonstrate how this framework can help auditors align their sample evaluation with the situation in practice and, as a result, provide an audit opinion that is specifically tailored to the audit and the auditee. Because, to the best of our knowledge, Bayesian generalized linear modeling has not been addressed in the auditing literature before, the main contribution of this chapter is to describe how it can be used in an audit environment and what the main benefits are for auditors in practice.

While much research discusses how analytical procedures can be applied to obtain audit information (Brown-Liburd et al., 2015; Yoon et al., 2015), an underrepresented area of research is the integration of this information into later stages of the audit (e.g., tests of details). Certainly, advances in data analytics now allow auditors to inspect some populations completely without the need for tests of details (Appelbaum et al., 2017; Huang et al., 2022). However, if the should-be position ("soll") to compare the data to be audited ("ist") against cannot easily be made integrally available in electronic form, tests of details remain the primary method of obtaining reasonable assurance about the misstatement of a population. For example, the auditor can use an analytical procedure to cross-check payments for consistency with payment orders, but then subsequently also needs to verify the validity of the payment orders using tests of details. Another example is a situation where the auditor must form an opinion on a large number of items that are not digitally recorded. For such populations that require much audit effort, analytical procedures will not entirely replace but instead inform tests of details (Cockcroft and Russell, 2018; Yoon and Pearce, 2021).

A classic example is when an auditor performs an analytical procedure to determine the risk of material misstatement for a population as "low" (Blokdijk, 2004). Auditing standards prescribe that this pre-existing information be incorporated into the auditor's statistical approach to tests of details, commonly referred to as audit sampling, to reduce the amount of work that needs to be performed (ISA 530, International Auditing and Assurance Standards Board (IAASB), 2018). The aim is to ensure that the information is used to its full potential. Pre-existing audit information can be integrated into audit sampling in one of two ways: qualitatively or quantitatively. In a qualitative approach, the pre-existing information enters into the statistical analysis via the audit risk (Stewart, 2012). For example, the aforementioned risk assessment enables the auditor to reduce the required audit risk for the population, which consequently leads to less audit work that needs to be performed to reduce this risk to an acceptably low level. Unfortunately, this approach is rigid because it does not allow auditors to differentiate between auditees based on information other than audit risk. In practice, there may be additional information that cannot be captured by adjusting this risk, resulting in some information being discarded in this approach. Consequently, the qualitative approach results in a suboptimal way of integrating the result of analytical procedures with that of tests of details. In a quantitative approach, all information (including integrally available data) enters the statistical analysis through the statistical model. This approach is more flexible than a qualitative approach because it enables auditors to develop a statistical model that aligns with the situation in practice. Since all available data is integrated into one estimate of the population misstatement, the quantitative approach exploits all data to its full potential.

By incorporating pre-existing audit information into the statistical model, auditors can provide an opinion that is tailored to the audit and the auditee. This comes with two practical benefits. First, it improves their ability to form an opinion about the financial statement misstatement because they can statistically explain the impact of the incorporated information on the misstatement. This enables auditors to explain to stakeholders in a transparent manner what characterizes misstatements in the population. Having such information enables the identification of problems at the highest possible organizational level (de Swart et al., 2013), as well as providing management with relevant insight into potential improvements. Suppose an auditor obtained information on the aging of open invoices. If the auditor discovers that the booking delay on an open invoice partly predicts whether the invoice is misstated, the auditor can relay this information to the auditee, who can then act on it. Both parties benefit from this increase in knowledge. Second, incorporating audit information into the statistical model for audit sampling improves auditors' ability to detect misstatements because they can more precisely determine which invoices in the population are likely to be misstated. As more data is used to differentiate the invoices, more fine-grained decisions about (parts of) the population can be made. For example, incorporating the booking delay into the statistical model allows the auditor to statistically differentiate between invoices in the population. They can then focus their remaining audit efforts on invoices that are most likely to be misstated. Despite these two considerable advantages, the quantitative approach is not commonly used in auditing practice. This is because incorporating information into the statistical model can be difficult and time-consuming to justify. Consequently, statistical audit software such as ACL (Dilligent, 2022) and IDEA (CaseWare Analytics, 2022) mainly focuses on a statistical model that estimates the misstatement in a population based on the number of misstatements encountered in a sample alone. To overcome the hurdles of incorporating more information when drawing quantitative conclusions about the misstatement in a population, we will walk through the steps and considerations involved in developing more complex statistical models.

In this chapter, we discuss statistical models in the context of Bayesian statistics. The Bayesian statistical framework, as an alternative to the currently dominant frequentist statistical framework (Stewart, 2012), can have considerable advantages for auditors (van Batenburg et al., 1994; Steele, 1992; Stewart, 2013). For example, Bayesian statistics allows for the inclusion of pre-existing information about the misstatement in the sample evaluation (Derks et al., 2021a), and it facilitates the continuous monitoring of evidence over time (Rouder, 2014; Wagenmakers et al., 2008). Furthermore, the Bayesian approach is advocated as a good fit for auditing practice because it is relatively simple to understand for auditors and stakeholders of the audit (Stewart, 2013; Derks et al., 2022a).

This chapter is structured as follows. Section 3.2 provides a brief overview of the building blocks of Bayesian inference in the context of audit sampling. Next, Section 3.3 introduces the reader to the fundamental concepts underpinning Bayesian generalized linear modeling. Using a practical example, we demonstrate how this approach enables auditors to incorporate multiple sources of information into the statistical model. The following section discusses the sensitivity and robustness of the proposed models to the prior distribution. Section 3.5 delves deeper into two practical cases to demonstrate the advantages of Bayesian generalized linear modeling over a more traditional approach. Section 3.6 contains guidelines and suggestions for putting this technique into practice. In the final part, we provide our concluding comments.

3.2 The Bayesian approach to audit sampling

When evaluating an audit sample, the goal of the auditor is to make a statement about a particular characteristic, θ, of a population of items. Generally, in a financial audit θ represents the probability of misstatement, that is, the probability that an item or a monetary unit (e.g., a dollar) from the population is misstated. Since the auditor does not inspect the entire population but only the data coming from a sample of this population, y, the information from the sample has to be extrapolated to the entire population. Hence, the auditor ought to acknowledge uncertainty and make a probabilistic statement about θ.

The Bayesian way of making a probabilistic statement about the parameter θ, given the data y, is through the posterior distribution $p(\theta \mid y)$. The posterior distribution is defined by Bayes' theorem as proportional to the product of two distributions, the prior distribution $p(\theta)$ and the likelihood function $l(y \mid \theta)$, given the sample data y:

$$\underbrace{p(\theta \mid y)}_{\text{Posterior}} \propto \underbrace{l(y \mid \theta)}_{\text{Likelihood}} \times \underbrace{p(\theta)}_{\text{Prior}}. \tag{3.2.1}$$

The symbol $\propto$ indicates that the expression to the right of this sign is equal to the probability distribution to the left of this sign except for scaling. As Equation 3.2.1 illustrates, a Bayesian model has three fundamental components: the prior distribution, the likelihood, and the posterior distribution. In the following subsections, we will briefly discuss these three components of a Bayesian model.

3.2.1 The prior distribution

The prior distribution $p(\theta)$ reflects the auditor's pre-existing information about the probability of misstatement before any information from the sample has been seen.

The prior distribution is a probability distribution, which means that an adequate prior distribution assigns a relative plausibility to each possible value of θ such that the sum of the probabilities of every possible value of θ equals one. Note that because the prior distribution contains the auditor's pre-existing information, it should also be substantiated as such. For example, because the aforementioned risk assessments provide information about the probability of misstatement in the population, they can be incorporated as information into the prior distribution (Derks et al., 2021a, pp. 628–629).

A commonly used prior distribution is a uniform Beta(1, 1) distribution for θ (Figure 3.1, dashed line), implying that every value of the population misstatement is equally likely a priori. There are two reasons why this uniform prior distribution is often used in practice. First, it is easy to understand; second, with this prior, the Bayesian approach produces nearly identical results as a frequentist audit sampling procedure (Derks et al., 2021a). Therefore, we will use the Beta(1, 1) distribution as a prior for θ in the remainder of this section. However, it is important to note that this prior distribution lacks explicit information about the population misstatement (Derks et al., 2021a, pp. 626–627).

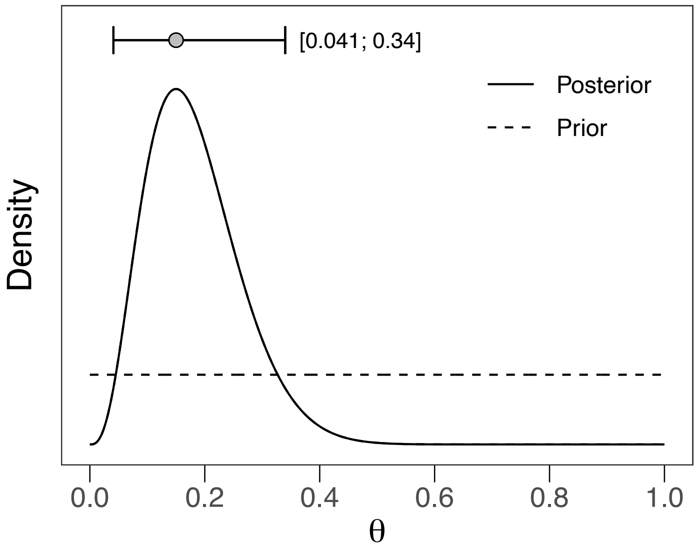


Figure 3.1: Illustration of the Beta(1, 1) prior distribution and corresponding Beta(4, 18) posterior distribution for after seeing a sample of $n = 20$ items containing $k = 3$ misstatements. The posterior mode and 95 percent highest density interval are provided on top. Since the Beta(1, 1) prior is uniformly distributed, the posterior distribution also equals the likelihood function.

3.2.2 The likelihood

The likelihood represents the information that the observed sample data y contains about the population misstatement θ. The likelihood function $l(y \mid \theta)$ quantifies the probability that the sample outcomes will occur under specific values of θ (Etz, 2018). Hence, the role of the likelihood function is to describe the support in the data for the possible values of the parameter θ.

In audit sampling, the probability of observing k misstatements in a sample of n items, with an underlying parameter θ, is binomially distributed and is typically denoted as $\text{Binomial}(k \mid n, \theta)$ (Stewart, 2012). Suppose the auditor has inspected a sample of $n = 20$ items and discovered that $k = 3$ items contain a misstatement. Assuming the binomial likelihood implies that the data from this sample is distributed as $\text{Binomial}(k = 3 \mid n = 20, \theta)$, see Figure 3.1. The value of θ that maximizes the likelihood function is called the maximum likelihood estimate (Myung, 2003), which in this example is equal to $\frac{3}{20} = 0.15$.

In a frequentist audit sampling procedure, the likelihood function is the sole source of information for the auditor's inferences about θ in the population. In a Bayesian audit sampling procedure, the information in the likelihood function is combined with the information in the prior distribution to form the posterior distribution. As we will demonstrate in this chapter, the likelihood also accommodates incorporating more information than just the observed number of misstatements in the sample.

3.2.3 The posterior distribution

Using Bayes' theorem, the auditor can combine the information in the prior distribution with the information in the likelihood to form the posterior distribution. Hence, the posterior distribution $p(\theta \mid y)$ contains the auditor's updated knowledge about the probability of misstatement θ after observing the sample data y. Conceptually, Bayes' theorem states that values of θ that predict the sample data relatively well become more likely than they were under the prior distribution, while values of θ that predict the sample outcomes relatively poorly become less likely. Since the posterior distribution follows from the prior distribution and the likelihood, it is the sole component in the model that cannot be modified by the auditor. Nonetheless, the posterior distribution of θ is the main source of information for a Bayesian auditor because it contains all the information the auditor had prior to sampling and all the information obtained from the sample.

Because the posterior distribution, like the prior distribution, is a probability distribution, the updated knowledge about the population misstatement θ can be summarized by statistics such as the mean, median, and mode of the posterior distribution. For example, the auditor can make a statement about the most likely misstatement in the population by determining the mode of the posterior distribution (i.e., the value of with the highest probability). Similarly, percentiles of the posterior distribution can be interpreted in terms of credibility (Kruschke and Liddell, 2018). For example, the 95 percent highest posterior density (HPD; Kruschke, 2015, pp. 87–89) interval contains the 95 percent most likely values of θ under the posterior distribution.

The posterior distribution using a Beta(1, 1) prior after seeing the binomially distributed sample of $n = 20$ items containing $k = 3$ misstatements is a Beta($1 + k = 4$, $1 + n - k = 18$) distribution (Figure 3.1, solid line). The mode of this posterior distribution is $\frac{k}{n} = 0.15$ (Myung, 2003), which means that the most likely probability of misstatement in the population is estimated to be 15 percent. The 95 percent HPD interval for the posterior distribution is [0.041; 0.340]. This means that, with a 95 percent probability, the probability of misstatement in the population is estimated to be between 4.1 percent and 34 percent.

To summarize, the Bayesian approach to audit sampling entails specifying a prior distribution for the probability of misstatement, updating the prior distribution with information provided by the sample data using Bayes' theorem, and using the posterior distribution to perform inference about the misstatement in the population. One considerable advantage of the Bayesian approach over the frequentist approach is that it allows the auditor to incorporate pre-existing information about the probability of misstatement into the statistical analysis via the prior distribution. However, the Bayesian approach to audit sampling discussed in this section is limited because inferences about θ are based on a single characteristic of the sample data, that is, whether an item is misstated or not. The following section introduces Bayesian generalized linear modeling as a method for incorporating multiple source of data into the auditor's approach to audit sampling.

3.3 Bayesian generalized linear modeling

This section explains how Bayesian generalized linear modeling (Dey et al., 2000; Gelman et al., 2013, Chapter 16) can be applied in the context of audit sampling. In contrast to the basic approach described in the preceding section, in which inferences about the population are based on a single characteristic of the sample data, a Bayesian generalized linear model can be applied if the auditor wants to use multiple sources of information to estimate the probability of misstatement. Assume an auditor has obtained data concerning the control effectiveness that applies to an item or the booking delay of an item. This data is typically ignored during the sample evaluation. However, the auditor has the option to make use of this data by expanding the statistical model with additional parameters. By doing so, the auditor can tailor their statistical model and sample evaluation to the audit and the auditee, resulting in a more fine-grained opinion. Expanding the model has the additional advantage that pre-existing information about the impact of the incorporated data on the probability of misstatement can be integrated into the prior distribution for the new parameters. There is, of course, a trade-off: While a Bayesian generalized linear model enables auditors to incorporate many types of integrally available data into the statistical model, it is also more difficult to set up. At the same time, even if the added information does not increase the accuracy of the estimate of misstatement, it does not hurt too much since the cost of retrieving extra data is getting lower and the output of the statistical model will reveal when data has been incorporated in vain. To aid the effective application of this approach in auditing practice, we describe how to construct and evaluate a Bayesian generalized linear model for audit sampling step by step.

Consider the common scenario in which an auditor is tasked with obtaining reasonable assurance about a population of items representing accounts receivable not being materially misstated. To illustrate this example, consider the fictitious data set in Table 3.1, which contains $N = 3000$ items originating at $S = 3$ branches of the auditee. Assume that, prior to tests of details, the auditor performed an analytical procedure to obtain data about the population. In this instance, the auditor used data mining on logging data to determine how long it took (in minutes) for each item in the population to be approved by an administrator after logging in. The auditor wants to use this time spent on controls as a proxy for control effectiveness. Hence, for each item i in the population, two characteristics are known: the control intensity of the item (denoted x_i in Table 3.1) and the branch of the auditee from which the item originated (denoted s_i). Assume that the auditor has taken a sample of $n = 20$ random items from the population and discovered that $k = 3$ items contain a misstatement (column k_i in Table 3.1). For simplicity, assume that both the control intensity and the number of items per branch in the table are representative of those in the population. The auditor wants to estimate the probability of misstatement in the population using the data from this sample.

An intuitive and efficient way for the auditor to estimate the probability of misstatement in the population is to divide the number of misstatements (3) by the number of items in the sample (20) to arrive at the proportion of misstatement in the sample. In the absence of any additional information about the items in the population, the sample proportion of $\frac{3}{20} = 0.15$ is the auditor's most logical guess of the probability of misstatement in the population. Now, suppose that the auditor is not only interested in estimating the probability of misstatement in the population but also in estimating the probability of misstatement for each item that was not included in their sample. Then, for each of these items, the sample proportion of 0.15 is also the auditor's most logical estimate of the probability of misstatement. Although intuitive, this approach is not ideal because it does not explicitly state how the misstatements in the sample (i.e., the data) are related to the probability of misstatement in the population (i.e., the parameter). This means that the auditor cannot quantify the uncertainty associated with their estimates. To quantify this uncertainty, the auditor must explicitly state their approach to estimation by means of a statistical model.

3.3.1 The basic statistical model for audit sampling

A statistical model defines a functional relationship $f(k \mid \theta)$ between data k and a parameter θ. For instance, the data k can represent measurements of the misstatement in a sample of items, while the parameter θ represents the probability of misstatement in the population. In audit sampling, the relationship f is typically assumed to be a Bernoulli distribution that accounts for the uncertainty due to only measuring a small sample of a larger population (Gelman and Hill, 2007). To make this concrete, the statistical model $k_i \sim \text{Bernoulli}(\theta)$, where the symbol $\sim$ means "is a stochastic variable distributed as", captures the auditor's intuitive approach (described in the preceding paragraph) to estimating the probability of

Table 3.1: Fictitious data set containing a population of $N = 3000$ items from which a random sample of $n = 20$ items containing $k = 3$ misstatements has been inspected. The left-hand columns display the item index (i), control intensity (x_i), branch index (s_i), and misstatement outcome (k_i). The remaining columns display the estimated probability of misstatement for each item, θ_i, (with 95 percent HPD interval in brackets) based on six models. For clarity, the estimates for the in-sample items are omitted, but they are used to calculate the population estimate (bottom row).

				Estimated θ_i					
i	x_i	s_i	k_i	No covariate No differences	No covariate No similarities	No covariate Multilevel	With covariate No differences	With covariate No similarities	With covariate Multilevel
1	0	1	1						
2	0	1	1						
3	0	1	-	0.150 [0.041; 0.340]	0.333 [0.081; 0.685]	0.156 [0.029; 0.505]	0.207 [0.046; 0.514]	0.436 [0.116; 0.827]	0.230 [0.046; 0.659]
4	0	2	1						
5	0	2	0						
6	1	1	0						
7	1	2	0						
8	1	3	0						
9	1	3	0						
10	1	3	0						
11	2	1	0						
12	2	2	0						
13	2	3	0						
14	3	2	0						
15	3	2	0						
16	3	3	0						
17	3	3	0						
18	4	1	0						
19	4	1	-	0.150 [0.041; 0.340]	0.333 [0.081; 0.685]	0.156 [0.029; 0.505]	0.060 [0.008; 0.272]	0.198 [0.025; 0.615]	0.066 [0.007; 0.401]
20	4	1	0						
21	4	2	0						
22	4	2	-	0.150 [0.041; 0.340]	0.143 [0.011; 0.476]	0.127 [0.011; 0.368]	0.060 [0.008; 0.272]	0.063 [0.003; 0.411]	0.050 [0.001; 0.294]
23	4	3	-	0.150 [0.041; 0.340]	0.000 [0.000; 0.312]	0.086 [0.000; 0.308]	0.060 [0.008; 0.272]	0.014 [0.000; 0.260]	0.025 [0.000; 0.243]
24	5	2	-	0.150 [0.041; 0.340]	0.143 [0.011; 0.476]	0.127 [0.011; 0.368]	0.038 [0.003; 0.268]	0.042 [0.001; 0.409]	0.028 [0.000; 0.287]
25	5	2	-	0.150 [0.041; 0.340]	0.143 [0.011; 0.476]	0.127 [0.011; 0.368]	0.038 [0.003; 0.268]	0.042 [0.001; 0.409]	0.028 [0.000; 0.287]
26	5	2	-	0.150 [0.041; 0.340]	0.143 [0.011; 0.476]	0.127 [0.011; 0.368]	0.038 [0.003; 0.268]	0.042 [0.001; 0.409]	0.028 [0.000; 0.287]
27	5	3	-	0.150 [0.041; 0.340]	0.000 [0.000; 0.312]	0.086 [0.000; 0.308]	0.038 [0.003; 0.268]	0.011 [0.000; 0.247]	0.016 [0.000; 0.234]
28	5	3	-	0.150 [0.041; 0.340]	0.000 [0.000; 0.312]	0.086 [0.000; 0.308]	0.038 [0.003; 0.268]	0.011 [0.000; 0.247]	0.016 [0.000; 0.234]
29	5	3	0						
…	…	…	…	…	…	…	…	…	…
3000	5	3	-	0.150 [0.041; 0.340]	0.000 [0.000; 0.312]	0.086 [0.000; 0.308]	0.038 [0.003; 0.268]	0.011 [0.000; 0.247]	0.016 [0.000; 0.234]
Pop				0.150 [0.041; 0.340]	0.221 [0.084; 0.366]	0.149 [0.040; 0.331]	0.142 [0.035; 0.311]	0.197 [0.086; 0.351]	0.138 [0.037; 0.302]

Note. The number 1 in the column k_i implies that the item is misstated. The number 0 implies that the item does not contain a misstatement. The '-' sign implies that an item was not included in the sample.

misstatement in the population. Note that, because the aggregated data of n independent Bernoulli(θ) distributed observations form a Binomial($k \mid n, \theta$) distribution, this approach is equal to that described in Section 3.2, with the exception that the data in Section 3.2 are aggregated before being analyzed. In Section 3.5, we will discuss examples of statistical models formulated at a higher level than the item level i.

This basic statistical model explains the data k with a single parameter θ, which can be interpreted as the probability of misstatement in the population. In other words, this model assumes that the misstatement k_i in each item i is a function of the probability of misstatement in the population. It is worth noting that this basic model is widely used in auditing (American Institute of Certified Public Accountants (AICPA), 2019; Stewart, 2012). However, a more flexible way to specify this model is to reformulate the problem of estimating θ in terms of log-odds via a transformed parameter $\xi = \ln(\frac{\theta}{1-\theta}) = \text{logit}(\theta)$, as follows: $k_i \sim \text{Bernoulli}(\text{logit}^{-1}(\xi))$. Following the conventions for graphical representations of statistical models used in Lee and Wagenmakers (2013, p. 18), we have depicted this basic statistical model in Figure 3.2. This transformation of the model parameters is the key idea behind a generalized linear model. We refer to Appendix 3.A for a more detailed discussion of the log-odds transformation, but we want to emphasize that this model is equivalent to the previous model, with the exception that it can be expanded more easily at a later point. For explanatory purposes, the R (R Core Team, 2022) code for fitting this model and the other models in this chapter to the data in Table 3.1 can be found in Appendix 3.B.

$$k_i \sim \text{Bernoulli}(\theta)$$
$$\theta \leftarrow \text{logit}^{-1}(\xi)$$
$$\xi \sim \text{Logistic}(0, 1)$$

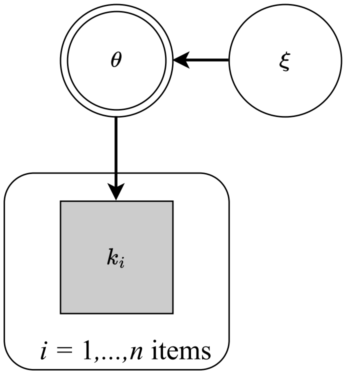


Figure 3.2: Graphical representation of the basic statistical model to estimate the probability of misstatement θ in a population. The single white circle represents an unobserved continuous parameter in the model and the gray square represents observed discrete data. The double white circle represents θ, which is deterministically determined (unobserved and continuous) once ξ is known. Since the observed misstatements depend on the parameter θ, the arrow is directed towards k.

After specifying a uniform prior distribution transformed to the log-odds scale (i.e., if the prior on θ is Uniform(0, 1), a Logistic(0, 1) prior on $\xi = \ln(\frac{\theta}{1-\theta})$ is induced (Lin and Hu, 2008, p. 1148)) and collecting data, the auditor can compute the posterior distribution for $\theta = \text{logit}^{-1}(\xi)$. This allows the auditor to do two things. First, they can make statistical inferences about the probability of misstatement in the population. For instance, the auditor can determine the most likely value of the posterior distribution of θ and quantify the uncertainty about θ using the HPD interval. To put this into context, the most likely probability of misstatement in the population is 15 percent (see the bottom row of Table 3.1). The 95 percent HPD interval around this estimate ranges from 4.1 percent to 34 percent. Second, given the posterior distribution of θ, the auditor can estimate the probability of misstatement for each unseen item in the population, θ_i, based on the characteristics of that item. To illustrate, the estimated probability of misstatement for each unseen item in the population is shown in the fifth column of Table 3.1. Item 3, for instance, is estimated to have a probability of misstatement of 15 percent, with a 95 percent HPD lower bound of 4.1 percent and an upper bound of 34 percent.

Every statistical model is a simplification of reality. Consequently, the output of a statistical model is predicated on the assumption that the model accurately describes the situation in practice. In other words, the auditor's choice of statistical model will shape their conclusions. To demonstrate this, we revisit the estimates of the basic model in Table 3.1. Because this model assumes a single parameter that explains the misstatement of all items in the population, it does not distinguish between specific items in the population, and thus the estimated probability of misstatement is the same across all items. To make this concrete, all items in the population have an estimated probability of misstatement of 15 percent, with a 95 percent HPD lower bound of 4.1 percent and an upper bound of 34 percent. Such output forces the auditor to conclude that all items in the population have the same probability of being misstated, even though the items may vastly differ in their vulnerability to misstatement.

With this in mind, the basic statistical model accurately describes the situation in practice under one condition: The auditor's only source of data is the misstatement in the sample items. However, if multiple sources of data about the (probability of) misstatement are available, this model is less appropriate and even restrictive. In our running example, for instance, it is possible that items that are quickly approved by an administrator are more likely to contain a misstatement than items that took a longer time to approve, or that items from a specific branch are more likely to be misstated than items from other branches. In this case, the statistical model should ideally take the additional information into account to better align with the situation in practice. In other words, because the statistical model shapes the auditor's conclusions, it is in their best interest to incorporate relevant information in the form of data——also known as covariates——into the statistical model. By including the logit transformation in the model, the covariates can be added and processed in a statistically sound way.

3.3.2 Incorporating covariates into the statistical model

Covariates are additional data that the auditor believes contains information about the (probability of) misstatement in the items. Typically, this data can be continuous (e.g., the length of the time period an invoice is outstanding) or categorical (e.g., the branch of the auditee from which an item originates, or a risk category assigned to an item). In the following subsections, we will cover how to incorporate continuous and categorical covariates into the statistical model for audit sampling.

3.3.2.1 Incorporating continuous covariates into the statistical model

In this subsection, we demonstrate how continuous covariates can be incorporated into the statistical model for audit sampling. Continuous covariates are data that can take on any numerical value. In an audit context, this could be the number of days a payment was (or is) outstanding, the total sales revenue per branch of the auditee, or the number of FTEs who can modify a payment through the internal system. In our running example, the auditor used data mining to perform an analytical procedure to determine the control intensity that applies to each item. Suppose that the auditor discovers that some items in the population are subject to higher control intensity than others, for example, because items that administrators are less familiar with are checked more thoroughly. The statistical model can express the relationship between the misstatements, the control intensity x_i and the probability of misstatement.

To incorporate the control intensity as a covariate in the statistical model, a new model term that quantifies the relationship between the misstatement and the control intensity is added. In practice, this means that the basic model is expanded with an additional parameter β_x and the new data x_i as follows: $k_i \sim \text{Bernoulli}(\text{logit}^{-1}(\beta_0 + \beta_x \cdot x_i))$, see Figure 3.3. In other words, the expanded model assumes that the misstatement k_i of each item i is a function of the probability of misstatement in the population and the control intensity x_i of that item. Note that, because this is a Bayesian analysis, all parameters that are to be estimated must have a prior distribution, which we will cover in Section 3.4.

$$k_i \sim \text{Bernoulli}(\theta_i)$$
$$\theta_i \leftarrow \text{logit}^{-1}(\xi_i)$$
$$\xi_i \leftarrow \beta_0 + \beta_x \cdot x_i$$
$$\beta_0 \sim \text{Logistic}(0, 1)$$
$$\beta_x \sim \text{Normal}(0, 1)$$
$$\theta \leftarrow \sum_{i=1}^{N} \frac{\theta_i}{N}$$

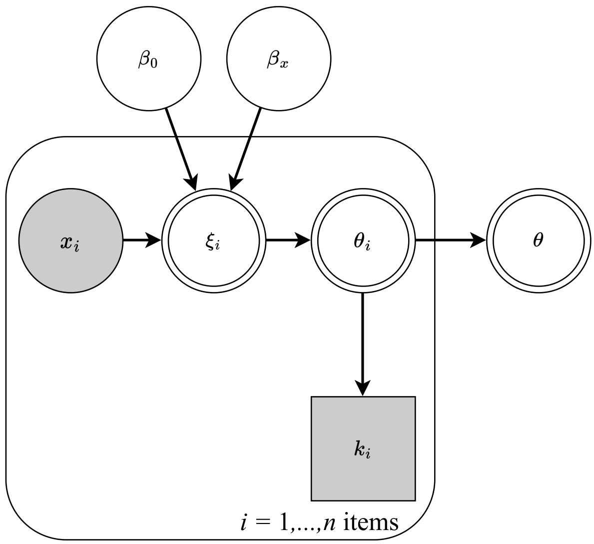


Figure 3.3: Graphical representation of a statistical model to estimate the probability of misstatement θ in a population while taking the control intensity x into account. The gray circle represents observed continuous data.

When the control intensity is fed into the statistical model as an input, the parameter β_0 represents the log-odds of the probability of misstatement in the population for items with zero control intensity. It should be noted that items with zero control intensity may not exist in the data, which means that the interpretation of β_0 is potentially meaningless. For this reason, it is recommended to standardize continuous covariates in the model. We follow the conventions for standardization and transform the covariates to have a mean of zero and a standard deviation of $\frac{1}{2}$ (for a motivation of this choice, see Gelman (2008)). By standardizing the covariates, the parameter β_0 represents the log-odds of misstatement in the population for items with an average control intensity, providing a more meaningful interpretation than before the covariates were standardized.

Next, the parameter β_x can be interpreted as the deviation from β_0 as a function of the standardized control intensity. For example, if $\beta_x = -1.2$, an increase by one in the standardized control intensity is associated with a multiplication of the log-odds of misstatement by -1.2 and, equivalently, with a multiplication of the odds of misstatement by $e^{-1.2} = 0.30$. In other words, if β_x is negative, the probability of misstatement is higher for items with low control intensity than for items with high control intensity, and vice versa. While not the main focus of this chapter, it is worth noting that the statistical model is not limited to a single β parameter and can be expanded if more continuous covariates are available (Gelman and Hill, 2007, pp. 32—34).

In comparison to the basic model, the auditor gains two benefits by including the control intensity as a covariate. First, the auditor can explain the impact of this integrated information on the misstatement in a statistical manner because the expanded model statistically estimates the relationship between the control intensity and the probability of misstatement. This, in turn, can provide the auditee with useful information. For example, if the auditor has convincing evidence that the probability of misstatement is high for items with low control intensity, this information can be communicated to the auditee, who can then, for example, improve internal control by increasing the minimal control intensity. Second, unlike the basic statistical model, which assumes that the probability of misstatement is the same across all items in the population, this statistical model allows the auditor to estimate the probability of misstatement of an item based on the control intensity of that item. To illustrate, the estimated probability of misstatement for the unseen items in the population under the expanded model is shown in the eight column of Table 3.1. Because this model includes information to differentiate items in the population based on control intensity, the estimated probability of misstatement for each item varies depending on the control intensity of that item. For example, item 3 has zero minutes between login and approval and thus has a relatively high estimated probability of misstatement when compared to item 19, which has four minutes between login and approval and thus has a relatively low estimated probability of misstatement. Because auditors can make this distinction, they can more precisely identify potential misstatements in the population and, as a result, make finer-grained decisions about (parts of) the population. Even in the case that the control intensity does not influence the probability of misstatement, the output of the statistical will inform the auditor accordingly.

3.3.2.2 Incorporating categorical covariates into the statistical model

In this subsection, we will look at how to incorporate categorical covariates into the statistical model for audit sampling. Categorical covariates are data whose value can take on one of a limited number of possible values, classifying each item in the population as belonging to one of several unordered groups. In audit practice, parts of the population having the same value for the categorical covariate are typically referred to as strata (Roberts, 1978, p. 95). For instance, in our running example, the auditor has divided the population into three distinct strata based on the origin of each item across the branches of the auditee. In this case, the branch from which an item originates can be viewed as a type of stratum membership.

The data from the branches, like the data from continuous covariates, can be included in the statistical model by introducing a new model term that quantifies the relationship between the misstatements and the stratum membership. The form of this new model term is determined by the auditor's assumptions about how different or similar the probability of misstatement is in each branch. This can be thought of as a spectrum: On one end, the misstatement in the branches could be identical, while on the other end, the misstatement in the branches could be completely unrelated. To best align the new model term with the situation in practice, the auditor must decide where to place their model on this spectrum. In the following sections, we will first look at what the two extreme ends of this spectrum imply for the statistical model. Following that, we demonstrate how to strike a more realistic balance between these two extremes.

3.3.2.2.1 Incorporating no differences between strata

First, assume the auditor has information indicating that the probability of misstatement is likely to be equal across all branches of the auditee. For example, they are aware that all branches of the auditee use the same formula, internal systems, and have rotating staff. Then, a statistical model in which misstatements are not dependent on the originating branch might be used to describe the situation in practice. The fundamental assumption of such a model is that there are no differences between the misstatements in the branches.

The auditor can explicitly state this assumption in the statistical model by including a single β_0 parameter that represents the overall probability of misstatement in the population: $k_{i,s} \sim \text{Bernoulli}(\text{logit}^{-1}(\beta_0 + \beta_x \cdot x_{i,s}))$, see Figure 3.4. Conceptually, this means that the samples from all branches are interchangeable; in other words, the distinction between branches is unnecessary. As a result, the samples can be aggregated across branches and analyzed as a whole. This model is identical to the previous model, except for the addition of the subscript s, which indicates from which branch item i originates.

As previously stated, if the auditor has information indicating that the probability of misstatement is equal across branches of the auditee, this statistical model might be used to describe the situation in practice. However, the assumption underlying this model is at the extreme end of the spectrum, and the statistical results reflect this. By choosing a model with a common probability of misstatement for the population, the auditor is unable to statistically distinguish between items in the population based on the branch from which they originate. Consider items 19 and 23 in Table 3.1. Even though these items originate from different branches of the auditee, the eight column shows that they both have the same estimated probability of misstatement. This statistical model raises the question:

$$
\begin{aligned}
k_{i,s} &\sim \text{Bernoulli}(\theta_{i,s}) \\
\theta_{i,s} &\leftarrow \text{logit}^{-1}(\xi_{i,s}) \\
\xi_{i,s} &\leftarrow \beta_0 + \beta_x \cdot x_{i,s} \\
\beta_0 &\sim \text{Logistic}(0, 1) \\
\beta_x &\sim \text{Normal}(0, 1) \\
\theta &\leftarrow \sum_{s=1}^{S} \sum_{i=1}^{N_s} \frac{\theta_{i,s}}{N}
\end{aligned}
$$

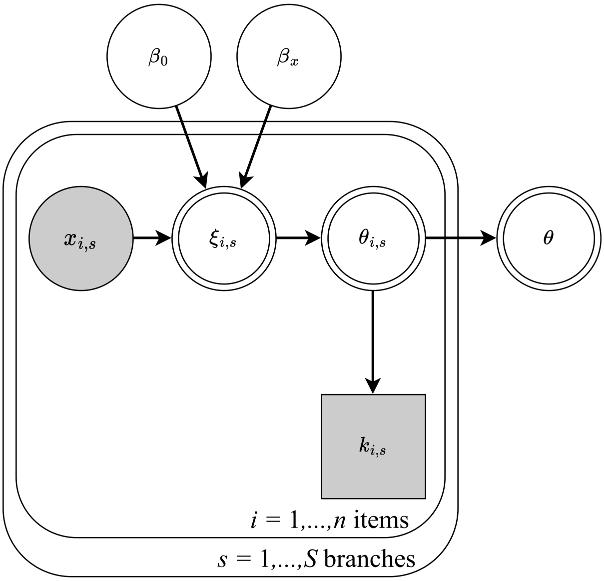


Figure 3.4: Graphical representation of a statistical model to estimate the probability of misstatement θ in a population while taking the control intensity x into account and assuming no differences between branches s.

Why should an auditor incorporate the branch into the statistical model after stating that the branch does not influence the probability of misstatement? In practice, this model may occur for pragmatic reasons. For example, a group auditor, knowing that all operating companies use the same processes and IT systems, may still want to distribute the audit work amongst the auditors of the individual operating companies and use s as an index indicating the operating companies.

3.3.2.2.2 Incorporating no similarities between strata

Second, suppose the auditor has information indicating that the misstatements in the auditee's branches come about in a completely different manner. For example, they are aware that the items in the population are sourced from franchised branches of the auditee, each with their own internal systems and staff. Then, a statistical model with no relationship between the misstatements in the branches might be used to describe the situation in practice. The fundamental assumption of such a model is that there are no similarities between the misstatements in the branches.

The auditor can explicitly state this assumption in the statistical model by including an independent parameter β_0 for each branch s that represents its own probability of misstatement: $k_{i,s} \sim \text{Bernoulli}(\text{logit}^{-1}(\beta_{0,s} + \beta_x \cdot x_{i,s}))$, see the right panel of Figure 3.5. In other words, an item's misstatement is assumed to be a function of the branch's probability of misstatement and the control intensity of that item. Because the $\beta_{0,s}$ parameters are independent, no information can be shared between the branches, and each branch's samples should be analyzed independently. For illustrative purposes, we only introduce independent $\beta_{0,s}$ parameters and assume that the relationship between the control intensity and the misstatement is equal for all items, but this is not required.

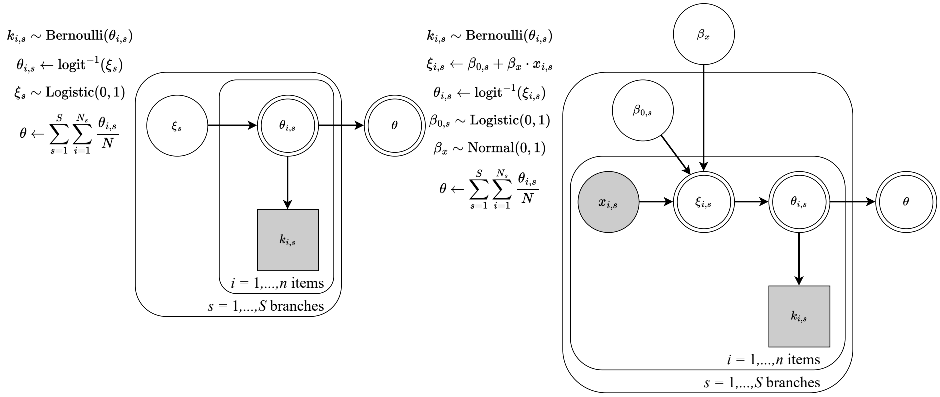


Figure 3.5: Graphical representation of a statistical model to estimate the probability of misstatement θ in a population while assuming no similarities between branches s. In addition to this assumption, the model in the right panel takes into account the control intensity x.

As previously stated, if the auditor has information indicating that the misstatements across branches are completely unrelated, this statistical model might be used to describe the situation in practice. The assumption underlying this model is at the extreme end of the spectrum, and the statistical results reflect this. To illustrate, the estimated probability of misstatement for each unseen item in Table 3.1 is shown in the second to last column. To clearly illustrate the effect of assuming independent strata, Table 3.1 also contains the results of the statistical model without the control intensity displayed in the left panel of Figure 3.5. Unlike the previous approach, this model enables the auditor to distinguish between items in the population based on the originating branch. Consider items 3 and 19 in the population, which both have a higher estimated probability of misstatement under this model than under the previous model because, in addition to having only one control, they also belong to a relatively misstated branch of the auditee. Vice versa, item 23 now has a lower estimated probability of misstatement because it is part of a branch containing relatively little misstatement. However, compared to the previous model, more parameters need to be estimated using the same amount of data, which leads to a reduction in the amount of data available to estimate each parameter and an increase in the uncertainty of the estimates. To illustrate, the uncertainty in the estimate for item 19 is 59 percent under this model, which is considerably higher than in the previous model (26.4 percent).

3.3.2.2.3 Incorporating differences and similarities between strata

Third, assume the auditor has information indicating that the probability of misstatement is likely to be different in all branches of the auditee but that the misstatements in the branches come about in a similar manner. For example, they are aware that items in the population are payments from franchises of the auditee that are different in nature but still processed by a centralized purchasing system. In this case, both the extreme assumptions that the misstatements between the branches are the same and that the misstatements between the branches are completely unrelated are unrealistic. Then, a statistical model that reflects a compromise between these assumptions is a more accurate description of the situation in practice. The fundamental assumption of such a model is that there are differences and similarities between the misstatements in the branches.

The auditor can explicitly state this assumption in the statistical model by introducing a set of parameters with a hierarchical structure: $k_{i,s} \sim \text{Bernoulli}(\text{logit}^{-1}(\mu + \sigma\alpha_s + \beta_x \cdot x_{i,s}))$, see the right panel of Figure 3.6. At the highest level, this hierarchical structure defines a population distribution for the probability of misstatement, parameterized by the hyperparameters μ, which represents the location of this distribution, and σ, which represents the standard deviation of this distribution (i.e., the heterogeneity between the branches). At the lowest level, the hierarchical structure includes a unique parameter α_s for each branch s that represents the branch's standardized estimate of the log-odds probability of misstatement (Betancourt and Girolami, 2015). Because this type of model defines parameters at multiple levels, it is generally referred to as a multilevel model (Gelman and Hill, 2007, p. 251).

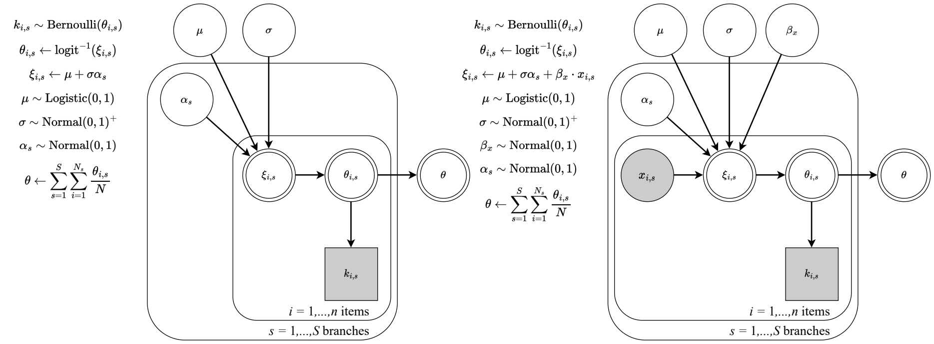


Figure 3.6: Graphical representation of a statistical model to estimate the probability of misstatement θ in a population assuming differences and similarities between branches s. In addition to this assumption, the model in the right panel takes into account the control intensity x.

A multilevel model is a conceptual compromise between the two extreme assumptions of the preceding models. In a multilevel model, the auditee's branches are explicitly related to each other via the population distribution, which means that information can be statistically shared between them. This has the effect of shifting the estimates of the misstatement in the branches towards the population estimate by some amount. This can be demonstrated by revisiting the estimate of item 3 in Table 3.1. To clearly illustrate the effect of the multilevel model, Table 3.1 also contains the results of the statistical model without control intensity, see the left panel of Figure 3.6. In comparison to the high estimate of the previous model, the estimate for this item under the multilevel model is substantially lower and more realistic. This shows that the multilevel model assists in bringing these unrealistic estimates—based on relatively little information—closer to the population estimate. Furthermore, because information in one branch can be combined with information from other branches, there is more information available to estimate each parameter in the model, and therefore the item estimates will become less uncertain (Efron and Morris, 1977). Take, for example, item 19 in the population. The previous model estimates this item with an uncertainty of 59 percent, but the multilevel model estimates this item with an uncertainty of 39.4 percent. Another example of this is item 3000, which under the multilevel model has the least uncertain estimate of all models with an uncertainty of 23.4 percent. This shows that multilevel models introduce some bias into the statistical model, resulting in lower variance in the estimates (Gelman et al., 2013, p. 101; Hastie et al., 2001, Chapter 2). For this reason, these types of models are applied in a variety of scientific fields, including auditing (Laws and O'Hagan, 2002), to account for underlying relationships between groups in the data (Gelman et al., 2013, Chapter 5). Note that, while not the focus of the current chapter, a multilevel model is not restricted to including only two levels and can be expanded when more than two levels are defined (Gelman and Hill, 2007, pp. 32–34).

As previously stated, if the auditor has information indicating that the probability of misstatement is likely to be different across the auditee's branches but that the misstatements come about in a similar manner, this statistical model might be used to describe the situation in practice. The two extreme assumptions, that misstatement across branches is identical and that misstatement between branches is completely unrelated, are often unrealistic. Therefore, incorporating a multilevel structure into the statistical model can strike a balance between these two extremes and help the auditor better align their statistical model with the situation in practice. However, including a multilevel component in the model also complicates the model structure and makes the required calculations more difficult. This means that the auditor must work well with a statistician to ensure the statistical model can be adequately explained and justified.

3.3.3 Consolidating item estimates to a population estimate

So far, all that has been accomplished is a statistical model that, given the characteristics of the items in the population, estimates the probability of misstatement for each item in the population. Before the auditor can use this model to perform inference about the population, they must consider whether the sample is representative of the population with respect to the covariates. In other words, they must determine whether the distribution of covariates in the sample matches that of the population. If this is the case, the auditor can use the model parameters directly to perform inference about the probability of misstatement in the population. However, if this is not the case, and the auditor extrapolates the parameter estimates as is, their estimate of the probability of misstatement in the population will be biased.

In practice, auditors will almost always find themselves in the last situation. The reason for this is that auditors commonly use stratification in their sample evaluation (Hall et al., 2021). Stratification is the division of an audit population into groups (i.e., strata) followed by disproportional sampling of items within the strata (Cochran, 1977, Chapter 5; Roberts, 1978, Chapter 6). In the previous section, we have demonstrated how stratification can be accounted for in the statistical model. However, because stratification involves disproportional sampling, the representation of the strata in the sample will likely differ from their representation in the population. For instance, in the running example, the auditor stratified the items according to which branch they originated from, but instead of selecting samples proportionally, the auditor selected samples equally across the branches.

To arrive at a representative estimate of the probability of misstatement in the total population, the item estimates of the statistical model should be averaged across all items in the population. This is the key idea behind poststratification (Gelman and Little, 1997; Park et al., 2004). In the Bayesian framework, because the auditor has specified a prior distribution, their estimate is not a point estimate but a posterior probability distribution. Averaging these estimates means that the poststratification weight for the posterior distribution of each item in the population is simply $\frac{1}{N}$, where N is the total number of items in the population.

For larger data sets, the estimates for each item can be cumbersome to compute, even with a powerful computer. However, if items can be grouped into subgroups that are similar in all aspects (e.g., items from the same branch with the same control intensity), poststratification can be made less computationally intensive by weighting the estimate for a specific subgroup by its proportion in the population (e.g., our fictitious data set contains 300 items from the first branch with zero minutes between login and approval, and so the posterior distribution for this subgroup is weighed by $\frac{300}{3000} = 0.1$). Thus, in general, the poststratification procedure can be stated as $\theta = \frac{\sum_{s=1}^{S} \theta_s N_s}{\sum_{s=1}^{S} N_s}$, where θ is the poststratified population estimate of the probability of misstatement, s is a running index in the set of S subgroups, θ_s is the estimate of the probability of misstatement for a specific subgroup, and N_s corresponds to the number of items in that subgroup. Table 3.2 shows the poststratification weights for all unique combinations of the originating branch s and the control intensity x in the population from Table 3.1.

The bottom row of Table 3.1 shows the population estimates for the preceding models. Because each population estimate is an average of all item estimates, they are subject to the implications of the model at the item level. First, the most likely population estimate for the model without the covariate is 15 percent, with an associated uncertainty of 29.9 percent. Second, when control intensity is factored into the model that assumes to differences between the branches, the most likely population estimate is 14.2 percent, with an uncertainty of 27.6 percent. In this case, the incorporated data adjusts the population estimate downwards because there are relatively more items in the population with high control intensity than

Table 3.2: Poststratification weights $N_s/\sum N_s$ for the unique combinations in Table 3.1. The value NA implies that there are no items in the population with these characteristics.

	Control intensity (x)					
Branch (s)	0	1	2	3	4	5
1	$\frac{300}{3000}$	$\frac{100}{3000}$	$\frac{100}{3000}$	NA	$\frac{300}{3000}$	NA
2	$\frac{200}{3000}$	$\frac{100}{3000}$	$\frac{100}{3000}$	$\frac{200}{3000}$	$\frac{200}{3000}$	$\frac{300}{3000}$
3	NA	$\frac{300}{3000}$	$\frac{100}{3000}$	$\frac{200}{3000}$	$\frac{100}{3000}$	$\frac{400}{3000}$

in the sample. These items have a low estimated probability of misstatement. Furthermore, because of the incorporated data, the population estimate has also become 2.3 percent less uncertain. Third, when a further distinction between independent strata is made in the model, the population estimate is 19.7 percent, with a 26.5 percent uncertainty. When compared to the previous model, this estimate is higher because the items in the first stratum with low control intensity are predicted to have a relatively high probability of misstatement. Finally, in the multilevel model that takes into account control intensity, the most likely population estimate is 13.8 percent, with an uncertainty of 26.5 percent. Compared to the previous model, the population estimate has been adjusted downwards. The reason for this is that the multilevel approach reduces the variance in the item estimates, which means that the predicted probability of misstatement for items in the first stratum with low control intensity is lower than before. Furthermore, compared to the second model, the population estimate is adjusted downwards because there are relatively more items in the sample from the first branch than there are in the population. On average, these items have a higher estimated probability of misstatement. This example shows how a multilevel model can provide an accurate population estimate while also allowing for fine-grained estimates at the item level.

So far, we have demonstrated how to set up and interpret a Bayesian generalized linear model in the context of audit sampling. The key advantage of this approach is that it enables auditors to incorporate multiple sources of information into the statistical analysis. Hence, by taking this approach, auditors can bridge the gap between analytics on integrally available data and analytics on data that is available on a sample basis and, as a result, derive more knowledge from their samples. However, because we consider these models in the context of Bayesian statistics, a critical step is to specify a prior distribution for each parameter in the model that needs to be estimated. The following section goes over several commonly used prior distributions that can be used in a Bayesian generalized linear model for audit sampling.

3.4 Prior distributions for generalized linear models

In this section, we will look at a few commonly used prior distributions that can be used in a Bayesian generalized linear model for audit sampling. As previously stated, the prior distribution is a probability distribution that incorporates the auditor's pre-existing information about the model parameters. This means that the prior distribution, like the statistical model, is based on assumptions. Once again, there is a trade-off that the auditor needs to consider: Incorporating information into the prior can reduce uncertainty in the estimates because more information is available to estimate the parameters, but it also means that the incorporated information must be properly justified. Hence, it is in the auditor's best interest to consider if there is information available about the parameters in the model and whether or not this information is worthwhile to incorporate in the prior distribution. However, specifying a prior distribution that is aligned with the situation in practice can be difficult. If the auditor does not wish to construct a prior distribution themselves, there are several recommendations from the literature that are easy to understand and justify. For this reason, we recommended a prior distribution for two types of parameters: those related to continuous covariates and those related to the population distribution in a multilevel model. Note that if there is sufficient data and a reasonable prior is applied, the exact choice of prior should have no meaningful influence on the parameter estimates. Nonetheless, we recommend performing a robustness analysis to see if the prior distribution has a substantial impact on the parameter estimates (see Figure 3.7). If this is the case, the auditor should provide arguments in favor of choosing one prior over the other or gather more data to overcome the influence of the prior distribution.

3.4.1 Prior distributions related to continuous covariates

The symbol β typically represents the model parameters related to continuous, standardized covariates (except for β_0). In our example, for instance, the parameter β_x quantifies the relationship between the control intensity x and the misstatement in the data. In a Bayesian model, each of the (potentially many) β parameters to be estimated is assigned a prior distribution.

Prior distributions are typically classified into three types: noninformative prior distributions; weakly informative prior distributions; and informative prior distributions. The idea behind a noninformative prior distribution is that it contains no explicit information about the parameter of interest, in this case β_x. There is considerable debate about what constitutes a noninformative prior (e.g., Kerman, 2011a; Yang and Berger, 1996), but a prior distribution that assumes all possible values of β_x to be equally likely is a uniform prior on the range $[-\infty, \infty]$ (Carpenter et al., 2017). It should be noted that this prior distribution is not a proper probability distribution (i.e., it does not integrate to one) and thus cannot be meaningfully interpreted. However, this can be avoided by restricting the uniform prior distribution to a specific range of plausible values. A weakly informative prior distribution is proper, but it is intentionally designed to contain information that is weaker than the prior information available (Gelman et al., 2013, p. 55). Hence, when there is no explicit information available, a weakly informative prior distribution is often preferable over a noninformative prior. Some recommendations for weakly informative prior distributions for continuous covariates are the Cauchy(0, 2.5) distribution (Gelman et al., 2008), the standard normal distribution (Gao et al., 2021), or the Student–t distribution with degrees of freedom greater than one (Ghosh et al., 2018). The advantage of incorporating little information into the prior distribution is that the prior distribution is relatively simple to understand and justify; the disadvantage is that the estimates come with relatively much uncertainty as a result.

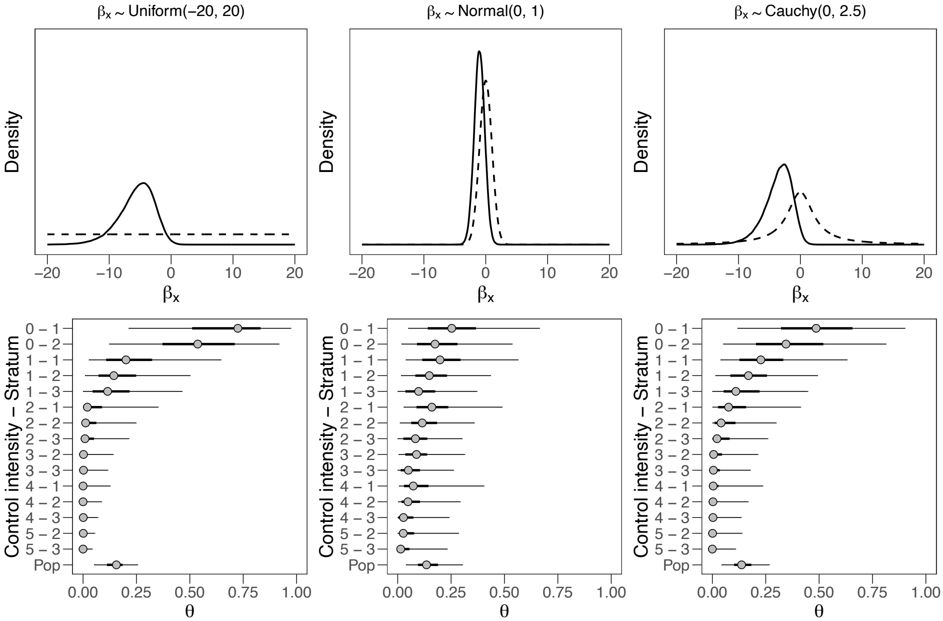


Figure 3.7: The top panels show the posterior distributions (solid lines) of the parameter β_x using three different prior distributions: (a) a Uniform(−20, 20) prior distribution; (b) a Normal(0, 1) prior distribution; and (c) a Cauchy(0, 2.5) prior distribution. The corresponding prior distribution for β_x is provided (dashed lines). The bottom panels show the posterior modes, the 50 percent and the 95 percent HPD intervals of the estimates of θ for the unique combinations of the control intensity (0–5) and stratum (1–3) alongside the population (Pop).

Different prior distributions can, of course, be chosen if the auditor can justify their choice with relevant audit information. For example, if the auditor knows that the relationship between the control intensity and the misstatement must be negative, the prior distribution can be limited to only those reasonable values of the parameter space. However, if the auditor wishes to define an informed prior distribution, it is recommended to perform a robustness analysis, comparing the results of the informed prior distribution to the results of the weakly informative prior distributions presented above. If different prior distributions produce the same estimates, the estimates can be considered robust, and the impact of the prior can be considered negligible. This generally holds true for large sample sizes because the posterior distribution is a compromise between the prior distribution and the data, and this compromise is increasingly controlled by the data as sample size grows (Gelman et al., 2013, p. 32). However, if the prior proves to be critical for the auditor's conclusions, this should be accounted for in the justification of the final prior distribution.

To show how a robustness analysis can be performed on the multilevel model that includes control intensity (the last column in Table 3.1), Figure 3.7 depicts inferences from posterior distributions using three different priors: a Uniform(−20, 20) prior, a Normal(0, 1) prior, and a Cauchy(0, 2.5) prior. When compared to the normal prior (middle column), the simplest approach—a uniform prior restricted to the range [−20, 20]—results in relatively high variance between the item estimates (left column). Furthermore, the Cauchy prior (right column) also results in a relatively high variance when compared to the normal prior. The reason for this effect is that both the uniform and the Cauchy priors assign a large prior probability to values of β_x further away from zero, which induces a bimodal prior distribution for θ_i with the two modes near 0 and 1. Such a prior on the individual items implies an expectation that θ_i is likely to be near 0 or 1. Note that the impact of the prior is relatively large in this example because it is based on a small sample size. However, this impact will decrease as more data becomes available. Nonetheless, despite the rather small sample size of twenty, none of the prior distributions had a substantial impact on the population estimate. Hence, in the remainder of this chapter, we follow Gao et al. (2021) and use a Normal(0, 1) prior distribution for the β parameters in the statistical model related to continuous covariates.

3.4.2 Prior distributions related to categorical covariates

The parameters that determine the prior for the population distribution in a multilevel model are known as hyperparameters (Gelman et al., 2013, p. 101). In our running example, for instance, the hyperparameters are μ and σ, which represent the location and standard deviation of the population distribution for the log-odds of the misstatement probabilities for each level of the categorical covariate, respectively. Using stratified sampling terminology, μ can be interpreted as the mean of the stratum means and σ as the between-stratum standard deviation. All hyperparameters in a Bayesian model that need to be estimated require a prior distribution, which is referred to as a hyperprior. For comparableness, we choose a standard Logistic(0, 1) distribution as a hyperprior for μ because it is the log-odds transformation of a uniform prior (as used in Section 3.2). More information on adequate prior distributions for the probability of misstatement in a population can be found in Derks et al. (2021a). Instead, in this section, we will concentrate on the hyperprior for σ. In practice, the prior distribution for σ influences the degree to which the stratum estimates are all shifted towards the population estimate, a phenomenon known as shrinkage (Gelman and Hill, 2007, p. 477).

Again, three types of hyperpriors for σ are possible: noninformative hyperpriors; weakly informative hyperpriors; and informative hyperpriors. Like before, a prior distribution that assumes all possible values of σ to be equally likely is a uniform prior on the parameter space ranging from zero to some finite maximum (Gelman et al., 2013). Furthermore, Gelman (2006) suggests using a half-Cauchy prior distribution or a half-normal prior distribution as a weakly informative prior for σ. These two weakly informative prior distributions have a relatively high probability mass at zero, indicating that branches are more likely to be similar than different. Hence, they contain some information about the variation between the branches, whereas the uniform prior has no preference for any plausible value of σ. As previously stated, the advantage of using a weakly informative prior is that the prior distribution is relatively easy to understand and justify, with the disadvantage being that there is relatively little information about the relationship between the branches.

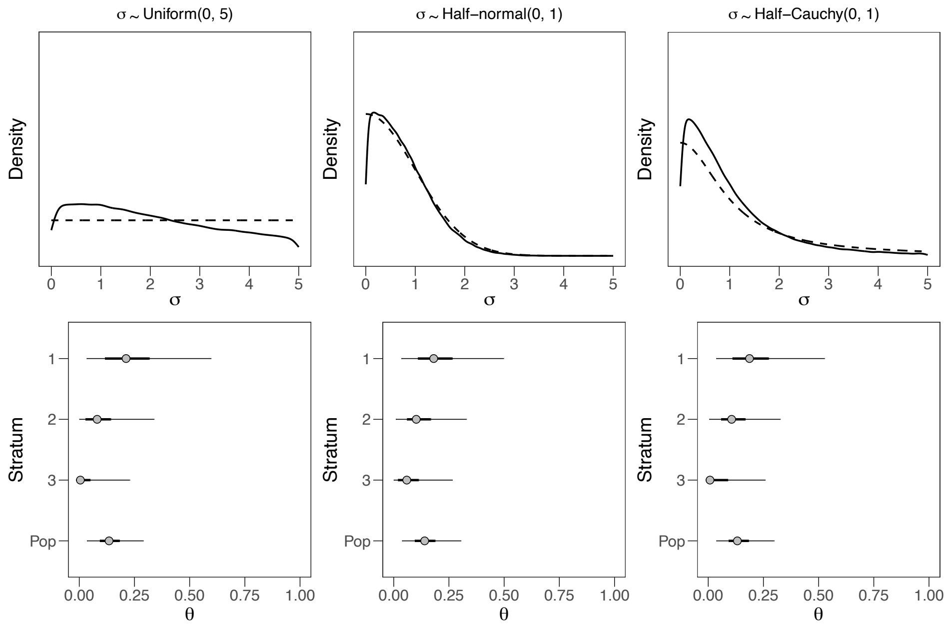


Figure 3.8: The top panels show the posterior distributions (solid lines) of the between-stratum standard deviation σ using three different prior distributions: (a) a Uniform(0, 5) prior distribution; (b) a Half-normal(0, 1) prior distribution; and (c) a Half-Cauchy(0, 1) prior distribution. The corresponding prior distribution for σ is provided (dashed lines). The bottom panels show the posterior modes, 50 percent and 95 percent HPD intervals of the estimates of θ for the unique strata (1–3) and the population (Pop).

Of course, different choices for the prior distribution are possible if the auditor can justify their choice with relevant audit information. For example, if the auditor has information indicating that the branches are relatively similar (e.g., a series of branches with staff rotating across the branches and one central AO/IC and one ERP system), they can specify a hyperprior with a high probability mass at low values of σ. On the other hand, if they have information indicating that the branches are relatively different from each other (e.g., branches in a franchise formula with only central purchasing), they can specify a hyperprior with a high probability mass at large values of σ. Nonetheless, as previously discussed, a robustness analysis is still recommended. If the prior turns out to be critical for the auditor's conclusions, the auditor should be able to justify their choice of prior distribution.

To demonstrate how such a robustness analysis can be performed on the multilevel model that includes control intensity, Figure 3.8 depicts inferences from posterior distributions using three different priors: a Uniform(0, 5) prior; a Half-normal(0, 1) prior; and a Half-Cauchy(0, 1) prior. When compared to the half-normal prior (middle column), the simplest approach—a uniform prior in the range [0, 5]—results in relatively low shrinkage for the branch estimates (left column). The use of a half-Cauchy prior (right column) also results in less shrinkage in the estimates but not in an increase in uncertainty. Furthermore, none of the prior distributions appeared to have an impact on the population estimate. In the following sections, we will therefore follow Gelman (2006) and use a Half-normal(0, 1) prior distribution for the parameter σ in a multilevel model.

3.5 Practical examples

In this section, we apply Bayesian generalized linear modeling to two examples based on real-world cases. First, we will look at an example of a group audit in which an auditor decides to use stratification. We use this example to demonstrate the benefits of including a multilevel structure in the model. Second, we will look at an example of monetary unit sampling in which an auditor has access to multiple sources of data for each item. We use this example to demonstrate how incorporating additional information into the statistical model can increase the auditor's efficiency.

3.5.1 Example 1: Group audit on misstatements in a retail company

We will start with an example from Derks et al. (2022b) that describes an audit of group financial statements. The organization in question is a retail company consisting of twenty branches across the country. In this case, a group auditor is tasked with forming an opinion on the level of the organization, while twenty component auditors are required to form an opinion on the level of each individual branch. For this reason, the group auditor has decided to implement a stratified sampling procedure. Furthermore, at the branch level, the group auditor anticipates that the number of FTEs responsible for administration within each branch of the auditee will impact the probability of misstatement. We will investigate if incorporating this information makes a difference in the sample evaluation.

3.5.1.1 Data

The component auditors collected a sample from each of the twenty branches. However, because the group auditor expected to find more misstatements at some branches, a larger sample was drawn from these branches than at others. The component auditors report the number of misstatements per branch alongside the number of FTEs, as displayed in Table 3.3. For every branch in the population, one additional data characteristic is known: the number of FTEs e responsible for the administration in that specific branch of the auditee.

Table 3.3: Data set containing the branch index (s) number of FTEs (e), number of items (N), and sample results (n and k) per branch.

Branch (s)	FTEs (e_s)	Items (N_s)	Samples (n_s)	Misstatements (k_s)
1	5	5000	300	21
2	4	5000	300	16
3	4	5000	300	15
4	3	5000	300	14
5	4	5000	300	16
6	3	5000	150	5
7	2	5000	150	4
8	2	5000	150	3
9	2	5000	150	4
10	3	5000	150	5
11	3	10000	50	2
12	5	10000	50	3
13	3	10000	50	2
14	2	10000	50	1
15	1	10000	50	0
16	1	10000	15	0
17	1	10000	15	0
18	1	10000	15	0
19	3	10000	15	1
20	5	4000	15	3

3.5.1.2 Models

To illustrate the effect of the statistical model on the group– and component auditors' conclusions, we specify four generalized linear models for the misstatements k_s in each branch s (see Figure 3.9). First, we define the model $k_s \sim \text{Binomial}(n_s, \text{logit}^{-1}(\xi))$, which assumes that all branches have the same probability of misstatement. Second, we describe the model $k_s \sim \text{Binomial}(n_s, \text{logit}^{-1}(\xi_s))$, which assumes the misstatements in the branches are completely unrelated. Third, we formulate the model $k_s \sim \text{Binomial}(n_s, \text{logit}^{-1}(\mu + \sigma\alpha_s))$, which assumes that there are differences and similarities between the misstatements in the branches. For comparison, none of these models include the number of FTEs as a covariate. Therefore, we specify a final model $k_s \sim \text{Binomial}(n_s, \text{logit}^{-1}(\mu+\sigma\alpha_s+\beta_e \cdot e_s))$ that assumes differences and similarities between the misstatements in the branches and incorporates the number of FTEs as a covariate.

As discussed in Section 3.4, we specify the following prior distributions: $\xi, \xi_s, \mu \sim \text{Logistic}(0, 1)$, $\sigma \sim \text{Normal}(0, 1)^+$, $\alpha_s \sim \text{Normal}(0, 1)$ and $\beta_e \sim \text{Normal}(0, 1)$. After fitting the model to the data, we apply poststratification to arrive at a representative estimate of the probability of misstatement θ in the population. For explanatory purposes, the R (R Core Team, 2022) code for fitting these models to the data in Table 3.3 and applying poststratification can be found in Appendix 3.B.

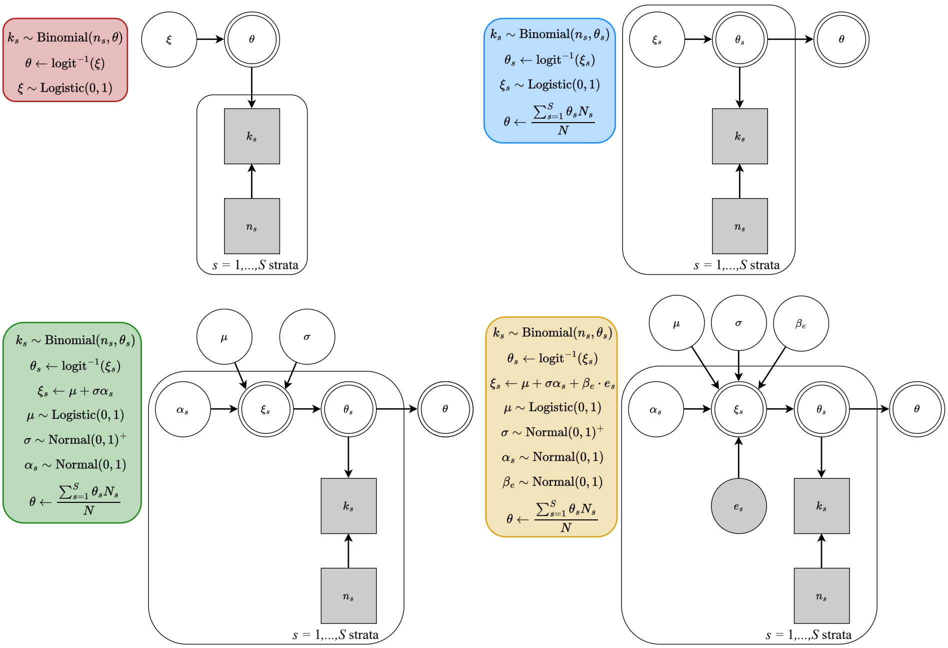


Figure 3.9: Graphical representations of four models to estimate the probability of misstatement θ in a population. The figure shows a model that incorporates no differences between strata (red), a model that incorporates no similarities between strata (blue), a model that incorporates differences and similarities between strata (green), and a model that incorporates differences and similarities and takes into account FTEs (yellow).

3.5.1.3 Comparison of results

Figure 3.10 depicts the population and branch estimates for the four models. The top panel displays the 95 percent HPD intervals (i.e., the 95 percent most likely values of the posterior distribution) and the population estimate of the probability of misstatement under each of the four models. The model that assumes no differences between the branches yields an estimated probability of misstatement of 4.5 percent and a 5.3 percent upper bound. The 95 percent HPD interval of the posterior distribution comes with an uncertainty of 1.6 percent. In comparison, the model that assumes no similarities between the strata yields an estimate of 5.7 percent and an upper bound of 8.1 percent. The uncertainty in the 95 percent HPD interval is 4 percent. Relative to the second model, the estimate of the first model is $\frac{4-1.6}{4}$ = 60 percent less uncertain on the population level. The second model produces not only a more uncertain estimate but also a considerably higher estimate. The latter trait can be understood by the fact that the second model employs the rather conservative uniform prior for every branch. Since the second model can be interpreted as 20 isolated small samples, the dilution of this conservatism by adding observations from the sample is much less in the second model compared to the other models. In comparison, the multilevel model decreases the uncertainty by 52.5 percent over the second model. When additional data, in this case, the number of administrative FTEs, is included in the multilevel model, the population estimate is 62.5 percent less uncertain than under the second model. This demonstrates that the efficiency of the group auditor can be improved by using a multilevel model, and that it can be further improved by including relevant covariates.

The bottom four panels depict the 95 percent HPD intervals around the branches' most likely probability of misstatement. These panels demonstrate how the estimates of the first two models are different from those of the multilevel models. The branch estimates in the first model are identical and come with an average uncertainty of 1.6 percent, while the branch estimates in the second model vary but have an average uncertainty of 11 percent. In comparison, this average is 4 percent for the multilevel model, while including the FTEs results in the estimates having an average uncertainty of 2.8 percent. The panels in the bottom row show how, in the two models that include a multilevel structure, the estimates in the branches (dotted line) are all shifted towards the population estimate (dashed line). This occurs to a greater extent in branches with few samples taken and to a lesser extent in branches with many samples taken. Consider branch 20, in which 15 items were inspected and the auditor had the unfortunate experience of discovering three misstatements. According to the second model, the auditor must conclude that this branch has an exceptionally high probability of misstatement ($\frac{3}{15}$ = 20 percent), which is unlikely based on the sample data from the other branches (Efron and Morris, 1977). However, when estimating the probability of misstatement in this branch, the multilevel model gives a high weight to the population estimate. This is a natural consequence. After all, when there is little information available in a branch, it is preferable to rely more on what is already known from the population. This shows that a multilevel model can not only give the group auditor an accurate estimate but also give the component auditors a more certain estimate of the misstatement in their branch.

In sum, creating a statistical model that incorporates data from various sources can help auditors become more efficient and can increase transparency because auditors can explain what separates misstated items in the population.

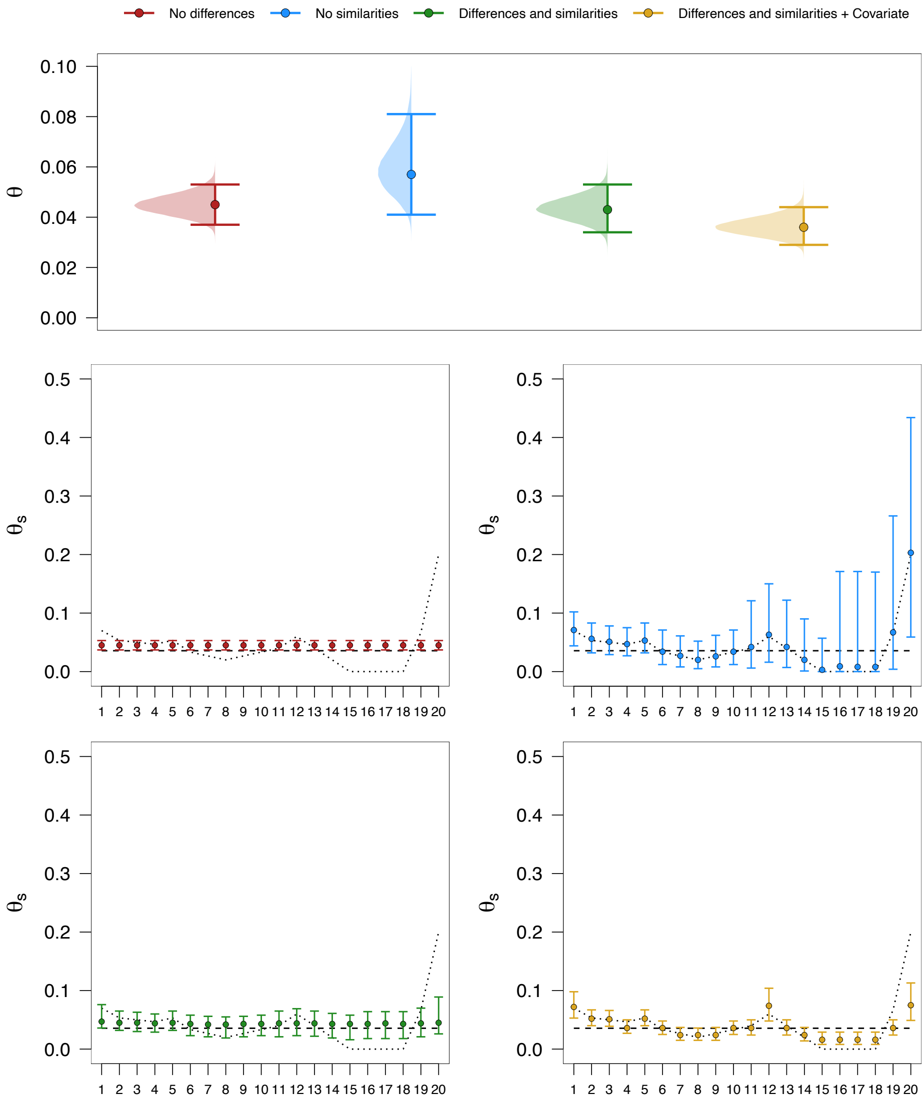


Figure 3.10: Estimates (modes, 95 percent HPD intervals and posterior distribution) for the probability of misstatement in the population (θ, top panel) and in the twenty branches (θ_s, bottom panels, posterior distributions omitted) of a model that incorporates no differences between strata (red), a model that incorporates no similarities between strata (blue), a model that incorporates differences and similarities between strata (green), and a model that incorporates differences and similarities and takes into account FTEs (yellow). The black lines in the bottom panels show, respectively, the proportion of misstatement found in the aggregated sample (dashed line) and the proportion of misstatement found per branch in the sample (dotted line).

3.5.2 Example 2: Audit on legitimacy of subsidy payments

In this example, we will look at an instance of determining the legitimacy of subsidy payments. The scenario is as follows: An auditor is asked to declare with 95 percent certainty that no more than 10 percent of subsidy payments does not comply with the rules and regulations set out in the subsidy scheme. To quantify the uncertainty associated with their conclusion, the auditor must specify a statistical model that connects the data from the sample to a parameter that represents the overstatement in the population. In monetary unit sampling, this statistical model typically only takes into account the data about the overstatements in the sample (Stewart, 2012). However, if there are multiple sources of data available, the statistical model can take this data into account as well. In this example, prior to performing tests of details, the auditor obtained two relevant characteristics for each payment in the population: the length of the time period (in days) between the delivery date of the service and the booking date and the number of FTEs with access to the auditee's computer system who could modify that payment. We will investigate if including this information in the sample evaluation makes the auditor more efficient.

3.5.2.1 Data

The population for this example consists of $N = 3500$ items representing subsidy payments. For illustrative purposes, suppose that the auditor expects 4.6 percent of the sample to contain misstatements. In monetary unit sampling, these options typically correspond to a planned sample size of 100 samples (Stewart, 2012; American Institute of Certified Public Accountants (AICPA), 2019). Because the auditor did not conduct any risk assessment procedures, there is no explicit prior information about the risk of material misstatement in the population. Hence, they decide to select and audit the full sample of 100 payments. In monetary unit sampling, the misstatement in a payment is typically quantified using the taint. The taint t_i of a payment i is the proportional overstatement of a payment, which is defined as a function of the booked value, b_i, and the actual value, a_i, as $t_i = \frac{b_i - a_i}{b_i}$. In the sample of $n = 100$ payments, the auditor discovers ten payments that are partially overstated. Table 3.4 displays these ten partial overstatements in the sample, which result in a total taint of $k = \sum_{i=1}^{n} t_i = 6$. Hence, the average taint in this sample is 6 percent.

3.5.2.2 Models

To construct a statistical model for this scenario, we define a functional relationship between the data and a parameter representing the overstatement in the population. In monetary unit sampling, the data t is not binary but continuous, which means that the Bernoulli distribution cannot be used to describe the data. In particular, because taints lie in the unit interval, we need to select a continuous likelihood that confines this range. The beta distribution conveniently possesses these characteristics, making it a suitable candidate for the likelihood. A typical way to specify the Beta(α, β) likelihood is to define the α and β parameters as

Table 3.4: The ten payments in the sample of $n = 100$ items with a non-zero taint. The table shows each payment's index in the sample (i), booked value (b_i), actual value (a_i), taint (t_i), the booking delay (d_i) and the number of FTEs (e_i) that can modify that payment in the internal computer systems. The ninety items in the sample with a taint of zero are not shown in this table.

i	Book value (b_i)	Actual value (a_i)	Taint (t_i)	Booking delay (d_i)	FTEs (e_i)
2	307.29	184.37	0.4	20	9
17	388.45	155.38	0.6	20	9
22	246.90	123.45	0.5	15	8
45	102.68	51.34	0.5	21	9
46	542.56	108.51	0.8	26	9
57	290.38	174.23	0.4	22	9
63	381.64	114.49	0.7	16	9
74	239.26	47.85	0.8	23	9
88	423.30	126.99	0.7	18	9
99	226.86	90.74	0.6	22	8

$\alpha = \phi\nu$ and $\beta = (1-\phi)\nu$ (Smithson and Verkuilen, 2006; Verkuilen and Smithson, 2012), yielding the statistical model $t_i \sim \text{Beta}(\phi\nu, (1-\phi)\nu)$.

In this model, the parameter $\phi = \frac{\alpha}{\alpha+\beta}$ can be interpreted as the average taint, and $\nu = \alpha + \beta$ is a precision parameter that controls the concentration of the distribution. In this formulation, the parameter ϕ is the one that is most relevant for the auditor's inferences about the overstatement in the population. Because the parameter ϕ is constrained to the unit interval, it is convenient to apply the log-odds transformation $\ln(\frac{\phi}{1-\phi}) = \text{logit}(\phi)$. The advantage of this setup is that a set of parameters can be defined to describe the log-odds of the average taint.

To illustrate the effect of including multiple sources of data into the statistical model, we specify three generalized linear models for the taint t_i in each item i (see Figure 3.11). First, we specify the model $\phi = \text{logit}^{-1}(\beta_0 + \beta_d \cdot d_i)$, which assumes that the taint of a payment is a function of its booking delay. Second, we specify the model $\phi = \text{logit}^{-1}(\beta_0 + \beta_e \cdot e_i)$, which assumes that the taint of a payment is a function of the number of FTEs that can modify that payment. Finally, we specify the model $\phi = \text{logit}^{-1}(\beta_0 + \beta_d \cdot d_i + \beta_e \cdot e_i)$, which assumes that both covariates play a role in determining the taint of a payment.

Because in this scenario there is no pre-existing information about the average taint in the population, we specify a logistic prior distribution for β_0: $\beta_0 \sim \text{Logistic}(0, 1)$. After standardizing the covariates to have a mean of zero and a standard deviation of $\frac{1}{2}$, we specify independent prior distributions for the remaining β parameters in the model: $\beta_d, \beta_e \sim \text{Normal}(0, 1)$. Finally, we specify a Pareto prior distribution for ν: $\nu \sim \text{Pareto}(1, \frac{3}{2})$ (Carpenter, 2016; Derks et al., 2022b).

We transform the taints in the sample of size n in accordance with a standard procedure for dealing with the beta likelihood (Smithson and Verkuilen, 2006, p. 61), yielding a set of transformed taints $t_i = \frac{t_i(n-1)+\frac{1}{2}}{n}$. After fitting the model to the data, we apply poststratification to arrive at a representative estimate of the average taint ϕ in the population. For explanatory purposes, the R (R Core Team, 2022) code for fitting these models to the data partly shown in Table 3.4 and applying poststratification can be found in Appendix 3.B.

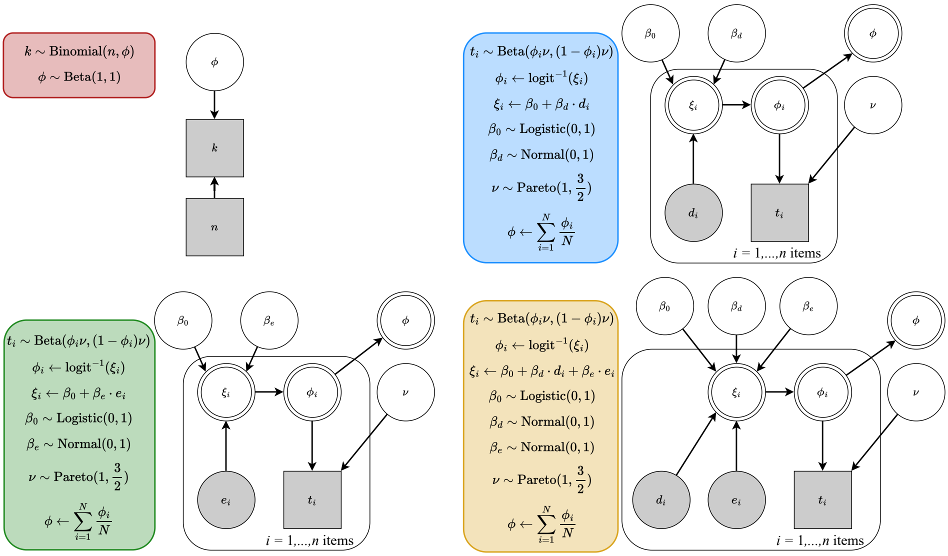


Figure 3.11: Graphical representations of four models to estimate the average taint ϕ in a population. The figure shows the basic model that does not incorporate any covariates (red), a model that incorporates booking delay (blue), a model that incorporates FTEs (green), and a model that incorporates booking delay and FTEs (yellow).

3.5.2.3 Comparison of results

To investigate if including the additional information in the sample evaluation has made the auditor more efficient, we use the results of three basic methods as a baseline. First, evaluating this sample as prescribed by guidance on auditing standards (e.g., American Institute of Certified Public Accountants (AICPA), 2019; Stewart, 2012, pp. 11–12) yields a most likely misstatement of 6 percent and a 95 percent upper bound of 11.84 percent. Second, using a basic Bayesian approach with a uniform prior yields a Beta($1 + k = 7$, $1 + n - k = 95$) posterior distribution that, in comparison with the first baseline, has a reduced 95 percent upper bound of 11.39 percent. Third, evaluation using the commonly used Stringer bound (Bickel, 1992) reduces the 95 percent upper bound further to 11.19 percent. Thus, of the three basic approaches, the Stringer bound yields the most efficient 95 percent upper bound but does not yet permit the auditor to conclude that they have obtained sufficient information to conclude that the misstatement is below the performance materiality of 10 percent. This means that, if the auditor can bring the 95 percent upper bound of their estimate below 10 percent, they will have used their existing sample more efficiently because they will have gathered sufficient information to determine that there are no material misstatements in the population.

Because auditors typically base their conclusion on the 95 percent upper bound of the posterior distribution, they can increase their efficiency by reducing the uncertainty in the posterior distribution for ϕ. Incorporating multiple sources of data achieves this because there is more information available to estimate the average taint in the population. To illustrate, Figure 3.12 shows the posterior distributions for ϕ under the three generalized linear models alongside the baseline Beta(7, 95) distribution. The two models including a single covariate produce a posterior distribution that is less wide and peaks at a lower value than the baseline, which is associated with a reduction in the upper bound. These estimates are less uncertain because more information is available, and they are lower because the sample mostly consists of misstatements in items with high values of the covariates while the population only contains a small number of these items. Furthermore, the model in which both covariates are included produces the most efficient estimate: The most likely average taint in the population is estimated to be 1.9 percent, with a 95 percent upper bound of 2.33 percent. Thus, including both covariates provided the auditor with the greatest increase in efficiency, lowering the upper bound by at least 8.86 percent when compared to a more traditional approach. To put this into perspective, if the auditor wanted to reduce their upper bound by this much using a traditional approach, they would have needed to observe at least 390 additional, error-free samples. Furthermore, by incorporating the covariates into the model, the auditor has obtained sufficient information to conclude that the misstatement in the population is below the performance materiality of 10 percent. This means that the auditor has used their sample more efficiently when compared to a traditional approach.

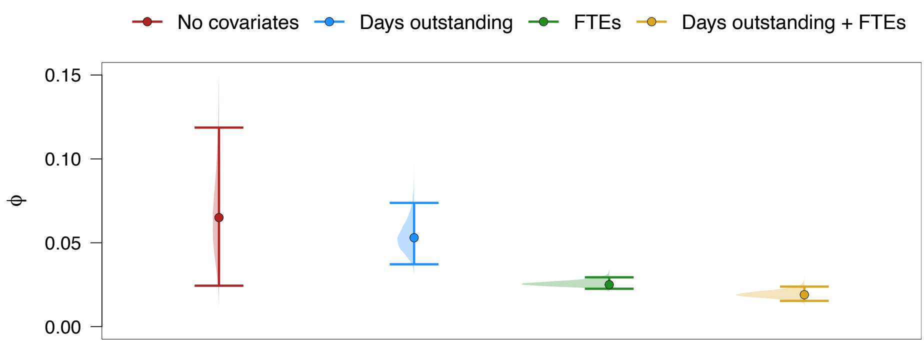


Figure 3.12: Estimates (modes, 95 percent HPD intervals and posterior distribution) for the average taint ϕ in the population of the basic model without covariates (Beta(7, 95), red), a model that includes booking delay (blue), a model that includes FTEs (green), and a model that includes booking delay and FTEs (yellow). In comparison to the basic approach, the inclusion of additional data results in lower and less uncertain estimates and an increase in efficiency.

In the remainder of this subsection, we focus on the results of the model that includes both covariates. Because the auditor statistically estimates the relationship between the number of FTEs, the booking delay, and the average taint, the auditor can use the statistical results to gain an understanding of the misstatement in the population. For example, the parameter β_d is estimated to be 0.456 [0.194; 0.701]. This means that, when the number of FTEs is kept constant, a one-standard-deviation increase in booking delay will likely multiply the odds of the average taint (i.e., $\frac{\phi}{1-\phi}$) by about $e^{0.456} = 1.578$ [1.214; 2.016]. Furthermore, the parameter β_e is estimated to be 1.587 [1.381; 1.786]. Similarly, when the booking delay is kept constant, an increase of one standard deviation in the number of FTEs is likely to multiply $\frac{\phi}{1-\phi}$ by approximately $e^{1.587} = 4.889$ [3.979; 5.966]. From these results, the auditor can infer that the average taint is more strongly influenced by the number of FTEs that have access to the auditee's computer system and could alter a payment than by the booking delay of the payment. Using these findings, the auditor can reveal what distinguishes misstatements in the population and communicate this information to the auditee, who can then respond appropriately. For example, the auditee may decide to limit employees' capabilities to modify payments in their computer systems. If the auditor had evaluated the sample without including the covariates in the statistical model, neither the auditee nor the auditor would have had this information.

Also in a monetary unit sampling context where no stratification is applied, creating a statistical model that incorporates data from various sources can help auditors become more efficient and can increase transparency because auditors can explain what separates misstated items in the population.

3.6 Practical recommendations

In the previous sections, we have discussed and demonstrated the merits of a Bayesian generalized linear modeling approach to audit sampling. However, implementing this approach in practice may come with several challenges for auditors. For instance, it may be difficult to weigh the pros and cons of a Bayesian generalized linear model or it may be difficult to report on the statistical model and its outcomes. A lack of guidance on these issues may unnecessarily make the use of a Bayesian generalized linear model less appealing to auditors. For this reason, we provide recommendations for implementing this approach in practice.

The advantages of this approach have been discussed in detail in this chapter. In previous sections, we have demonstrated how a Bayesian generalized linear modeling approach to audit sampling can assist auditors in providing an audit opinion that is specifically tailored to the audit and the auditee. Because auditors can explain the impact of the data on the misstatement, the auditee can gain valuable insights, and the auditor can add value to the auditee. Moreover, we have shown that a Bayesian generalized linear model can also help to increase efficiency in real-world scenarios.

One aspect of this approach to take into account is that obtaining the data and justifying the statistical model may take time and effort. Hence, the auditor must carefully make an assessment. For example, if evaluating a sample is expensive or if there is a lot of data easily obtainable, the use of a Bayesian generalized linear model can be worthwhile. For instance, if the auditor knows there is information that plays a role in determining the misstatement, this data is easy to obtain and that sampling an extra unit is costly, the benefits of incorporating the information most likely outweigh the time and effort spent justifying the statistical model. However, if they are conducting an audit where collecting more data is costly, the time and money spent justifying the statistical model might not be worth the time and money spent selecting and auditing the samples that are potentially reduced. For example, if they know that obtaining this information is time-consuming or difficult, and that selecting and auditing extra samples is cheap, it may be more pragmatic to evaluate a potentially larger sample using the basic statistical model in Figure 3.2. We recommend that auditors carefully weigh the pros and cons of incorporating data into the statistical model to determine whether the benefits outweigh the time and effort required.

Furthermore, to ensure that the statistical model can be explained and interpreted correctly, the use of a Bayesian generalized linear modeling approach requires auditors to be at least somewhat familiar with statistical inference. In practice, this means that the auditor must collaborate closely with a statistician to construct and explain the statistical model. If this is not possible or desirable, a more straightforward method should be used. We advise auditors to openly discuss these issues with the auditee in order to properly balance the pros and cons of a Bayesian generalized linear model.

Finally, statistical reporting demands more effort when a statistical model is more complex. If auditors decide to use a Bayesian generalized linear model for audit sampling, they must be able to clearly report the statistical results to the auditee. This means that the audit report should include information about the statistical model and its results. When reporting on the statistical outcomes, we recommend providing a description, visualization and justification of the statistical model and the prior distribution, that is, the family of the distribution and its parameter values (van Doorn et al., 2021). In the presentation of parameter estimates, we advise including figures of the prior and posterior distributions along with summaries of the distributions, such as the most likely value and 95 percent HPD interval (see Appendix 3.C for an example report). We also recommend performing and reporting a robustness analysis as described in Section 3.4. Finally, a final and important part of reporting is making the computer code and data available to the auditee in a way that is not only transparent but also enables easy reproduction of the results (Kruschke, 2021). Reporting on the main conclusions in this manner increases transparency for stakeholders of the audit as it allows for a skeptical assessment of the statistical claims in the audit report.

3.7 Concluding comments

In this chapter, we have shown that Bayesian generalized linear modeling can be used to include multiple sources of data in the sample evaluation and, as a result, align the statistical model for audit sampling with the situation in practice. In short, this comes with two concrete advantages for auditors. First, it improves auditors' ability to form an opinion about the misstatement in the population because they can explain the impact of the integrated information on the misstatement in a transparent manner. Second, it improves auditors' ability to detect misstatements because they can more accurately estimate which items in the population are likely to be misstated. Because of these practical benefits, we argue that implementing this approach in practice may assist auditors in overcoming the data-analytic precipice they are currently facing. That is, especially in today's information-rich audits, these techniques are critical for facilitating the effective and efficient integration of audit data into tests of details. Applying these techniques in practice does not have to be difficult; they are supported by user-friendly open-source software such as R (R Core Team, 2022) and JASP (JASP Team, 2022), and code to replicate all examples in this manuscript is available in the appendices to this chapter. Furthermore, we hope that by making these methods available in the open-source software JASP for Audit (Derks et al., 2021b), auditors will use Bayesian modeling to evaluate audit samples more often.

It is worth noting that we have demonstrated an approach to audit sampling that requires the auditor to specify a single statistical model to describe both integrally available data, as well as data that can only be retrieved on a sample basis. However, in practice, the appropriate model for the situation may not always be obvious. It can be difficult, for instance, to determine the relationship between the strata in the data or which covariates do or do not play a role in determining the misstatement. Auditors may want to compare several models and select the one that best describes the data. For this purpose, the Bayesian framework allows auditors to calculate the Bayes factor, a measure that quantifies the statistical evidence in favor of one model over another model (Kass and Raftery, 1995; Fragoso et al., 2018). By quantifying and comparing the evidence for multiple models (e.g., a model with and a model without a specific covariate), auditors can assess whether the addition of a parameter to the model has a meaningful impact on the model predictions. For this reason, the Bayes factor can help auditors select the most appropriate model out of many possible options.

In sum, the use of Bayesian inference, and particularly the use of Bayesian generalized linear modeling, in audit sampling can help auditors better meet expectations for the role of data in an audit. The reason for this is that Bayesian generalized linear modeling allows data to inform the auditor's approach to tests of details. If these techniques manage to gain a foothold in auditing theory, it is not difficult to predict that auditors will evaluate their samples using more advanced statistical models—and reap their benefits—in the near future. It should be noted that the use of a Bayesian (generalized) linear model to account for multiple sources of data is not limited to the context of audit sampling. For instance, the auditor may want to estimate the auditee's cost of sales on the basis of an industry benchmark data set containing multiple data characteristics, or they may want to estimate an employee's salary based on multiple data characteristics. Even though the context is different, Bayesian (generalized) linear models retain their advantages; that is, they can incorporate both pre-existing information and data into the statistical model, resulting in an increase in efficiency for the auditor.

3.A The logistic transformation

The combined output of a set of parameters in a linear model can take on any continuous value in the range $[-\infty, \infty]$. However, in audit sampling, the auditor is interested in estimating the probability of misstatement or the average taint, a quantity that is restricted to the range [0, 1]. Without applying a transformation to the output of the linear model, most output values do not lie in the unit interval. A generalized linear model solves this problem by introducing a link function that converts the output of the statistical model from an unbounded quantity to the unit interval.

Statistically, the link function $g(\theta) = \xi$ connects the value θ in the range [0, 1] to the output ξ of a linear model in the range $[-\infty, \infty]$. Several options for the link function exist (any function whose domain is the unit interval can be used), but for illustrative purposes we have focused on an intuitive link function: the logistic function $g(\theta) = \ln(\frac{\theta}{1-\theta})$, see Figure 3.13. This link function is convenient because it allows for the interpretation of the output of the linear model ξ as the log-odds of the value θ. For example, the value $\theta = 0.1$ corresponds to odds of 19 and therefore log-odds of $\xi = \ln(19) = -2.197$. The inverse link function $g^{-1}(\xi)$ can be expressed as $\frac{1}{1+e^{-\xi}}$ and is used to transform the output ξ of the linear model to the value θ. For instance, a model output of $\xi = -2.197$ is transformed to a value $\theta = \frac{1}{1+e^{2.917}} = 0.1$. We refer the interested reader to Gelman and Hill (2007, Chapter 6) for further reading on the link function.

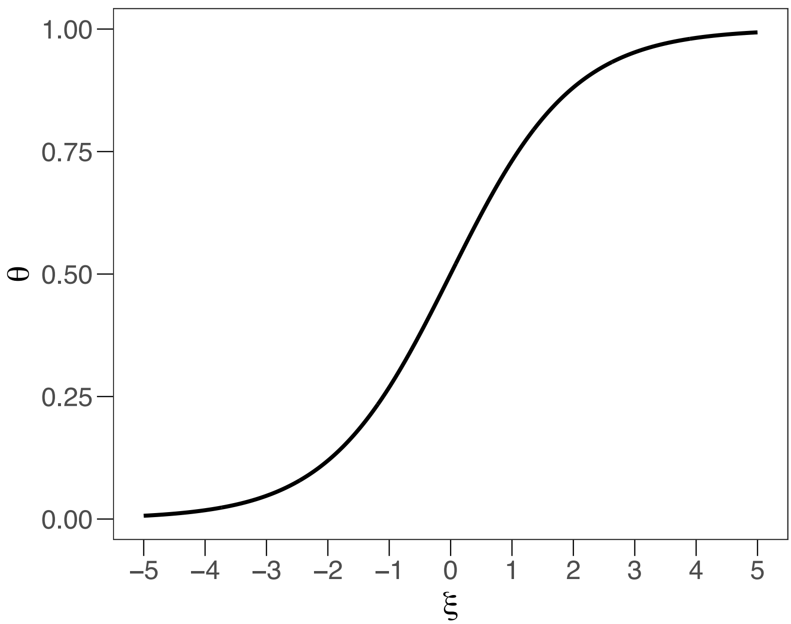


Figure 3.13: The (inverse) logistic link function that transforms the unbounded output ξ of the linear model to the value θ in the unit interval.

3.B R code to reproduce results

This appendix contains R (R Core Team, 2022) and Stan (Carpenter et al., 2017) code for specifying the models discussed in this chapter and reproducing their results. The online appendix containing the R files and the data can be found at `https://osf.io/m7xu5/`.

3.B.1 Example 0: Illustrative

The R code below loads the population from the illustrative example.

```
library(rstan) # Install using install.packages("rstan")

standardize <- function(x) {
  (x - mean(x)) / (sd(x) * 2) # Standardizes input to mean 0 and sd 0.5
}

# Read and standardize data
population <- read.csv("https://osf.io/y9jmc/download")
population$x_std <- standardize(population$x)
sample <- population[!is.na(population$k), ]

# Create variables of relevant data
n <- nrow(sample)
k <- sample$k
S <- length(unique(population$s))
j <- sample$s
x <- sample$x_std
N <- nrow(population)
J <- population$s
X <- population$x_std
```

The model can be fitted to the data by filling in the model code and calling `stan`.

```
model_code <- "" # Fill in the code of the Stan model (see below)
fit <- stan(model_code, data = list(n = n, ...)) # See "data { }" for ...
```

3.B.1.1 No covariate — No branch differences

```
data {
  int<lower=0> n;              //# Number of items in sample
  int<lower=0, upper=1> k[n]; //# Misstatement of items in sample
  int N;                       //# Number of items in population
}
parameters {
  real xi;
}
model {
  xi ~ logistic(0, 1);
  for (i in 1:n) {
```

```
    k[i] ~ bernoulli_logit(xi);
  }
}
generated quantities {
  vector[N] theta_i;
  real theta;
  for (i in 1:N) {
    theta_i[i] = inv_logit(xi);
  }
  theta = mean(theta_i);
}
```

3.B.1.2 No covariate — No branch similarities

```
data {
  int<lower=0> n;              //# Number of items in sample
  int<lower=0, upper=1> k[n]; //# Misstatement of items in sample
  int N;                       //# Number of items in population
  int j[n];                    //# Vector of branch indices in sample
  int J[N];                    //# Vector of branch indices in population
  int S;                       //# Number of branches in population
}
parameters {
  real xi_s[S];
}
model {
  for (s in 1:S) {
    xi_s[s] ~ logistic(0, 1);
  }
  for (i in 1:n) {
    k[i] ~ bernoulli_logit(xi_s[j[i]]);
  }
}
generated quantities {
  vector[N] theta_i;
  real theta;
  for (i in 1:N) {
    theta_i[i] = inv_logit(xi_s[J[i]]);
  }
  theta = mean(theta_i);
}
```

3.B.1.3 No covariate — Multilevel

```
data {
  int<lower=0> n;              //# Number of items in sample
  int<lower=0, upper=1> k[n]; //# Misstatement of items in sample
  int N;                       //# Number of items in population
```

```
  int j[n];                    //# Vector of branch indices in sample
  int J[N];                    //# Vector of branch indices in population
  int S;                       //# Number of branches in population
}
parameters {
  real mu;
  real<lower=0> sigma;
  real alpha_s[S];
}
model {
  mu ~ logistic(0, 1);
  sigma ~ normal(0, 1);
  for (s in 1:S) {
    alpha_s[s] ~ normal(0, 1);
  }
  for (i in 1:n) {
    k[i] ~ bernoulli_logit(mu + sigma * alpha_s[j[i]]);
  }
}
generated quantities {
  vector[N] theta_i;
  real theta;
  for (i in 1:N) {
    theta_i[i] = inv_logit(mu + sigma * alpha_s[J[i]]);
  }
  theta = mean(theta_i);
}
```

3.B.1.4 With covariate — No branch differences

```
data {
  int<lower=0> n;              //# Number of items in sample
  int<lower=0, upper=1> k[n]; //# Misstatement of items in sample
  int N;                       //# Number of items in population
  vector[n] x;                 //# Control intensity in sample
  vector[N] X;                 //# Control intensity in population
}
parameters {
  real beta_0;
  real beta_x;
}
model {
  beta_0 ~ logistic(0, 1);
  beta_x ~ normal(0, 1);
  for (i in 1:n) {
    k[i] ~ bernoulli_logit(beta_0 + beta_x * x[i]);
  }
}
generated quantities {
```

```
  vector[N] theta_i;
  real theta;
  for (i in 1:N) {
    theta_i[i] = inv_logit(beta_0 + beta_x * X[i]);
  }
  theta = mean(theta_i);
}
```

3.B.1.5 With covariate — No branch similarities

```
data {
  int<lower=0> n;                    //# Number of items in sample
  int<lower=0, upper=1> k[n]; //# Misstatement of items in sample
  int N;                             //# Number of items in population
  vector[n] x;                       //# Control intensity in sample
  vector[N] X;                       //# Control intensity in population
  int j[n];                          //# Vector of branch indices in sample
  int J[N];                          //# Vector of branch indices in population
  int S;                             //# Number of branches in population
}
parameters {
  real beta_0[S];
  real beta_x;
}
model {
  for (s in 1:S) {
    beta_0[s] ~ logistic(0, 1);
  }
  beta_x ~ normal(0, 1);
  for (i in 1:n) {
    k[i] ~ bernoulli_logit(beta_0[j[i]] + beta_x * x[i]);
  }
}
generated quantities {
  vector[N] theta_i;
  real theta;
  for (i in 1:N) {
    theta_i[i] = inv_logit(beta_0[J[i]] + beta_x * X[i]);
  }
  theta = mean(theta_i);
}
```

3.B.1.6 With covariate — Multilevel

```
data {
  int<lower=0> n;                    //# Number of items in sample
  int<lower=0, upper=1> k[n]; //# Misstatement of items in sample
  int N;                             //# Number of items in population
```

```
  vector[n] x;                //# Control intensity in sample
  vector[N] X;                //# Control intensity in population
  int j[n];                   //# Vector of branch indices in sample
  int J[N];                   //# Vector of branch indices in population
  int S;                      //# Number of branches in population
}
parameters {
  real mu;
  real<lower=0> sigma;
  real alpha_s[S];
  real beta_x;
}
model {
  mu ~ logistic(0, 1);
  sigma ~ normal(0, 1);
  beta_x ~ normal(0, 1);
  for (s in 1:S) {
    alpha_s[s] ~ normal(0, 1);
  }
  for (i in 1:n){
    k[i] ~ bernoulli_logit(mu + sigma * alpha_s[j[i]] + beta_x * x[i]);
  }
}
generated quantities {
  vector[N] theta_i;
  real theta;
  for (i in 1:N) {
    theta_i[i] = inv_logit(mu + sigma * alpha_s[J[i]] + beta_x * X[i]);
  }
  theta = mean(theta_i);
}
```

3.B.2 Example 1: Group audit on misstatements in a retail company

The R code below loads the population from the example concerning a group audit on misstatements in a retail company.

```
library(rstan) # Install using install.packages("rstan")

standardize <- function(x) {
  (x - mean(x)) / (sd(x) * 2) # Standardizes input to mean 0 and sd 0.5
}

# Read and standardize data
sample <- read.csv("https://osf.io/akrvj/download", sep = ";")
sample$x_std <- standardize(sample$x)

# Create variables of relevant data
S  <- nrow(sample)
```

```
n  <- sample$n
k  <- sample$k
Ns <- sample$N
e  <- sample$x_std
```

The model can be fitted to the data by filling in the model code and calling `stan`.

```
model_code <- "" # Fill in the code of the Stan model (see below)
fit <- stan(model_code, data = list(S = S, ...)) # See "data { }" for ...
```

3.B.2.1 No covariate — No branch differences

```
data {
  int<lower=1> S;    //# Number of branches in population
  int<lower=0> n[S]; //# Number of items per branch
  int<lower=0> k[S]; //# Number of misstatements per branch
}
parameters {
  real xi;
}
model {
  xi ~ logistic(0, 1);
  k ~ binomial_logit(n, xi);
}
generated quantities {
  real<lower=0, upper=1> theta;
  theta = inv_logit(xi);
}
```

3.B.2.2 No covariate — No branch similarities

```
data {
  int<lower=1> S;         //# Number of branches in population
  int<lower=0> n[S];      //# Number of items per branch
  int<lower=0> k[S];      //# Number of misstatements per branch
  vector<lower=0>[S] Ns; //# Number of items in population per branch
}
parameters {
  real xi_s[S];
}
model {
  xi_s ~ logistic(0, 1);
  k ~ binomial_logit(n, xi);
}
generated quantities {
  vector<lower=0, upper=1>[S] theta_s;
  real<lower=0, upper=1> theta;
  for (i in 1:i) {
```

```
    theta_s[i] = inv_logit(xi_s[s]);
  }
  theta = dot_product(Ns, theta_s) / sum(Ns);
}
```

3.B.2.3 No covariate — Multilevel

```
data {
  int<lower=1> S;          //# Number of branches in population
  int<lower=0> n[S];       //# Number of items per branch
  int<lower=0> k[S];       //# Number of misstatements per branch
  vector<lower=0>[S] Ns; //# Number of items in population per branch
}
parameters {
  real mu;
  real<lower=0> sigma;
  vector[S] alpha_s;
}
model {
  mu ~ logistic(0, 1);
  sigma ~ normal(0, 1);
  alpha_s ~ normal(0, 1);
  k ~ binomial_logit(n, mu + sigma * alpha_s);
}
generated quantities {
  vector<lower=0, upper=1>[S] theta_s;
  real<lower=0, upper=1> theta;
  for (i in 1:S) {
    theta_s[i] = inv_logit(mu + sigma * alpha_s[i]);
  }
  theta = dot_product(Ns, theta_s) / sum(Ns);
}
```

3.B.2.4 With covariate — Multilevel

```
data {
  int<lower=1> S;          //# Number of branches in population
  int<lower=0> n[S];       //# Number of items per branch
  int<lower=0> k[S];       //# Number of misstatements per branch
  vector<lower=0>[S] Ns;   //# Number of items in population per branch
  vector[S] e;             //# Number of FTEs per branch
}
parameters {
  real mu;
  real<lower=0> sigma;
  vector[S] alpha_s;
  real beta_e;
}
```

```
model {
  mu ~ logistic(0, 1);
  sigma ~ normal(0, 1);
  alpha_s ~ normal(0, 1);
  beta_e ~ normal(0, 1);
  k ~ binomial_logit(n, mu + sigma * alpha_s + beta_e * e);
}
generated quantities {
  vector<lower=0, upper=1>[S] theta_s;
  real<lower=0, upper=1> theta;
  for (i in 1:S) {
    theta_s[i] = inv_logit(mu + sigma * alpha_s[i] + beta_e * e[i]);
  }
  theta = dot_product(Ns, theta_s) / sum(Ns);
}
```

3.B.3 Example 2: Audit on legitimacy of subsidy payments

The R code below loads the population from the example concerning an audit on legitimacy of subsidy payments.

```
library(rstan) # Install using install.packages("rstan")

standardize <- function(x) {
  (x - mean(x)) / (sd(x) * 2) # Standardizes input to mean 0 and sd 0.5
}

# Read and standardize data
population <- read.csv("https://osf.io/vge5d/download")
population$d_std <- standardize(population$x1)
population$e_std <- standardize(population$x2)
sample <- population[population$insample == 1, ]

# Create variables of relevant data
n <- nrow(sample)
t <- (sample$bookValue - sample$auditValue) / sample$bookValue
d <- sample$d_std
e <- sample$e_std
N <- nrow(population)
D <- population$d_std
E <- population$e_std
t <- (t * (n - 1) + 0.5) / n
```

The model can be fitted to the data by filling in the model code and calling `stan`.

```
model_code <- "" # Fill in the code of the Stan model (see below)
fit <- stan(model_code, data = list(n = n, ...)) # See "data { }" for ...
```

3.B.3.1 Including booking delay

```
data {
  int<lower=0> n;      //# Number of items in sample
  real<lower=0> t[n]; //# Taints of items in sample
  int N;               //# Number of items in population
  vector[n] d;         //# Booking delay in sample
  vector[N] D;         //# Booking delay in population
}
parameters {
  real beta_0;
  real beta_d;
  real<lower=0> nu;
}
model {
  vector[n] phi_i;
  beta_0 ~ logistic(0, 1);
  beta_d ~ normal(0, 1);
  nu ~ pareto(1, 1.5);
  for (i in 1:n) {
    phi_i[i] = inv_logit(beta_0 + beta_d * d[i]);
    t[i] ~ beta(phi_i[i] * nu, (1 - phi_i[i]) * nu);
  }
}
generated quantities {
  vector[N] phi_i;
  real phi;
  for (i in 1:N) {
     phi_i[i] = inv_logit(beta_0 + beta_d * D[i]);
  }
  phi = mean(phi_i);
}
```

3.B.3.2 Including number of FTEs

```
data {
  int<lower=0> n;      //# Number of items in sample
  real<lower=0> t[n]; //# Taints of items in sample
  int N;               //# Number of items in population
  vector[n] e;         //# Number of FTEs in sample
  vector[N] E;         //# Number of FTEs in population
}
parameters {
  real beta_0;
  real beta_e;
  real<lower=0>nu;
}
model {
  vector[n] phi_i;
```

```
  beta_0 ~ logistic(0, 1);
  beta_e ~ normal(0, 1);
  nu ~ pareto(1, 1.5);
  for (i in 1:n) {
    phi_i[i] = inv_logit(beta_0 + beta_e * e[i]);
    t[i] ~ beta(phi_i[i] * nu, (1 - phi_i[i]) * nu);
  }
}
generated quantities {
  vector[N] phi_i;
  real phi;
  for (i in 1:N) {
     phi_i[i] = inv_logit(beta_0 + beta_e * E[i]);
  }
  phi = mean(phi_i);
}
```

3.B.3.3 Including booking delay and number of FTEs

```
data {
  int<lower=0> n;     //# Number of items in sample
  real<lower=0> t[n]; //# Taints of items in sample
  int N;              //# Number of items in population
  vector[n] d;        //# Booking delay in sample
  vector[n] e;        //# Number of FTEs in sample
  vector[N] D;        //# Booking delay in population
  vector[N] E;        //# Number of FTEs in population
}
parameters {
  real beta_0;
  real beta_d;
  real beta_e;
  real<lower=0>nu;
}
model {
  vector[n] phi_i;
  beta_0 ~ logistic(0, 1);
  beta_d ~ normal(0, 1);
  beta_e ~ normal(0, 1);
  nu ~ pareto(1, 1.5);
  for (i in 1:n) {
    phi_i[i] = inv_logit(beta_0 + beta_d * d[i] + beta_e * e[i]);
    t[i] ~ beta(phi_i[i] * nu, (1 - phi_i[i]) * nu);
  }
}
generated quantities {
  vector[N] phi_i;
  real phi;
  for (i in 1:N) {
```

```
    phi_i[i] = inv_logit(beta_0 + beta_d * D[i] + beta_e * E[i]);
  }
  phi = mean(phi_i);
}
```

3.C Example report

Statistical model

The statistical model assumes that the misstatement of an item is a function of the time between login and approval (a proxy for control intensity) of that item.

$k_i \sim \text{Bernoulli}(\text{logit}^{-1}(\beta_0 + \beta_x \cdot x_i)$

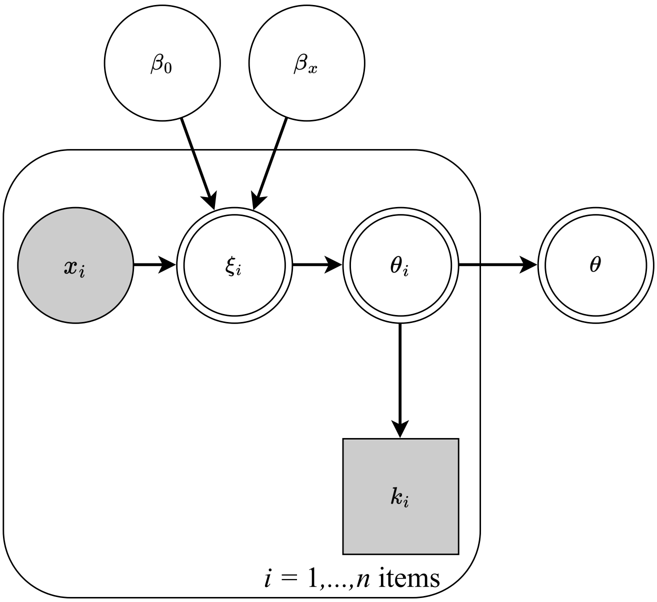
k_i = Misstatement of item i
x_i = Minutes between login and approval (a proxy for control intensity) of item i

Prior distributions for parameters

$\beta_0 \sim \text{Logistic}(0, 1)$
$\beta_x \sim \text{Normal}(0, 1)$

Graphical representation of the model

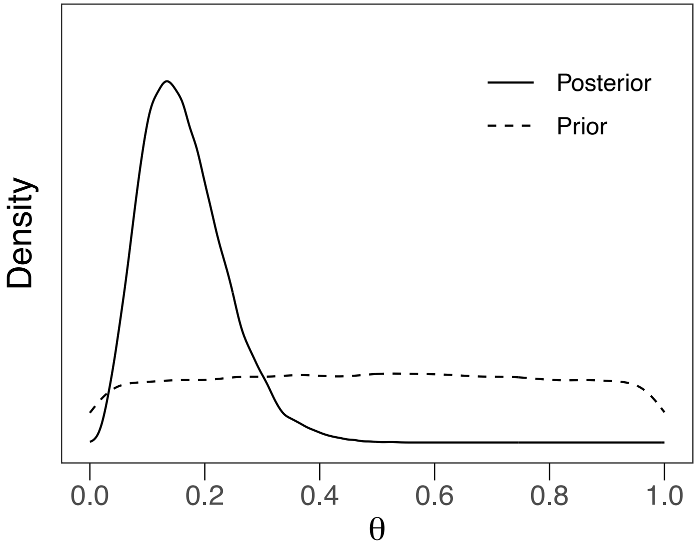
$$k_i \sim \text{Bernoulli}(\theta_i)$$
$$\theta_i \leftarrow \text{logit}^{-1}(\xi_i)$$
$$\xi_i \leftarrow \beta_0 + \beta_x \cdot x_i$$
$$\beta_0 \sim \text{Logistic}(0, 1)$$
$$\beta_x \sim \text{Normal}(0, 1)$$
$$\theta \leftarrow \sum_{i=1}^{N} \frac{\theta_i}{N}$$



Parameter estimates

		95% HDI	
Parameter	Estimate	Lower	Upper
β_0	-1.78	-3.18	-0.666
β_x	-0.954	-2.693	0.581
Population	0.142	0.035	0.311

Prior and posterior distribution



Part II

Bayesian Hypothesis Testing

Chapter 4

Quantifying Statistical Audit Evidence using the Bayes Factor

Abstract

The impact of statistical methods on the audit practice is growing because of the increasing availability of audit data and the statistical methods to analyze these data. A key aspect in the statistical approach to auditing is assessing the strength of evidence for or against a hypothesis. Unfortunately, the often-used frequentist statistical methods cannot provide the statistical evidence that audit standards demand directly nor easily. In this chapter we discuss an alternative approach that can provide this evidence: Bayesian inference. Firstly, we explore the philosophical differences between frequentist and Bayesian inference. Secondly, we discuss misconceptions in the interpretation of frequentist statistical evidence, and finally we discuss how Bayesian inference allows the auditor to obtain and interpret statistical evidence in line with audit standards via its alternative to the p-value, the Bayes factor. We contribute to audit theory and practice by showing how Bayesian inference can quantify audit evidence.

Keywords: Audit evidence, analytical procedures, Bayes factor, substantive testing.

4.1 Introduction

In today's society, auditors play a key role in preserving the integrity of companies, (non-profit) organizations, and governments. The objective of the auditor is to provide stakeholders of the auditee with an opinion with reasonable assurance about the completeness, accuracy, and fairness of the assertions as presented by the auditee's management in the financial statements (ISA 200, International Auditing

This chapter is under review for publication as Derks, K., de Swart, J., Wagenmakers, E.–J., & Wetzels, R. (2022). The Bayesian approach to audit evidence: Quantifying statistical evidence using the Bayes factor. doi: `https://doi.org/10.31234/osf.io/kzqp5`

and Assurance Standards Board (IAASB), 2018; AU-C 200, American Institute of Certified Public Accountants (AICPA), 2021; AS 5, Public Company Accounting Oversight Boards (PCAOB), 2020). To obtain this reasonable assurance, auditors are expected to conduct an audit of the organization as prescribed by the International Standards on Auditing (ISA International Auditing and Assurance Standards Board (IAASB), 2018), Generally Accepted Auditing Standards (GAAS American Institute of Certified Public Accountants (AICPA), 2021), or the Auditing Standards (AS Public Company Accounting Oversight Boards (PCAOB), 2020). These audit standards mandate that the auditor's opinion must be based on persuasive (rather than conclusive) audit evidence consisting of information that can support or contradict management's assertions in the financial statements (ISA 500, paragraph 5c, International Auditing and Assurance Standards Board (IAASB), 2018; AU-C 500, paragraph 6, American Institute of Certified Public Accountants (AICPA), 2021; AS 1105, paragraph 2, American Institute of Certified Public Accountants (AICPA), 2021). Unfortunately, the often-used frequentist statistical methods cannot provide the statistical evidence that audit standards demand directly nor easily. In this chapter we introduce an alternative approach that can quantify this evidence in line with auditing standards: Bayesian inference.

Two types of audit evidence can be distinguished: non-statistical and statistical. Non-statistical audit evidence is collected from supervision, inquiry, or correspondence with the auditee (Bennett and Hatfield, 2013; Perry, 2011; Yin, 2020). Statistical audit evidence is collected from statistical procedures and analyses performed by the auditor (Gillett and Srivastava, 2000; van den Acker, 2000). For example, statistical evidence can be obtained by testing a subset of relevant control systems to determine whether or not they meet quality requirements (Li et al., 2020), by performing analytical procedures to assess relationships between financial and non-financial data in the auditee's databases (Appelbaum et al., 2018; Daroca and Holder, 1985), or by performing audit sampling to obtain reasonable assurance that auditee's financial statements are free of material misstatement (American Institute of Certified Public Accountants (AICPA), 2019; Dowling and Leech, 2007).

Traditionally, a large part of statistical audit evidence was obtained from audit sampling (Hirst and Koonce, 1996; Trompeter and Wright, 2010). However, in recent decades, the emergence of big data, as well as the growing use of data analytics in the audit practice has brought new opportunities for the use of statistical techniques other than audit sampling as an additional source of audit evidence (Appelbaum et al., 2017). For example, regression analysis can be used to identify and evaluate trends in the balance of payments over time or to verify the auditee's financial data against an industry benchmark (Appelbaum et al., 2018; Bednarek et al., 2016). However, regardless of the type of statistical analysis, the auditor requires data to perform statistical inference on a certain characteristic of the auditee.

Where statistical sampling is applied, probability theory is required (ISA 530, paragraph 5g, International Auditing and Assurance Standards Board (IAASB), 2018). There are two main schools of probability theory: frequentist and Bayesian (Wagenmakers et al., 2008). In current practice, frequentism is the dominant methodology to collect statistical audit evidence. Audit guides implicitly nudge auditors towards a frequentist hypothesis testing framework to, for example, evaluate their samples using confidence intervals or p-values (American Institute of Certified Public Accountants (AICPA), 2019; Stewart, 2012). However, frequentism has several well-known drawbacks. Most importantly, a frequentist hypothesis test does not give the auditor what the audit standards desire: statistical evidence that can support or contradict the auditor's conclusions. In particular its main decision-making tool, the p-value, is not a measure of statistical evidence, it strictly provides only indirect support against the conclusions made by the auditor. Moreover, the p-value is ineffective at quantifying evidence that can support the auditor's conclusions, and it is inefficient since monitoring evidence as the data come in and stop gathering data when a certain evidential threshold is reached (e.g., when the p-value is smaller than 0.05) is not allowed (Wagenmakers, 2007; van Batenburg, 2018e). For these reasons, the frequentist methodology has raised persistent concerns about efficiency, transparency, and applicability to the audit practice in the scientific literature (Beck et al., 1985; Hubbard and Lindsay, 2008; Johnstone, 1986, 1990; Kim et al., 2018; Kinney, 1975; Scott, 1973). In Section 4.3, we will go into more detail about the concerns of frequentist hypothesis testing when it comes to auditors' effectiveness and efficiency.

Both the frequentist and Bayesian approaches facilitate attribute sampling, monetary unit sampling, estimation using confidence or credible intervals, and—most importantly in an audit context—statistical hypothesis testing. Nonetheless, the two approaches differ in their approach to hypothesis testing and therefore in how the resulting statistical evidence can be interpreted. Most importantly, in a frequentist approach the auditor samples data with the aim to falsify a certain null hypothesis and ignores the likelihood of the data under an alternative hypothesis. There is a disconnect between this approach and the auditing practice because if the goal of the auditor is to quantify how much evidence the data provide for or against the null hypothesis it is unavoidable that an alternative hypothesis is specified (Goodman and Royall, 1988). On the other hand, in a Bayesian approach the auditor samples data with the aim of obtaining relative evidence for two competing hypotheses. The rationale is that there is evidence for the null hypothesis over the alternative hypothesis if the data are more likely to occur under the null hypothesis (and vice versa). Because a Bayesian approach considers the likelihood of the data under two competing hypotheses, it is possible to directly compare the relative evidence for the hypotheses. This allows auditors to obtain statistical evidence that can support or contradict their conclusions, something that is often desired but impossible using a frequentist or nonstatistical approach.

As an alternative to frequentist hypothesis testing, Bayesian hypothesis testing allows the auditor to quantify statistical evidence using the Bayes factor (Kass and Raftery, 1995). As mentioned, Bayesian hypothesis testing primarily focuses on quantifying evidence, and it has therefore been advocated as a more informed statistical framework, both for accounting research (Johnstone, 2021; Kim et al., 2018) and for the audit practice (Johnstone, 2018; Stewart, 2013). A key aspect of Bayesian inference is that the auditor needs to specify a so-called prior distribution. The prior distribution represents the information about an unknown parameter of interest to the auditor, which in turn is combined with the probability distribution of new data (e.g., the likelihood) to yield the posterior distribution. This updating of information from the prior to the posterior distribution forms the basis of any Bayesian inference.

Even though the audit is inherently a continuous process (Leslie, 1984), and audit evidence is considered "cumulative in nature" (ISA 200, paragraph A30, International Auditing and Assurance Standards Board (IAASB), 2018; AU-C 500, paragraph A3, American Institute of Certified Public Accountants (AICPA), 2021), the use of the Bayesian philosophy in the audit practice is scarce. This may, in part, be explained by the fact that literature discussing the role of Bayesian philosophy in the quantification of audit evidence is scarce. This is unfortunate since the Bayesian philosophy has several properties that fit well with the audit process. For example, in this chapter we show that the Bayes factor as a measure of audit evidence fits the auditor well because it can quantify evidence for and against hypotheses (Wagenmakers, 2007), it allows for sequential adding of information (Rouder, 2014), and Bayesian analyses allow for incorporation of expert knowledge or other existing information into the statistical analysis (Corless, 1972). For these reasons, we believe that the Bayes factor has the potential to enhance the way that auditors currently analyze and evaluate statistical evidence from a sample. While not the first mention of the Bayes factor in an audit context (see Johnstone, 2018, pp. 33–34), the main contributions of this chapter are to explain how the Bayes factor enables the auditor to intuitively quantify audit evidence, to show how it can be applied in practice using a variety of relevant examples, and to enable practical use by supplying easy-to-use, free, and open-source software to calculate the Bayes factor. In Section 4.3 we describe in detail how Bayesian hypothesis testing using the Bayes factor addresses the concerns of frequentist hypothesis testing when it comes to auditors' effectiveness and efficiency.

The structure of this chapter is as follows. We will first provide a theoretical introduction into statistical audit evidence. Next, we discuss and compare statistical audit evidence from a frequentist and Bayesian point of view. We show that—in contrast to the frequentist p-value—the Bayes factor can quantify evidence that supports or contradicts the auditor's conclusions regarding management's assertions in the financial statements. In the section thereafter we reason that the p-value limits the auditor in their activities and that the Bayes factor does not suffer from these limitations. Finally, to illustrate how the auditor can quantify evidence from a wide range of activities, we use the Bayes factor in a reanalysis of four typical audit questions. The last section presents our concluding comments.

4.2 Two approaches to statistical audit evidence

Audit evidence is the subject of auditing standards ISA 500 (IAASB), AU-C 500 (AICPA) and AS 1105 (PCAOB). IAASB standard ISA 500 describes audit evidence as "[i]nformation used by the auditor in arriving at the conclusions on which the auditor's opinion is based" (paragraph 5c) and states that audit evidence "comprises both information that supports and corroborates management's assertions, and any information that contradicts such assertions" (paragraph A.1, International Auditing and Assurance Standards Board (IAASB), 2018). AICPA standard AU-C 500 (paragraph 6) describes audit evidence as "[i]nformation used by the auditor in arriving at the conclusions on which the auditor's opinion is based". It also states that "Audit evidence is information to which audit procedures have been applied and consists of information that corroborates or contradicts assertions in the financial statements" (American Institute of Certified Public Accountants (AICPA), 2021). PCAOB standard AS 1105 (paragraph 2) describes audit evidence as "all the information, whether obtained from audit procedures or other sources, that is used by the auditor in arriving at the conclusions on which the auditor's opinion is based" and states that "Audit evidence consists of both information that supports and corroborates management's assertions regarding the financial statements or internal control over financial reporting and information that contradicts such assertions" (Public Company Accounting Oversight Boards (PCAOB), 2020). Unmistakably, the auditing standards desire that audit evidence should be able to support or contradict the auditor's conclusions about management's assertions in the financial statements.

For example, to obtain audit evidence to support or contradict the hypothesis that the auditee's recorded financial transactions do not contain misstatement that exceeds materiality, an auditor can inspect a subset of all financial transactions for their accuracy. However, because the auditor only inspects a subset of the population a statistical hypothesis test is needed to assess the level in which the evaluated subset of data supports or contradicts the auditor's hypothesis about all the data. To make concrete how the auditor can engage in statistical hypothesis testing, we will focus on a specific activity in the audit where statistical inference is common: audit sampling.

4.2.1 Hypothesis testing in audit sampling

Audit sampling enables the auditor to obtain evidence with respect to a specific hypothesis about the misstatement in the population from which items are selected. Because the auditor only inspects a sample of the population, the hypothesis cannot be evaluated with absolute certainty. However, because the auditor is required to obtain a reasonable assurance they must evaluate the hypothesis to a level of certainty, and therefore it needs to be clear how much information is required to reach this level. The audit standards prescribe that the information from a sample is sufficient when it has reduced the sampling risk to an acceptably low level (ISA 530, paragraph 5c, International Auditing and Assurance Standards Board (IAASB), 2018; AU-C 530, paragraph 5, American Institute of Certified Public Accountants (AICPA), 2021). There are two types of sampling risk that can lead to an incorrect conclusion about the financial misstatements (Elliott and Rogers, 1972). First, there is α: the risk of incorrectly deciding that the population contains material misstatement when in fact it does not (i.e., a Type-I error or the risk of underreliance). Second, there is β: the risk of incorrectly deciding that the population does not contain material misstatement when in fact it does (i.e., a Type-II error or the risk of overreliance). According to the audit standards, auditors are primarily concerned with reducing the second type of risk, β, since it affects effectiveness and their ability to provide an appropriate audit opinion. The α risk is mentioned in the context of efficiency as it "would usually lead to additional work to establish that initial conclusions were incorrect" (ISA 500, paragraph 5c, International Auditing and Assurance Standards Board (IAASB), 2018). For example, AICPA standard AU-C 200 states that "audit risk does not include the risk that the auditor might express an opinion that the financial statements are materially misstated when they are not. This risk is ordinarily insignificant" (AU-C 200, paragraph A7, American Institute of Certified Public Accountants (AICPA), 2021).

If the auditor engages in statistical audit sampling, the audit standards state that they can quantify the sampling risk (ISA 530, paragraph 5g, International Auditing and Assurance Standards Board (IAASB), 2018; AU-C 530, paragraph .05, American Institute of Certified Public Accountants (AICPA), 2021; AU 350, paragraph .46, Public Company Accounting Oversight Boards (PCAOB), 2020). However, when doing so the auditor is free to choose which philosophy of probability is applied: a frequentist or a Bayesian philosophy. To illustrate the differences between these two philosophies, we will evaluate a running example of an hypothesis test along the lines of both philosophies.

Suppose that an auditor is required to assess the financial statements of a publicly traded company. The auditee's financial statements incorporate, among other accounts, a subsidy provided by the government to hire temporary staff when necessary. In this example, a contract for temporary staff is only legal when it contains a valid signature. Furthermore, government norms mandate that for the full subsidy to be legal, no more than three percent of the temporary contracts can contain an invalid signature. Over the course of this year, the publicly traded company has used the money from the subsidy to employ 500 staff members on the basis of such a temporary contract. In this case, the auditor investigates if the subsidy provided to the company is fully legal. The auditor's one-sided null hypothesis stating that the misstatement in the temporary contracts does not exceed three percent can be formulated as $\mathcal{H}_0$: $\theta \leq 0.03$, whereas the one-sided alternative hypothesis, which is the opposite of the null hypothesis, reads $\mathcal{H}_1$: $\theta > 0.03$. In these hypotheses, θ represents the proportion of contracts with an invalid signature. A different way of testing these hypotheses is to include the value of the performance materiality in the alternative hypothesis, see Appendix 4.A for further details.

However, since these contracts have not been subject to an audit before the true value of θ is unknown. Therefore, the auditor would like to decide about the credibility of the hypothesis $\mathcal{H}_0$ by selecting several temporary contracts and determining the validity of the signature. Suppose that, after inspecting a sample of 99 temporary contracts from the auditee's archives, the auditor finds that none of these contracts contain an invalid signature.

4.2.2 Frequentist null hypothesis testing

The traditional method of analyzing these sample outcomes to arrive at a conclusion about θ is frequentist null hypothesis significance testing (NHST) (American Institute of Certified Public Accountants (AICPA), 2019; Elliott and Rogers, 1972; Stewart, 2012). In NHST, the auditor formalizes a so-called null hypothesis that represents the minimal value of θ at which their alternative hypothesis is not supported (Fisher, 1934), which in our example is the hypothesis $\mathcal{H}_0$: $\theta \leq 0.03$. Unfortunately, NHST only allows for quantifying evidence against the null hypothesis $\mathcal{H}_0$. This means that the auditor is unable to quantify evidence that can support the null hypothesis and, furthermore, is unable to quantify evidence that supports the alternative hypothesis if an alternative hypothesis is defined.

When testing the null hypothesis, the auditor assumes that $\mathcal{H}_0$ is true and sets out to gather data to evaluate $\mathcal{H}_0$. The rationale behind NHST is that increasingly stronger evidence will be obtained against $\mathcal{H}_0$ (the population is free of material misstatement) when the data become increasingly implausible assuming the truth of $\mathcal{H}_0$. If sufficient evidence is obtained that contradicts $\mathcal{H}_0$, it can be rejected with reasonable assurance. NHST allows the auditor to quantify the evidence against $\mathcal{H}_0$ using the p-value, which expresses the probability of seeing the observed sample outcome or more extreme—but unobserved—sample outcomes, assuming the truth of $\mathcal{H}_0$.

To perform the statistical inference, the data from the sample needs to be connected to the null hypothesis about the population parameter by means of a probability distribution (Lehmann and Romano, 2006). For example, if the auditor assumes that the contracts in the sample are all independent observations and that the only parameter that exerts influence on the sample outcomes is the probability of misstatement θ, a binomial distribution can be applied (Johnstone, 1990; Sorensen, 1969), see Equation 4.2.1.

$$\text{Probability of } k \text{ errors in } n \text{ items} = \binom{n}{k} \theta^k (1-\theta)^{n-k} \tag{4.2.1}$$

To determine when $\mathcal{H}_0$: $\theta \leq 0.03$ should be rejected for a given sample size n, the auditor must calculate the maximum number of invalid signatures k that can be observed while the risk of incorrectly rejecting the null hypothesis is still sufficiently low. Suppose that the auditor has determined α—the risk of incorrectly deciding that the population contains material misstatement when in fact it does not—to be five percent. This threshold is referred to as the significance level. In this case, the rule for rejection of $\mathcal{H}_0$ is $k \geq 7$ because if it is true that $\theta = 0.03$, then the probability of finding 7 or more invalid signatures in the sample of 99 items equals 2.98 percent (see Equation 4.2.1). This probability is lower than the required significance level of 5 percent. Using a lower cutoff, as in $k = 6$, violates the required α risk (i.e., 7.78 percent); using a higher cutoff, as in $k = 8$, is less efficient because the α risk is much lower than required (i.e., 1 percent). This procedure for rejecting $\mathcal{H}_0$ can also be regarded as rejecting the null hypothesis if the p-value associated with the observed value of k is less than or equal to the significance level α.

To continue our running example, the p-value is the probability of finding $k = 0$ to $k = 99$ invalid signatures in the sample, given that the temporary contracts contain three percent misstatement, and equals $p = 1$ (Equation 4.2.2). Since the calculated p-value of 1 is higher than the significance level $\alpha = 0.05$, the auditor cannot reject the null hypothesis.

$$p = \sum_{i=0}^{k=99} \binom{99}{k} 0.03^k (1 - 0.03)^{99-k} = 1 \tag{4.2.2}$$

A default conclusion for a p-value larger than the significance level α is to not reject—and thus maintain—the null hypothesis $\mathcal{H}_0$: $\theta \leq 0.03$. However, the interpretation of a larger p-value as support for the null hypothesis is flawed (Goodman, 2008). The fallacy in this statement is best described through the words "absence of evidence is not evidence of absence" (Altman and Bland, 1995; Keysers et al., 2020). Simply put, the finding that data contain no evidence against the null hypothesis does not imply that they contain evidence that supports the null hypothesis. Since the p-value is solely a measure of evidence against the null hypothesis, it fails to address the extent to which the sample provides support for the null hypothesis. Thus, based on the p-value, auditors are unable to quantify statistical evidence that can support their conclusion that the misstatement in the population does not exceed the performance materiality.

On the other hand, a p-value lower than the significance level α generally leads the auditor to reject the null hypothesis $\mathcal{H}_0$: $\theta \leq 0.03$, and to accept the alternative hypothesis $\mathcal{H}_1$: $\theta > 0.03$. However, as we have discussed, in NHST the p-value only concerns the null hypothesis $\mathcal{H}_0$ and not an alternative hypothesis. Therefore, if the auditor uses this procedure to substantiate their conclusion that the sample supports the opposing hypothesis, $\mathcal{H}_1$, they fall into a statistical trap (Berger and Sellke, 1987; Berkson, 1942; Wagenmakers, 2007). The (im)plausibility of the alternative hypothesis is not considered in the computation of the p-value, and since the computation of the p-value is solely based on the evaluation of the data in light of the null hypothesis it provides only an indirect argument for the alternative hypothesis. Thus, the p-value fails to address the extent to which the sample supports the alternative hypothesis $\mathcal{H}_1$. Based on the p-value, the auditor is unable to quantify statistical evidence that can support the conclusion that the misstatement in the population exceeds the performance materiality. This also implies that auditors are unable to statistically contradict the conclusion that the misstatement in the population does not exceed the performance materiality.

As mentioned in the auditing standards, the auditor can also calculate the sampling risk β—the risk of deciding that the population does not contain material misstatement when in fact it does. In order to calculate β for an alternative point hypothesis about the population misstatement, the auditor needs to make an assumption about θ. Suppose the auditor assumes that the population misstatement is equal to $\theta = 0.04$, which is slightly higher than the performance materiality. When assuming the truth of this hypothesis, the risk of failing to reject the null hypothesis $\mathcal{H}_0$: $\theta \leq 0.03$ is the probability of finding an outcome that would yield a p-value above five percent (e.g., $k = 7$ misstatements would give $p < .05$ and thus lead to rejection of $\mathcal{H}_0$). Hence, the sampling risk β can be calculated as the probability of finding $k = 0$ to $k = 6$ invalid signatures in the sample under the Binomial($k \mid n = 99, \theta = 0.04$) distribution and equals $\beta = 0.90$. This calculation shows that, if the population truly contains misstatement slightly higher than performance materiality, there is a 90 percent chance to incorrectly decide that the population does not contain material misstatements. The reason that this is high is because the assumed population material misstatement is very close to the upper bound of the null hypothesis. The sampling risk β would be lower if the sample size would be higher, or if the assumed population material misstatement would be higher.

Since the auditing standards state that audit evidence consists of information that can support or contradict the auditor's conclusions, the inability of the p-value to provide support for the null and alternative hypothesis makes it arguably unsuited for quantifying statistical evidence in an audit context. Unfortunately, this applies to any (analytical or substantive) procedure in which the auditor quantifies statistical audit evidence using the p-value. However, in the next section we show that, by using a Bayesian hypothesis test, the auditor can quantify the required statistical evidence via the Bayes factor.

4.2.3 Bayesian hypothesis testing

In contrast to NHST, where the auditor's evidence is solely based on the model for the null hypothesis $\mathcal{H}_0$, a Bayesian hypothesis test incorporates both hypotheses $\mathcal{H}_1$: $\theta > 0.03$ and $\mathcal{H}_0$: $\theta \leq 0.03$ into the statistical procedure. The driving force behind Bayesian inference is Bayes' theorem (Jeffreys, 1939), which stipulates how existing information about an event A can be updated using information from a new event B (Equation 4.2.3).

$$p(A \mid B) = p(A) \times \frac{p(B \mid A)}{p(B)} \tag{4.2.3}$$

On a conceptual level, Bayes' theorem embodies a fundamental principle in the audit: the notion that audit evidence is "cumulative in nature" (ISA 200, paragraph A.30, International Auditing and Assurance Standards Board (IAASB), 2018; AU-C 500, paragraph A3, American Institute of Certified Public Accountants (AICPA), 2021) and that the auditor can therefore aggregate audit evidence over the audit.

In line with accumulating evidence, it is important to state the current level of information before an analysis is performed. Hence, in the Bayesian framework, a prior distribution needs to be defined for every aspect of the statistical model that is to be estimated. The prior distribution is a probability distribution that represents the auditor's existing information about the possible values of parameters or hypotheses. In audit sampling, the prior distribution can be used to incorporate existing evidence about the possible values of the misstatement θ into the sampling procedure (Corless, 1972). For example, the auditor's risk assessments on the inherent risk and control risk are information that can be incorporated into the prior distribution (Derks et al., 2021a; Stewart, 2013).

To use Bayes' theorem for hypothesis testing, the auditor must also first quantify their existing evidence about the plausibility of the two competing hypotheses using so-called prior probabilities. The prior probability $p(\mathcal{H}_0)$ incorporates the auditor's existing evidence about the probability of the null hypothesis $\mathcal{H}_0$ occurring before seeing any information from a data sample. Vice versa, the prior probability $p(\mathcal{H}_1)$ incorporates the auditor's existing evidence about the probability of the alternative hypothesis $\mathcal{H}_1$. This makes it possible that information for or against a given hypothesis is evaluated and incorporated prior to choosing the sample to test. The ratio of prior probabilities is called the prior odds and is an indication of the relative plausibility of the hypotheses before analyzing the intended sample.

When performing an audit, new information y from a sample is observed and the auditor aims to update the prior probability $p(\mathcal{H}_i)$ of the hypothesis $\mathcal{H}_i$ to a posterior probability $p(\mathcal{H}_i \mid y)$. This is done via Bayes' rule, which allows the auditor to update their prior knowledge about the hypothesis $p(\mathcal{H}_i)$ with the evidence that is contained in the data for or against this hypothesis $p(y \mid \mathcal{H}_i)$, resulting in the posterior probability of this hypothesis (see Equation 4.2.4). Given the cumulative nature of audit evidence described in the audit standards, this philosophy of revising and aggregating evidence is more in line with the audit practice than that of NHST.

$$\underbrace{p(\mathcal{H}_i \mid y)}_{\text{Posterior probability}} = \underbrace{p(\mathcal{H}_i)}_{\text{Prior probability}} \times \underbrace{\frac{p(y \mid \mathcal{H}_i)}{p(y)}}_{\text{Evidence}} \tag{4.2.4}$$

The posterior probability $p(\mathcal{H}_i \mid y)$ represents the probability that the hypothesis $\mathcal{H}_i$ is true, conditioned on the existing (prior) information and the evidence from the sample. For example, a posterior probability $p(\mathcal{H}_0 \mid y) = 0.95$ implies that, given the existing audit evidence and the evidence in the sample, there is a 95 percent probability of correctly deciding that $\mathcal{H}_0$ is true. Hence, the posterior probabilities can be intuitively related to the sampling risks α and β. More concretely, when accepting $\mathcal{H}_0$, the posterior probability $p(\mathcal{H}_1 \mid y)$ can be interpreted as the β risk. Vice versa, when rejecting $\mathcal{H}_0$ and accepting $\mathcal{H}_1$, the posterior probability $p(\mathcal{H}_0 \mid y)$ can be interpreted as the α risk.

However, because the auditor is interested in comparing the evidence for two hypotheses, they can employ Bayes' theorem to obtain the ratio of posterior probabilities for $\mathcal{H}_0$ and $\mathcal{H}_1$, the posterior odds. The posterior odds can be denoted as the product of the prior odds and the relative evidence for the hypotheses (see Equation 4.2.5).

$$\underbrace{\frac{p(\mathcal{H}_0 \mid y)}{p(\mathcal{H}_1 \mid y)}}_{\text{Posterior odds}} = \underbrace{\frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)}}_{\text{Prior odds}} \times \underbrace{\frac{p(y \mid \mathcal{H}_0)}{p(y \mid \mathcal{H}_1)}}_{\text{Relative evidence}} \tag{4.2.5}$$

Since the posterior odds depend on the prior odds as well as the information from the sample, and because it can be very difficult to define the prior odds, it is common practice to quantify the relative evidence that the sample contains using the ratio of evidence for the two hypotheses. This ratio is called the Bayes factor, and it quantifies the change in prior to posterior odds brought about by the data (Kass and Raftery, 1995). Compared to the p-value, the Bayes factor is a direct comparison of the evidence for both hypotheses on the sample (see Equation 4.2.6).

$$BF_{01}(y) = \frac{p(y \mid \mathcal{H}_0)}{p(y \mid \mathcal{H}_1)} \tag{4.2.6}$$

Since the Bayes factor is a ratio, it can quantify evidence in both directions. It is this specific property of the Bayes factor that fits well with the audit standards' description of audit evidence because it enables the auditor to quantify evidence that can support their hypotheses as well as evidence that can contradict their hypotheses. For example, a Bayes factor in favor of $\mathcal{H}_0$ of 7 ($BF_{01} = 7$) indicates that the sample outcomes are 7 times more likely to occur under the null hypothesis $\mathcal{H}_0$ than under the alternative hypothesis $\mathcal{H}_1$. Furthermore, because $BF_{10} = \frac{1}{BF_{01}} = \frac{1}{7}$, $BF_{01} = 7$ also indicates that the sample outcomes are 7 times less likely to occur under the alternative hypothesis $\mathcal{H}_1$ than under the null hypothesis $\mathcal{H}_0$. Because of the ease of interpretation of the Bayes factor, it is rapidly being adopted in many areas of business and science such as Psychology (Heck et al., 2022; Ly et al., 2016), Sociology (Bollen et al., 2012; Lynch and Bartlett, 2019), and Economy (Cipriani et al., 2012; Richard and Vecer, 2021). Furthermore, Bayes factor calculations have been made very easy in many standard situations such as the (partial) correlation test (Wetzels and Wagenmakers, 2012), the t-test (Rouder et al., 2009; Wetzels et al., 2009), or the ANOVA (Rouder et al., 2012; Wetzels et al., 2012) and have been implemented in easy-to-use software such as JASP (JASP Team, 2022; Love et al., 2019).

In a similar fashion to the p-value there exist only subjective decision rules for what Bayes factor represents sufficient evidence since what may be considered convincing evidence in a low-risk audit might not be considered as convincing evidence in a high-risk audit. However, to aid practitioners in interpreting the Bayes factor, a collection of labels has been proposed and reiterated in the scientific literature (Jeffreys, 1961; Wetzels et al., 2011; van Doorn et al., 2021). Table 4.1 displays these evidential thresholds, which auditors can use to interpret the strength of evidence provided by the Bayes factor in practice. Because the Bayes factor is a ratio $BF_{01} = 7$ implies moderate evidence in favor of the null hypothesis and at the same time implies moderate evidence against the alternative hypothesis. In similar fashion, $BF_{01} = 20$ implies strong evidence in favor of the null hypothesis and at the same time strong evidence against the alternative hypothesis. Note that, although p-values and Bayes factors mostly agree about which hypothesis is supported by the data, they often disagree about the strength of this support (Wetzels et al., 2011).

In sum, because the Bayes factor can quantify audit evidence in both directions it is more in line with the philosophy of evidence described in the audit standards than the p-value. However, the Bayes factor is not only an attractive alternative to the p-value because of its intuitive theoretical interpretation, it also removes some of the practical limitations that the p-value brings. In the next section we describe these limitations of the p-value in more detail and explain why the Bayes factor does not suffer from these limitations. In Section 4.4 we will discuss four practical examples of Bayes factor calculations in an audit context.

Table 4.1: Bayes factor labels as proposed by Jeffreys (1961).

$BF_{01} = \frac{1}{BF_{10}}$	Strength of evidence
$< \frac{1}{100}$	Extreme evidence for $\mathcal{H}_1$
$\frac{1}{100} - \frac{1}{30}$	Very strong evidence for $\mathcal{H}_1$
$\frac{1}{30} - \frac{1}{10}$	Strong evidence for $\mathcal{H}_1$
$\frac{1}{10} - \frac{1}{3}$	Moderate evidence for $\mathcal{H}_1$
$\frac{1}{3} - 1$	Anecdotal evidence for $\mathcal{H}_1$
1	No evidence for $\mathcal{H}_1$ or $\mathcal{H}_0$
$1 - 3$	Anecdotal evidence for $\mathcal{H}_0$
$3 - 10$	Moderate evidence for $\mathcal{H}_0$
$10 - 30$	Strong evidence for $\mathcal{H}_0$
$30 - 100$	Very strong evidence for $\mathcal{H}_0$
> 100	Extreme evidence for $\mathcal{H}_0$

4.3 Practical implications

In this section we illustrate that the use of the p-value limits auditors in their effectiveness and in their efficiency in quantifying audit evidence. Next, we show that the Bayes factor does not suffer from these limitations and that it is therefore an attractive alternative to the p-value. To illustrate which improvements the Bayes factor brings, we focus on two properties of the p-value that we believe have practical implications for the auditor: the p-value cannot provide evidence for the null hypothesis, and it does not allow for sequential sampling (Wasserstein and Lazar, 2016; Rouder, 2014; Wagenmakers, 2007; Wagenmakers et al., 2019). We explain why the Bayes factor does not suffer from these limitations and what this implies for the auditor in practice.

4.3.1 The p-value cannot provide evidence for the null hypothesis

The p-value can lead to an ineffective audit when the auditor wishes to support the null hypothesis, but due to the nature of frequentist hypothesis testing they are unable to obtain statistical evidence that quantifies this support. This limits the auditor because it takes away many possibilities for statistical analyses in which a null hypothesis is the focus of the investigation. The Bayes factor does not suffer from this limitation because it allows the auditor to quantify evidence in both directions, thereby enabling the auditor to obtain evidence for the null hypothesis if desired.

One can think of many scenarios (other than the sampling scenario in the previous section) where supporting the null hypothesis is the main goal of the statistical analysis, as doing so can contribute to the existing evidential matter in the audit. Suppose that the auditor wants to support the null hypothesis that an auditee's inventory is valued fairly, wants to confirm the auditee's accounts receivable or wants to confirm the auditee's sales transactions. As we have discussed in Section 4.2, relying on the p-value makes supporting this null hypothesis impossible. This can be appreciated by noting that $p = .35$ will not be strong evidence for the null hypothesis when the sample size is small (i.e., absence of evidence), but the same p-value will be strong evidence for the same hypothesis when the sample size is large (i.e., evidence of absence). To illustrate, a sample of 14 items containing 1 misstatement and a sample of 284 items containing 10 misstatements both yield $p = .35$. Hence, it is not possible to gather evidence supporting the null hypothesis based on the p-value. In Section 4 we discuss two more examples in which the auditor wants to support the null hypothesis: a situation in which the auditor wants to support the conclusion that the data in the auditee's financial statements are subject to Benford's law (Example 2) and a situation where an auditor of a tax authority wants to support the conclusion that all taxable persons are treated equally (Example 4). We show in more detail how in these commonly occurring scenarios the p-value does not fit well with the audit question at hand because it is unable to quantify support for the null hypothesis.

However, by reporting a Bayes factor the auditor can quantify evidence directly in favor of the null hypothesis, thereby removing this limitation and providing a more fitting answer to their question. In a Bayesian approach the auditor can make a statement about how much more likely the data are under the null hypothesis versus the alternative hypothesis. This makes it possible to directly compare the evidence in the data for both hypotheses, thus allowing the auditor to support or contradict their conclusion. Hence, the Bayesian approach to audit evidence fits well with the audit question at hand because supporting a null hypothesis can contribute just as much to the existing evidential matter as rejecting a null hypothesis.

4.3.2 The p-value does not allow for sequential sampling

The p-value can lead to an inefficient audit when the auditor already has enough evidence to support a particular hypothesis, but due to the nature of frequentist hypothesis testing they still need to perform the remainder of the planned work. This limits the auditor in their efficiency because at this point, they are performing more work than necessary. The Bayes factor does not suffer from this limitation since it allows the auditor to monitor the evidence for any hypothesis while the data come in (Rouder, 2014). Having access to such information during the audit increases efficiency for the auditor because it allows them to modify sampling procedures at an early stage when necessary, or to stop sampling procedures when sufficient evidence is obtained.

In a frequentist analysis, the auditor must complete data collection before analyzing the sample results (Berger and Wolpert, 1988; Lindley, 1993). That is because, given the sampling risk α, each time the auditor looks at the intermediate results there is an α percent chance that a significant p-value is produced when the null hypothesis is true. As a result, the probability of incorrectly rejecting the null hypothesis increases as a function of the number of times that the auditor looks at the results (Armitage et al., 1969; Wagenmakers, 2007). Thus, to maintain control over this type of sampling risk in the frequentist framework, the auditor must finish their intended sampling plan before analyzing the sample. In contrast, because a Bayesian analysis is not dependent on a sampling plan, the auditor is allowed to monitor the evidence for a particular hypothesis and to stop data collection when enough evidence is obtained (Wagenmakers et al., 2019). From a Bayesian point of view "It is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience." (Edwards et al., 1963, p. 193).

To illustrate the benefits of sequential sampling, let's consider the following example. Suppose that the auditor wants to obtain evidence to support the assertion that a certain population contains misstatements lower than a certain threshold t. Statistically speaking, they can then define the null hypothesis as: $\mathcal{H}_0$: $\theta \geq t$. In this case, the auditor wants to sample until they can reject the null hypothesis. They have planned a sample size such that—when no misstatements are found—they can reject the null hypothesis that the population contains material misstatement with a sampling risk α of five percent. As it turns out, the sample contains a single misstatement, which means that the auditor cannot reject this null hypothesis. If the auditor still wants to be able to reject the null hypothesis using the p-value, they will need to plan an extension for their sample. Because there is an increase in the sampling risk α after looking at the data, one possible way to proceed is to plan a follow-up sample in which they adjust the maximum p-value. However, this practice generally results in a substantial extension of the sample. To make this concrete, in most audit guides it is prescribed that the auditor inspects at least an additional number of items equal to the initial sample (American Institute of Certified Public Accountants (AICPA), 2019, Appendix B). For this case, that would imply an increase in the sample size from $n = 99$ to at least $n = 198$.

In contrast, in the Bayesian framework the auditor is allowed to build upon the information from the previous sample without penalty. Therefore, they can coherently extend their sample from $n = 99$ to $n = 156$ (the sample size that they would have got when initially planning for one misstatement in the sample), or to any other n depending on the desired strength of evidence (i.e., the desired Bayes factor). Other than being more efficient in terms of the sample size, this Bayesian sample size extension is arguably more intuitive and easier to explain for the auditor than a frequentist one.

4.4 Applying the Bayes factor in a modern audit

To facilitate the use of the Bayes factor and illustrate the benefits of a Bayesian approach to audit evidence, we will now apply the Bayes factor to four typical audit questions. Please see Appendix 4.B for details about the derivations of various statistics and the calculations of the Bayes factors in this section. As a first example, we will reanalyze the audit sampling scenario in Section 4.2 from a Bayesian point of view. In the second example, we will apply digit analysis using Benford's law to a financial data set. In the third example, we will analyze historical data from an auditee's sales revenue to uncover evidence that supports a potential seasonal effect. In the final example, we analyze an auditee's classification algorithm to investigate to what degree the data reflect algorithmic fairness.

4.4.1 Example 1: Evaluating an audit sample

A Bayesian auditor starts their audit sampling procedure by first specifying a prior probability distribution $p(\theta)$ that reflects their existing information about the parameter θ: the misstatement in the population. The two hypotheses $\mathcal{H}_0$: $\theta \leq 0.03$ and $\mathcal{H}_1$: $\theta > 0.03$ are defined as the range of the prior distribution that corresponds to the hypotheses' restrictions with respect to θ. This means that the prior probability for the hypothesis $\mathcal{H}_0$ corresponds to the total probability under the prior distribution on θ in the range [0; 0.03]. Vice versa, the prior probability for the hypothesis $\mathcal{H}_1$ corresponds to the total probability under the prior distribution on θ in the range (0.03; 1]. To be able to compute the probability under both hypotheses the auditor needs to define the prior probability distribution that fits this situation.

If the auditor has assessed inherent risk and internal control risk according to the Audit Risk Model, they can incorporate this information into the prior distribution. For illustrative purposes it is convenient to specify a uniform Beta(1, 1) prior distribution that represents negligible information about the misstatement θ (Stewart, 2013). The auditor has assessed both inherent risk and internal control risk as "medium" which, according to their audit guide, translates into a reduction in the sample size of $\Delta n = 33$. These unseen samples are assumed to be correct and can be incorporated in the prior distribution by setting the β parameter of the prior distribution to $1 + \Delta n = 34$ (Derks et al., 2021a; Steele, 1992). For the Beta($\alpha = 1$, $\beta = 34$) distribution, the prior odds in favor of the hypothesis $\mathcal{H}_0$ are $\frac{0.645}{0.355} = 1.817$, see Figure 4.1.

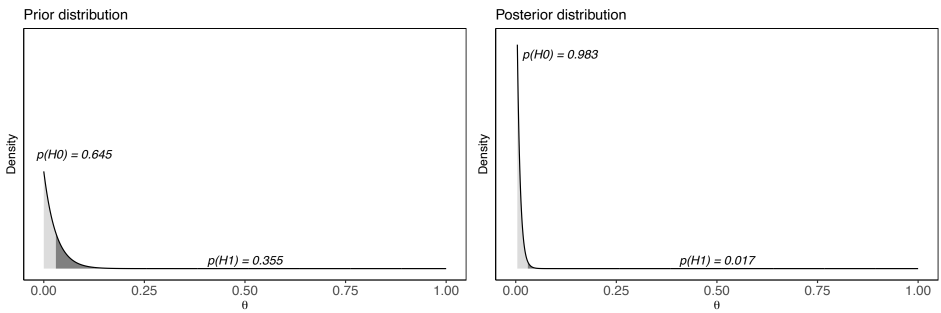


Figure 4.1: The Beta(1, 34) prior distribution (left panel) and Beta(1, 133) posterior distribution (right panel) on the misstatement proportion θ after seeing a sample of $n = 99$ items containing $k = 0$ misstatements. The prior and posterior probabilities for $\mathcal{H}_0$ (light) and $\mathcal{H}_1$ (dark) induced by the prior and posterior distributions are shown in numbers.

After seeing the information from a sample of n items of which k contain a misstatement, the prior distribution is updated by the binomial likelihood to a posterior distribution $p(\theta \mid n, k)$ according to Bayes' theorem (see Equation 4.4.1).

$$\underbrace{p(\theta \mid n, k)}_{\text{Posterior}} = \underbrace{p(\theta)}_{\text{Prior}} \times \underbrace{\frac{p(k \mid n, \theta)}{p(y = k, n)}}_{\text{Evidence}} \tag{4.4.1}$$

Similar to the prior distribution, the posterior distribution induces a probability for the occurrence of the hypotheses. The posterior probability for the hypothesis $\mathcal{H}_0$ corresponds to the probability mass assigned by the posterior distribution to the values of θ in the range [0; 0.03]. Vice versa, the posterior probability for the hypothesis $\mathcal{H}_1$ corresponds to the probability mass assigned by the posterior distribution to the values of θ in the range (0.03; 1].

After being updated by the sample of $n = 99$ items of which $k = 0$ contain a misstatement, the posterior distribution is the Beta$(1 + 0 = 1, 34 + 99 = 133)$ distribution. The posterior odds in favor of $\mathcal{H}_0$ induced by the posterior distribution are therefore $\frac{0.983}{0.017} = 57.824$, see Figure 4.1. The posterior probability $p(\mathcal{H}_0 \mid y) = 0.983$ implies that there is a 98.3 percent probability that the population does not contain misstatements that exceed the performance materiality. This means that, when accepting $\mathcal{H}_0$, there is a 1.7 percent probability that the auditor incorrectly judges that the population is free of material misstatement. This probability is sufficiently low to find the statement in the null hypothesis credible. Vice versa, this also implies that there is a 98.3 percent probability that the auditor correctly judges that the population is not materially misstated.

Because we know the prior odds and the posterior odds, we can calculate the Bayes factor by dividing the two. Thus, the Bayes factor in this example can be calculated as $BF_{01} = \frac{57.824}{1.817} \approx 31$, which implies that the data are about 31 times more likely to occur under $\mathcal{H}_0$ than under $\mathcal{H}_1$. This Bayes factor implies very strong evidence in favor of $\mathcal{H}_0$ (see Table 4.1).

4.4.1.1 Comparison of frequentist and Bayesian conclusions

Note that a frequentist analysis ($p = 1 > \alpha$) only facilitates a statement about the (im)plausibility of the data (or data more extreme) under the hypothesis $\mathcal{H}_0$, and forces the auditor to conclude that the null hypothesis cannot be rejected. As mentioned in the previous section, using this p-value the auditor cannot say that there is evidence in favor of the null hypothesis. The Bayes factor differs from the p-value in that it can quantify evidence directly in favor of the null hypothesis and that it provides an intuitive interpretation of this evidence. That is, the Bayes factor $BF_{01} \approx 31$ shows that $\mathcal{H}_0$ is many times more likely than $\mathcal{H}_1$ and that there is strong evidence in favor of the null hypothesis that the misstatement in the population does not exceed the performance materiality.

4.4.2 Example 2: Assessing Benford's law

Benford's law (Benford, 1938) has been advocated as a simple, (arguably) effective method for auditors to not only identify discrepancies in data, but to uncover potential data manipulation in financial statements (Durtschi et al., 2004), ERP systems (Ma'arif et al., 2020), or official information released by authorities (Wei and Vellwock, 2020). Simply put, Benford's law states that in many naturally occurring collections of numbers the leading digit is likely to be small. More concretely, a set of numbers is said to satisfy Benford's law if the leading digit $d \in \{1, \ldots, 9\}$ occurs with probability

$$p(d) = \log_{10}(1 + \frac{1}{d_i}). \tag{4.4.2}$$

Benford's law is, among other applications, mentioned as an analytical procedure in an early stage of the audit (Nigrini and Mittermaier, 1997). For example, small deviations from Benford's law may suggest that the data have passed a reasonableness test, while large deviations may be a sign of possible data manipulation or the need for further investigation (Drake and Nigrini, 2000). The goal of the analysis in this example is to determine how much evidence the data provide for the statement that the leading digits in the recorded values of a population of items follow Benford's law.

4.4.2.1 Data

The data for this example come from the financial statements of the Sino Forest Corporation's 2010 Report (Nigrini, 2012). For illustrative purposes, we will only analyze the leading digits of the recorded values. The frequencies of the leading digits in the sample are displayed in Table 4.2.

Table 4.2: Descriptive statistics for the first digits in the Sino Forest data set.

Leading digit	Count	Percentage	Benford's law
1	231	29.92%	30.1%
2	124	16.06%	17.61%
3	97	12.56%	12.49%
4	70	9.07%	9.69%
5	64	8.29%	7.92%
6	54	6.99%	6.69%
7	40	5.18%	5.8%
8	54	6.99%	5.12%
9	38	4.92%	4.58%

4.4.2.2 Frequentist analysis

In the NHST framework, the auditor wants to test the null hypothesis $\mathcal{H}_0$ that the first digits are distributed according to Benford's law. An example application of this procedure is described by Varma and Khan (2012), who used Benford's law to identify potential fraud in a similar population. The hypothesis $\mathcal{H}_0$: $p_d = \log_{10}(1 + \frac{1}{d})$ is assessed by means of the p-value.

Using a Chi-squared test ($X^2 = 7.652$, $df = 8$), the p-value for these data is 0.468. The interpretation of this p-value is: Assuming that the first digits are distributed according to Benford's law, there is a 46.8 percent probability that the auditor would have found the observed (or more extremely deviating) distribution of first digits in the data set. In a standard fashion, the conclusion would be to not reject, and thus maintain, the null hypothesis $\mathcal{H}_0$.

4.4.2.3 Bayesian analysis

In a Bayesian analysis of Benford's Law (Good, 1967; Sarafoglou et al., 2021), the null hypothesis is compared, by means of the Bayes factor, against the alternative hypothesis which states that the first digits are not distributed according to Benford's law (i.e., the digit probabilities are free to vary). The prior probabilities for the hypotheses are set to be equal: $p(\mathcal{H}_0) = p(\mathcal{H}_1) = 0.5$. The prior distribution for the alternative hypothesis is assumed to be a Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_9)$ distribution with all α parameters set to 1. Note that, in this case, a parameter α_d of the Dirichlet distribution reflects the prior count for the digit d and can be adjusted to incorporate prior information into the alternative hypothesis.

The corresponding Bayes factor in favor of the null hypothesis is $BF_{01} = 6899678$, which implies that the data are 6899678 times more likely (extreme evidence) to have occurred under the hypothesis that the first digits are distributed according to Benford's law than under the hypothesis that they are not. Because the prior probabilities are set to be equal, the Bayes factor equals the posterior odds, which implies that the posterior probability for the null hypothesis can be deduced as $p(\mathcal{H}_0 \mid y) = \frac{BF_{01} \times p(\mathcal{H}_0)}{BF_{01} \times p(\mathcal{H}_0) + (1 - p(\mathcal{H}_0))} = 0.999$. This means that, when accepting the null hypothesis $\mathcal{H}_0$, there is a 0.1 percent probability that the auditor incorrectly accepts the null hypothesis. Vice versa, there is a 99.9 percent probability that the auditor correctly accepts the null hypothesis.

4.4.2.4 Comparison of frequentist and Bayesian conclusions

The p-value of 0.468 leads the auditor to not reject the null hypothesis $\mathcal{H}_0$. Based on this low p-value, the auditor cannot say the data contain evidence that supports the conclusion that the auditee's data follow Benford's law. The Bayes factor $BF_{10} = 689978$ facilitates the conclusion that the data contain extreme evidence in favor of the conclusion that the auditee's data follow Benford's law.

4.4.3 Example 3: Uncovering seasonal patterns

We now turn to a situation where the auditor uses historical data in an analytical procedure. In particular, the auditor is concerned with the question of how much evidence there is that the sales of the auditee are influenced by seasonal factors. For example, yearly sales revenues may be increasing, but revenues in June might be lower than in September.

As part of the risk assessment process, the auditor can form expectations of patterns that can reasonably be anticipated in the current audit. Often, these expectations involve references to earlier years or industry benchmarks. For example, the auditor may be interested in whether the sales of the auditee are subject to seasonal effects. In addition to a seasonality effect, the auditor wants to know to what extent these historical data support a difference in sales between each season.

4.4.3.1 Data

The data for this example consist of monthly sales of the auditee over the course of the years 2013–2016 ($n = 48$). These data are plotted over time and categorized by season in Figure 4.2.

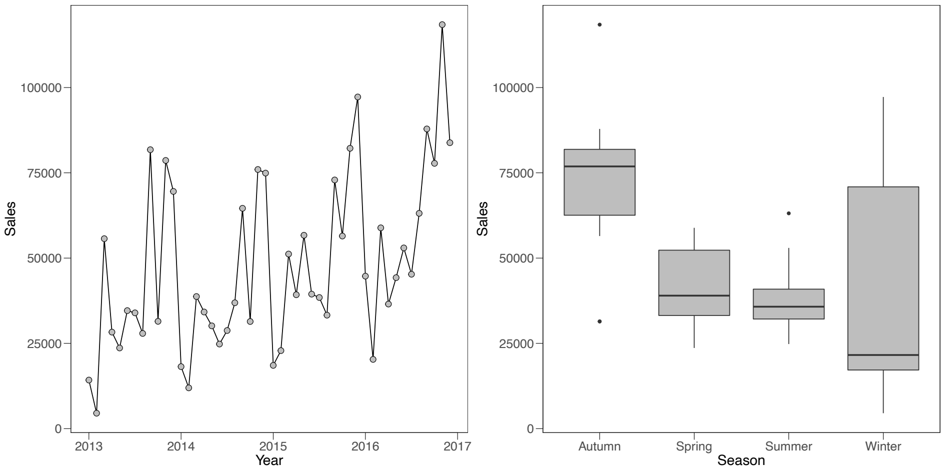


Figure 4.2: Monthly sales of the auditee over the course of the years 2013–2016. The left panel shows the sales over time, and the right panel shows the sales categorized by season.

4.4.3.2 Frequentist analysis

In the frequentist analysis, the null hypothesis $\mathcal{H}_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ is assessed via an ANOVA by means of the p-value. To get to know more about the specific seasons and to find out if they are significantly different, post-hoc tests are performed and assessed using Tukey's p-value, corrected for multiple testing.

Note that we apply ANOVA using a toy model as an illustrative example. In practice, there will be other variables that may also hold predictive value for the sales of the auditee. In all cases, the auditor must carefully construct their statistical model, but can include additional variables using an ANCOVA or a regression analysis.

We present the results of a frequentist ANOVA testing a main effect for season and post-hoc tests. The results of the ANOVA indicate a significant effect for season ($R^2 = 0.3042966, df = 3, F = 6.415, p < .01$). The interpretation of this p-value is: Assuming that there is no seasonal effect, there is less than one percent probability that the auditor would find the observed (or more extremely deviating) outcomes in the data set. However, as we have shown, the auditor cannot deduce the strength of evidence for the statement that there is a seasonal effect from these p-values. Post-hoc tests also indicate a significant difference ($p < .01$) in the autumn months when compared to the other months, see Table 4.3. However, when the auditor uses these p-values to substantiate this conclusion, they again fall into the aforementioned statistical trap.

Table 4.3: ANOVA post-hoc comparisons for season.

		Mean Diff.	SE	t	p tukey	Prior	Posterior	$BF10, U$
Autumn	Spring	30168.261	8876.126	3.399	0.008	0.414	16.053	38.755
	Summer	33339.340	8876.126	3.756	0.003	0.414	41.542	100.291
	Winter	31554.932	8876.126	3.555	0.005	0.414	1.775	4.285
Spring	Summer	3171.079	8876.126	0.357	0.984	0.414	0.183	0.442
	Winter	1386.672	8876.126	0.156	0.999	0.414	0.156	0.376
Summer	Winter	-1784.408	8876.126	-0.201	0.997	0.414	0.156	0.378

Note. P-value adjusted for comparing a family of 4. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall et al., 1997). Individual comparisons are based on the default t-test with a Cauchy(0, $r = 1/\sqrt{2}$) prior. The "U" in the Bayes factor denotes that it is uncorrected.

4.4.3.3 Bayesian analysis

In a Bayesian analysis, the null model of no effect is compared, by means of the Bayes factor, against an alternative model which incorporates the season as a predictive variable (Rouder et al., 2012; Wetzels et al., 2012). The post-hoc tests will also be evaluated using the Bayes factor based on the default Bayesian t-test (Rouder et al., 2009; Wetzels et al., 2009). In these post-hoc tests, the posterior odds have been corrected for multiple testing by fixing the prior probability that the null hypothesis holds across all comparisons to 0.5 (Westfall et al., 1997).

The Bayes factor for the model that includes a main effect for season over the model that does not is $BF_{10} = 33.923$. This Bayes factor implies that the observed data are 33.923 times more likely to have occurred under the hypothesis of a seasonal effect than under the hypothesis of no seasonal effect, which implies strong evidence for a seasonal effect (see Table 4.1). To answer the question how much more likely it is that, for example, the autumn season differs from the other seasons, the auditor must inspect the Bayes factors $BF_{10,U}$ obtained from the individual comparisons in Table 4.3. These Bayes factors indicate strong evidence for the statement that the autumn season differs from the spring and winter seasons. However, the data contain only moderate evidence for a difference in the autumn months compared to the winter months.

4.4.3.4 Comparison of frequentist and Bayesian conclusions

The fact that $p < .01$ leads the auditor to reject the null hypothesis $\mathcal{H}_0$. However, based on this low p-value the auditor cannot say there is evidence that supports the conclusion that there is a seasonal effect in the data. In contrast to the p-

value, the auditor can use the Bayes factor of $BF_{10} = 33.924$ to substantiate the conclusion that the data contain strong evidence in favor of a seasonal effect.

4.4.4 Example 4: Determining algorithmic bias

As a final example, we consider an increasingly relevant issue in the context of big data and artificial intelligence. With the rapid growth of information systems that collect and mine customer data, an increasing portion of auditees' business decisions is being guided by artificial intelligence (AI). On 21 April 2021, the European Commission presented a proposal for a regulation concerning artificial intelligence—the AI Act, for short (European Commission, 2021). One major focus of the AI Act is the classification of various types of AI systems according to the risks involved. One of the risks that has special attention is that application of AI might lead to unfair treatment and discrimination. Attention should therefore be given towards ensuring that decisions made with the aid of these algorithms remain fair (Kearns et al., 2018).

For example, such algorithms must avoid exhibiting discriminatory biases towards features such as gender, race, or age. Suppose that the auditor works with an auditee in the banking industry that employs an algorithm to predict whether people are going to default on a loan. Naturally, it is highly undesirable that, given that a customer is actually going to pay their loan, they are more likely to get classified by the algorithm as possibly defaulting on that loan as a result of their ethnicity. The following analytical procedure aims to test this algorithmic fairness with respect to ethnicity.

To illustrate this procedure, we focus on a relatively simple criterion of algorithmic fairness. This criterion requires equality of false positive or negative rates across all subgroups in the data (Hardt et al., 2016). In the context of the example algorithm, a false positive would imply that a person gets wrongly marked as a possible defaulter. A false negative on the other hand would mean that a customer will likely default on their loan, but no action will be taken by the bank as this person is not identified by the algorithm. The algorithm may possibly display racial bias if the probability that a customer gets wrongly marked as a possible defaulter is higher for some ethnic groups than for others. Statistically, this implies that the probability of a false positive should be the same across ethnic groups (i.e., the algorithm's classification is independent of a customer's ethnicity). To find out how much evidence there is for this hypothesis, we describe one possible analysis the auditor can use.

4.4.4.1 Data

We use a fictional benchmark data set ($n = 10,000$) from the field of credit risk prediction. The data contain information about a customer's ethnicity, a target variable that indicates defaulting behavior, and other financial information about the customer. Suppose that the auditor has fitted the auditee's predictive model to this benchmark data set to obtain the confusion matrix in Table 4. They calculate the false positive rate p_i for each ethnicity i as the number of false positives divided by the number of false positives plus the number of true negatives.

Table 4.4: The confusion matrix obtained from the auditee's classification algorithm.

		Predicted		
Ethnicity	Observed	Defaulted	Paid	Total
Asian	Paid	97	826	923
	Defaulted	11	66	77
	Total	108	892	1000
African	Paid	167	1678	1845
	Defaulted	12	143	155
	Total	179	1821	2000
Hispanic	Paid	195	1648	1843
	Defaulted	17	140	157
	Total	212	1788	2000
Caucasian	Paid	477	4137	4614
	Defaulted	55	331	386
	Total	532	4468	5000
Total	Paid	936	8289	9225
	Defaulted	95	680	775
	Total	1031	8969	10000

4.4.4.2 Frequentist analysis

In the frequentist analysis, the null hypothesis of independence $\mathcal{H}_0$: $p_1 = p_2 = p_3 = p_4$ will be tested using a Chi-squared test and assessed by means of the p-value.

The false positive rates for the categories p_1, p_2, p_3 and p_4 are 0.1051, 0.0911, 0.1058, and 0.1033. Using a Chi-squared test ($X^2 = 3.126$, $df = 3$), the p-value is 0.373. The interpretation of this p-value is: Assuming that the false positive rate is equal across all ethnic groups, there is a 37.3 percent probability of observing these (or more extremely deviating) false positive rates.

4.4.4.3 Bayesian analysis

In the Bayesian analysis, the null hypothesis of independence will be tested against the alternative hypothesis that the false positive rates are dependent on the subgroup (Gunel and Dickey, 1974; Jamil et al., 2017). The prior distribution for the alternative hypothesis is a Dirichlet$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ distribution with all α parameters set to 1.

The Bayes factor in favor of $\mathcal{H}_0$ is $BF_{01} = 11077.956$, which implies that the data are 11078 times more likely to have occurred under the hypothesis that the false positive rates are equal across ethnic groups than under the hypothesis that they are not. Using this Bayes factor, the auditor can quantify evidence in favor of the null hypothesis and support the statement that the false positive rates are equal across groups.

4.4.4.4 Comparison of frequentist and Bayesian conclusions

The p-value of 0.373 leads the auditor to not reject the null hypothesis $\mathcal{H}_0$. However, based on this p-value, the auditor cannot say the data shows evidence that supports the alternative hypothesis: the false positive rates are equal across ethnicities. In contrast to the p-value, the auditor can use the Bayes factor of $BF_{01} = 11077.956$ to substantiate the conclusion that the data contain extreme evidence in favor of equal false positive rates across ethnicities.

4.5 Concluding comments

From the perspective of an auditor, audit evidence plays a crucial role in providing an opinion about whether the assertions presented by the auditee's management in the financial statements are credible. However, the frequentist method (NHST) by which statistical audit evidence is currently often quantified in audits has raised legitimate concerns over the years. In this chapter, we have emphasized the fact that a frequentist hypothesis test does not produce the type of evidence that the audit standards demand, and that the p-value does not fit well with the nature of audit questions. We have shown that a Bayesian hypothesis test can produce a more fitting type of evidence for the auditor's conclusions about the financial statements, and that it does not suffer from the same limitations as the p-value when it comes to effectiveness and efficiency. Because the Bayes factor can quantify evidence in both directions, the Bayesian approach to audit evidence is more in line with the audit standards than that of a frequentist hypothesis test. We therefore propose the Bayes factor as an addition to the auditor's statistical toolbox. Since the auditing standards explicitly call for evidence that can support or contradict the auditor's conclusions, we expect that the Bayes factor will enhance the way that auditors are able to quantify and evaluate statistical evidence from a sample.

Moreover, Bayesian inference provides auditors the tools to aggregate audit evidence, and therefore to statistically accumulate audit evidence over the course of an audit. This makes the Bayes factor a good fit for today's audit practice because it can facilitate the growing use of complex data analytics by the auditor and the auditee. As data will become more complex, and statistical analyses will become more prevalent, the auditor will require an intuitive framework to integrate, quantify, and interpret the information from these procedures. This will be especially the case if they are to meet the constant demand for a more efficient audit. Since the Bayesian framework provides the flexibility to incorporate many types of prior information into the statistical analysis, we believe that it will be more useful for the auditor in the long term than the current frequentist methods.

Despite our arguments in favor of Bayesian hypothesis testing using the Bayes factor, it is not always practical to use this approach in favor of a frequentist hypothesis test. As discussed in Section 4.2, Bayes' theorem utilizes prior information in the form of the prior probabilities and the prior distribution on the parameters to perform inference. This means that, to get to the Bayes factor in practice, the auditor will have to think about how they incorporate their existing information into the statistical model. This has as a consequence that sometimes, a frequentist hypothesis test can be more beneficial to an auditor if translating prior information into Bayesian prior distributions is difficult, expensive or time-consuming. Although the Bayesian approach comes with advantages such as being able to quantify evidence in favor and against the auditor's conclusions or the ability to engage in sequential testing without penalty, the auditor needs to decide whether they think that the benefits of the Bayesian approach outweigh the costs of justifying that approach.

The sensitivity of the Bayes factor to the prior distribution is an avenue for further research in this area. For Bayesian hypothesis testing, default prior distributions exist and have been evaluated in a wide variety of designs and settings (Rouder et al., 2009; Wetzels and Wagenmakers, 2012; Wetzels et al., 2012). However, no default prior distribution exists or has been evaluated specifically in the context of audit sampling. Moreover, it remains to be investigated how auditors use and interpret Bayesian evidence in practice, and if it increases the ease of interpretation of statistical results for auditors.

The examples shown in this chapter show a selection of data-rich audit scenarios that the Bayes factor can be applied in, but in principle any statistical analysis can be performed in a Bayesian fashion. Most Bayesian analyses are easily accessible in a standardized format through open-source software packages such as R (R Core Team, 2022) or graphical user interfaces such as JASP (JASP Team, 2022). We have performed all statistical analyses in this chapter using JASP and have included reproducible examples in Appendix 4.B. Our proposition for a way forward is that next to their frequentist analyses auditors perform Bayesian equivalents of these analyses to become acquainted with these techniques, and to be able to compare the two measures of evidence (p-values and Bayes factors) in practice.

To conclude, we advocate the use of Bayesian inference in the audit because it fits well with the goals of the auditor. First, the Bayes factor embodies the audit standards' description of audit evidence and provides the auditor with a measure of statistical evidence that can support or contradict their conclusions. Second, the theoretical foundations underlying the Bayesian framework have long been argued to be beneficial for the audit, since they enable the auditor to quantify and aggregate evidence over the audit using the prior and posterior probabilities. In sum, Bayesian inference provides a fitting answer to the problems that today's auditors face.

4.A Two approaches to NHST in audit sampling

There are two approaches to frequentist null hypothesis significance testing (NHST) in audit sampling: the positive approach and the negative approach (Roberts, 1975). The key difference between the positive and the negative approach comes down to how the hypothesis of tolerable misstatement is defined, that is, whether the hypothesis of tolerable misstatement includes the value of the performance materiality or not. It can be demonstrated that the two approaches are equivalent (Roberts, 1978, p. 45).

In the positive approach, the value of the performance materiality is seen as the maximum tolerable misstatement. Hence, in this approach the hypothesis of tolerable misstatement includes the value of the performance materiality. This null hypothesis is encapsulated by the statistical scenario where $\theta \leq 0.03$, whereas the alternative hypothesis of intolerable misstatement is encapsulated by the scenario where $\theta > 0.03$. Because in NHST the auditor assumes that the null hypothesis is true, the positive approach implies a philosophy of auditing where the auditor accepts the population as not materially misstated unless there is evidence to the contrary. The sampling risks α and β are those defined in the chapter: α is the risk of incorrectly deciding that the population contains material misstatement when in fact it does not (i.e., a Type-I error) and β is the risk of incorrectly deciding that the population does not contain material misstatement when in fact it does (i.e., a Type-II error).

In the negative approach, the value of the performance materiality is seen as the minimum intolerable misstatement. Hence, in this approach the hypothesis of tolerable misstatement does not include the value of the performance materiality. This alternative hypothesis is encapsulated by the statistical scenario where $\theta < 0.03$, whereas the null hypothesis of intolerable misstatement is encapsulated by the scenario where $\theta \geq 0.03$. Because in NHST the auditor assumes that the null hypothesis is true, the negative approach implies a philosophy of auditing where the auditor rejects the population as being materially misstated unless there is evidence to the contrary. As is the case for the null and alternative hypotheses, the interpretation of the sampling risks α and β is reversed in this approach: α is the risk of incorrectly deciding that the population does not contain material misstatement when in fact it does (i.e., a Type-I error) and β is the risk of incorrectly deciding that the population contains material misstatement when in fact it does not (i.e., a Type-II error).

In the chapter accompanying this appendix, we have focused on the positive approach to NHST in audit sampling for two reasons. First, the text of the auditing standards directly links to the positive approach. For example, the definition of performance materiality is given as: "Performance materiality is determined to reduce to an appropriately low level the probability that the aggregate of uncorrected and undetected misstatements in the financial statements exceeds materiality for the financial statements as a whole" (ISA 320, paragraph A.13, International Auditing and Assurance Standards Board (IAASB), 2018). Furthermore, the definition of tolerable misstatement (i.e., the application of performance materiality to a particular sampling procedure) is given as: "A monetary amount set by the auditor in respect of which the auditor seeks to obtain an appropriate level of assurance that the monetary amount set by the auditor is not exceeded by the actual misstatement in the population" (ISA 530, paragraph 5i, International Auditing and Assurance Standards Board (IAASB), 2018; AU-C 530, paragraph 5, American Institute of Certified Public Accountants (AICPA), 2021). These texts highlight the intention of the auditor to assert whether the actual misstatement (i.e., θ) does not exceed (i.e., $\leq$) the performance materiality or tolerable misstatement (i.e., 0.03). The fact that θ must not exceed the performance materiality implies that the value of the performance materiality is included in the hypothesis of tolerable misstatement (i.e., $\theta \leq 0.03$). A second argument for the positive approach is that this approach is prominent in the auditing literature, see for example Elliott and Rogers (1972), Johnstone (1994), Martel-Escobar et al. (2018), and Edmonds et al. (2019).

However, there are also several arguments to be made in favor of the negative approach. First, in hypothesis testing, the probability of incorrectly rejecting the null hypothesis is generally considered to be more important and is therefore referred to as primary risk, Type-I error, α, or audit risk. This preference for the α risk as the conceptualization of audit risk is consistent with the negative approach described above. In an audit context, the negative approach can be seen as a form of 'guilty-until-proven-innocent', where 'guilty' means: the financial statements contain material misstatement. Thus, in the negative approach the auditor audits the financial statements from a conservative point of view. A second argument for the negative approach is that the standard sample size tables in American Institute of Certified Public Accountants (AICPA) (2019, Appendix A and Appendix C) can only be replicated using this approach. Hence, because practitioners may be more familiar with the negative approach to NHST, we describe the calculations underlying the running example from the chapter in the remainder of this appendix.

To determine when $\mathcal{H}_0$: $\theta \geq 0.03$ should be rejected for a given sample size, the auditor must calculate the maximum number of invalid signatures that can be observed while the risk of incorrectly rejecting the null hypothesis is still sufficiently low. Suppose that the auditor has determined α—the risk of incorrectly deciding that the population does not contain material misstatement when in fact it does—to be five percent. In this case, the rule for rejection of $\mathcal{H}_0$ is $k = 0$ because if it is true that $\theta = 0.03$, then the probability of finding 0 invalid signatures in the sample of 99 items equals 4.99 percent (see Equation 4.A.1). Using a higher threshold, as in $k = 1$, would violate this sampling risk since the probability of finding 1 or less invalid signatures is 19.91 percent. As described in the chapter, this procedure for rejecting $\mathcal{H}_0$ can also be regarded as rejecting the null hypothesis if the p-value associated with the observed value of k is less than or equal to the significance level α.

Let's revisit the running example in which a sample of 99 items was selected and no misstatements were found. In this case, the p-value is the probability of finding $k = 0$ invalid signatures in the sample, given that the temporary contracts contain three percent misstatement, and equals $p = 0.049$ (Equation 4.A.1). Since the calculated p-value of 0.049 is lower than the significance level $\alpha = 0.05$, the auditor can reject the null hypothesis.

$$p = \binom{99}{0} 0.03^0 (1 - 0.03)^{99-0} = 0.049 \tag{4.A.1}$$

The auditor can also calculate the sampling risk β—the risk of deciding that the population contains material misstatement when in fact it does not. Like in the positive approach, to calculate β for an alternative point hypothesis about the population misstatement, the auditor needs to make an assumption about θ. Suppose the auditor assumes that the population misstatement is equal to $\theta = 0.02$, which is slightly lower than the performance materiality. When assuming the truth of this hypothesis, the risk of failing to reject the null hypothesis $\mathcal{H}_0$: $\theta \geq 0.03$ is the probability of finding an outcome that would yield a p-value above five percent (e.g., $k = 1$ misstatements would give $p > .05$ and thus lead the auditor to not reject $\mathcal{H}_0$). Hence, the sampling risk β can be calculated as the probability of finding $k = 1$ to $k = 99$ invalid signatures in the sample under the Binomial($k \mid n = 99, \theta = 0.02$) distribution and equals $\beta = 0.86$.

From a Bayesian point of view, the two hypotheses $\mathcal{H}_0$: $\theta \geq 0.03$ and $\mathcal{H}_1$: $\theta < 0.03$ are defined as the range of the prior distribution that corresponds to the hypotheses' restrictions with respect to θ. This means that the prior probability for the hypothesis $\mathcal{H}_0$ corresponds to the total probability under the prior distribution on θ in the range [0.03; 1]. Vice versa, the prior probability for the hypothesis $\mathcal{H}_1$ corresponds to the total probability under the prior distribution on θ in the range [0; 0.03). Like in the chapter, the auditor chooses to specify a Beta(0, 34) prior distribution for θ.

After being updated by the sample of $n = 99$ items of which $k = 0$ contain an error, the posterior distribution is the Beta($1 + 0 = 1$, $34 + 99 = 133$) distribution. The posterior odds in favor of $\mathcal{H}_1$ induced by the posterior distribution are therefore $\frac{0.983}{0.017} = 57.824$. The posterior probability $p(\mathcal{H}_1 \mid y) = 0.983$ implies that there is a 98.3 percent probability that the population does not contain misstatements equal to or above the performance materiality. This means that, when accepting $\mathcal{H}_1$, there is a 1.7 percent probability that the auditor incorrectly judges that the population is free of material misstatement. This probability is sufficiently low to find the statement in the alternative hypothesis credible. Vice versa, this also implies that there is a 98.3 percent probability that the auditor correctly judges that the population is not materially misstated.

Because we know the prior odds and the posterior odds, we can again calculate the Bayes factor by dividing the two: $BF_{10} = \frac{57.824}{1.817} \approx 31$. This Bayes factor implies that the data are about 31 times more likely to occur under $\mathcal{H}_1$ than under $\mathcal{H}_0$.

4.B Bayes factor calculations

4.B.1 Example 1: Evaluating an audit sample

4.B.1.1 Technical details

The observed misstatements in the sample, k, are assumed to be binomially distributed (Equation 4.B.1).

$$k \sim \text{Binomial}(n, \theta) \tag{4.B.1}$$

The prior distribution for the misstatement parameter θ in $\mathcal{H}_0$: $\theta < 0.03$ is a conjugate Beta(1, 34) distribution truncated to the interval [0; 0.03] and the prior distribution for the parameter θ in $\mathcal{H}_1$: $\theta > 0.03$ is a conjugate Beta(1, 34) distribution truncated to the interval [0.03; 1].

The evidence in the data for each hypothesis (i.e., the marginal likelihood) is computed by multiplying the likelihood of the data ($n = 99$, $k = 0$) for each value of θ with the probability of each value of θ under the prior distribution, and then integrating this distribution over θ. The Bayes factor is the ratio of the computed marginal likelihoods. R code for computing these marginal likelihoods and the corresponding Bayes factor is given below.

```r
###########################################
# Example 1 - Evaluating an audit sample #
###########################################

library(jfa)
library(truncdist)

# Data
n           <- 99   # Sample size
k           <- 0    # Number of misstatements
materiality <- 0.03 # Performance materiality

# Prior parameters
alpha <- 1
beta  <- 34

# Multiply the likelihood with the prior
integrandMin <- function(theta) {
  likelihood <- dbinom(x = k, size = n, prob = theta)
  prior      <- dtrunc(x = theta, spec = "beta",
                       a = 0, b = materiality,
                       shape1 = alpha, shape2 = beta)
  return(likelihood * prior)
}

integrandPlus <- function(theta) {
  likelihood <- dbinom(x = k, size = n, prob = theta)
  prior      <- dtrunc(x = theta, spec = "beta",
```

```
                        a = materiality, b = 1,
                        shape1 = alpha, shape2 = beta)
  return(likelihood * prior)
}

# Integrate over theta
margLikMin <- integrate(f = integrandMin,
                         lower = 0, upper = materiality)$value
margLikPlus <- integrate(f = integrandPlus,
                          lower = materiality, upper = 1)$value

# Compute the Bayes factor
BF01 <- margLikMin / margLikPlus
BF01
# [1] 31.0765

# Alternatively, the "jfa" package can be used
evaluation(materiality = 0.03, n = 99, x = 0,
           prior = auditPrior(method = "param", likelihood = "binomial",
                              alpha = 1, beta = 34))

#   Bayesian Audit Sample Evaluation
#
# data:  0 and 99
# number of errors = 0, number of samples = 99, taint = 0, BF10 = 31.076
# alternative hypothesis: true misstatement rate is less than 0.03
# 95 percent credible interval:
#  0.00000000 0.02227252
# estimate:
#  0
# estimates obtained via method "binomial" + "prior"
```

4.B.1.2 Practical evaluations

This example can be reproduced in JASP via the Bayesian evaluation analysis in the Audit module (Derks et al., 2021b). The interface of the Bayesian evaluation analysis in JASP and the options required to reproduce this example are shown in the figure at the bottom of this section. An example JASP file containing the analysis and results is available in the online appendix at `https://osf.io/wtn9g/`.

The Bayesian evaluation analysis allows the auditor to evaluate their sample on the basis of summary statistics only, and it is therefore not required to load a data set. After opening the analysis, the auditor can check the option 'Performance materiality' and specify the performance materiality θ_{max} as 3 percent. In the 'Prior' section, the auditor can manually set the parameter of the prior distribution to 34. Finally, the auditor can fill in the sample outcomes $n = 99$ and $k = 0$ as the 'Sample size' and the 'Number of errors' respectively. The results of the statistical analysis are automatically computed.

The first table in the output shows the sample size and number of found misstatements in the sample. By default, the total tainting (i.e., proportional error) in the sample is shown alongside the most likely error in the population (i.e., the mode of the posterior distribution), in this case 0. The fifth column in the table displays the 95 percent upper credible bound (i.e., the 95$^{\text{th}}$ percentile of the posterior distribution), in this case 0.022. The table also shows the Bayes factor in favor of the hypothesis of tolerable misstatement. In this example, the Bayes factor in favor of tolerable misstatement is $BF_{-+} = 31.076$. In the 'Tables' section the auditor is able to request a table containing descriptive information of the prior and posterior distribution by clicking 'Prior and posterior'. Moreover, a figure displaying the prior and posterior distribution can be displayed by clicking the 'Prior and posterior' option in the 'Plots' section.

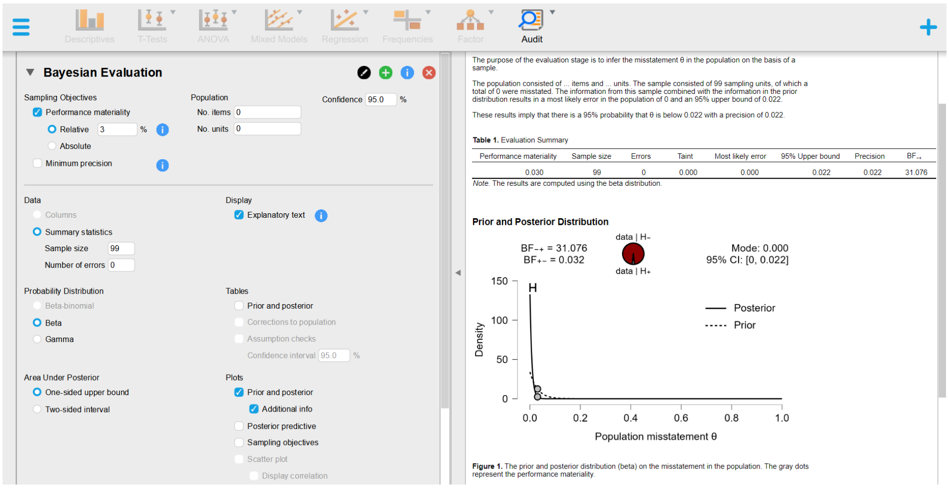


Figure 4.3: The interface of the Bayesian evaluation analysis in JASP and the options required to reproduce the audit sampling example.

4.B.2 Example 2: Assessing Benford's law

4.B.2.1 Technical details

The null hypothesis $\mathcal{H}_0$ implies an expected probability for all $D = 9$ digits $d \in \{1, \ldots, 9\}$. Specifically, the expected category proportion for digit d is:

$$
\begin{aligned}
c &= \log_{10}(1 + \frac{1}{d}) = \theta_1, \theta_2, \ldots, \theta_9 \qquad (4.B.2)\\
&= 0.301, 0.176, 0.125, 0.097, 0.079, 0.067, 0.058, 0.051, 0.046
\end{aligned}
$$

The prior distribution for $\mathcal{H}_1$ is a Dirichlet(α_1, α_2, …, α_9) distribution with all α parameters set to 1. After seeing the observed counts x_d of each digit d, the posterior distribution for $\mathcal{H}_1$ is a Dirichlet($\alpha_1 + x_1$, $\alpha_2 + x_2$, ..., $\alpha_9 + x_9$) distribution.

The Bayes factor in favor of $\mathcal{H}_0$ can be obtained by calculating the Savage-Dickey density ratio (Wagenmakers et al., 2010), that is, the height of the posterior distribution $p(\theta = c \mid x, \mathcal{H}_1)$ at the point of interest c divided by the height of the prior distribution $p(\theta = c \mid \mathcal{H}_1)$ at this point, see Equation 4.B.3 (Sarafoglou et al., 2021). R code for computing this ratio is given below.

$$BF_{01} = \frac{p(\theta = c \mid x, \mathcal{H}_1)}{p(\theta = c \mid \mathcal{H}_1)} = \frac{\frac{\Gamma(\sum_{d=1}^{D} \alpha_d + x_d)}{\prod_{d=1}^{D} \Gamma(\sum_{d=1}^{D} \alpha_d)} \prod_{d=1}^{D} \theta_d^{x_d + \alpha_d - 1}}{\frac{\Gamma(\sum_{d=1}^{D} \alpha_d)}{\prod_{d=1}^{D} \Gamma(\alpha_d)} \prod_{d=1}^{D} \theta_d^{\alpha_d - 1}} = \frac{6899678}{1} \tag{4.B.3}$$

```r
#########################################
# Example 2 - Assessing Benford's law #
#########################################

library(digitTests)

# Data
N        <- 772
digits   <- 1:9
p        <- log10(1 + 1 / digits)
expected <- p * N
observed <- c(231, 124, 97, 70, 64, 54, 40, 54, 38)

# Prior parameters
alpha <- 1

# Logarithm of Eq. 4.A.3
lbeta.xa <- sum(lgamma(alpha + observed)) - lgamma(sum(alpha + observed))
lbeta.a <- sum(lgamma(rep(alpha, 9))) - lgamma(sum(rep(alpha, 9)))

# Compute the Bayes factor
logBF10 <- (lbeta.xa - lbeta.a) + (0 - sum(observed * log(p)))
BF01    <- 1 / exp(logBF10)
BF01
# [1] 6899678

# Alternatively, the "digitTests" package can be used
distr.btest(x = rep(digits, observed), BF10 = FALSE)

#   Digit distribution test
#
# data: rep(digits, observed)
# n = 772, BF01 = 6899678
# alternative hypothesis: leading digit(s) are not distributed according
# to the benford distribution.
```

4.B.2.2 Practical evaluations

This example can be reproduced in JASP via the Benford's law analysis in the Audit module. The interface of the Benford's law analysis in JASP and the options required to reproduce this example are shown in the figure at the bottom of this section. An example JASP file containing the data, analysis, and results is available in the online appendix at `https://osf.io/wtn9g/`.

After loading the data into JASP and opening the Benford's law analysis, the auditor can drag the variable that contains the numbers to be inspected (in the example data this variable is named 'value') to the 'Variable' box in the interface. The leading digits are automatically extracted from the data, and the results for a test against Benford's law are computed.

By default, the first table in the output shows the sample size ($n = 772$), the Chi-squared statistic ($X^2 = 7.652$), the degrees of freedom ($df = 8$), the p-value ($p = 0.468$), and the Bayes factor in favor of the alternative hypothesis $\mathcal{H}_1$. However, the Bayes factor in favor of the null hypothesis $\mathcal{H}_0$ can be requested by changing the option 'Bayes Factor' in the interface from 'BF_{10}' to 'BF_{01}'. This results in the Bayes factor in favor of $\mathcal{H}_0$ shown in the chapter ($BF_{01} = 6899678$). The second table in the output contains descriptive statistics regarding the leading digits in the population. The table shows the observed percentage of each leading digit versus the expected percentage under Benford's law. For clarity, a figure displaying these observed and expected percentages can be created by clicking the option 'Observed vs. expected'.

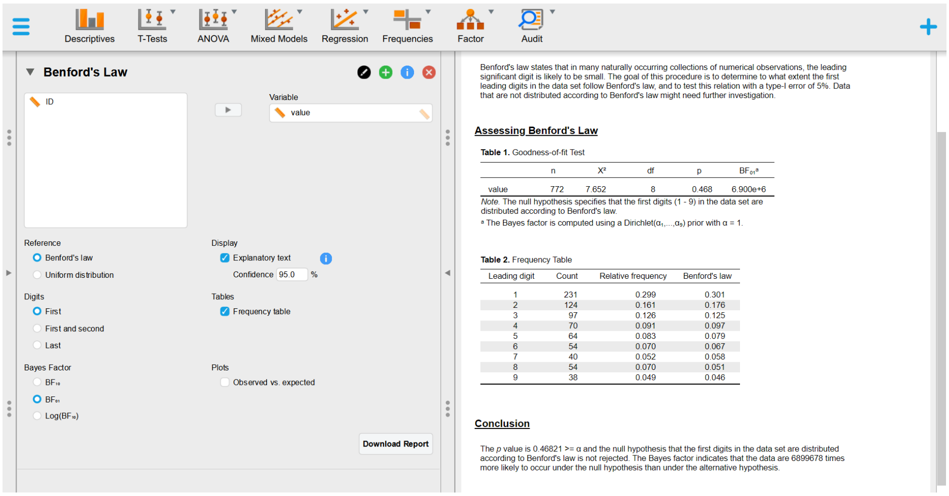


Figure 4.4: The interface of the Benford's law analysis in JASP and the options required to reproduce the Benford's law example.

4.B.3 Example 3: Uncovering seasonal patterns

4.B.3.1 Technical details

The model representing the null hypothesis is $\mathcal{M}_0$: $y = \mu + \epsilon$. The model representing the alternative hypothesis is $\mathcal{M}_1$: $y = \mu + \sigma X_a \alpha + \epsilon$, where α is a vector of a effects (one for each season), X_a is the design matrix with dimensionality $N \times \alpha$ reflecting which season each observation falls into, and σ is a scaling factor for the effects. A Jeffreys's prior is placed on the parameters μ and σ^2, see Equation 4.B.4.

$$\pi(\mu, \sigma^2) = \frac{1}{\sigma^2} \tag{4.B.4}$$

Furthermore, the effects are modeled as in Equation 4.B.5.

$$\alpha \,|\, g \sim \text{Normal}(0, gI_a), \tag{4.B.5}$$

In Equation 4.B.5, g is the variance of the effects and I is the identity matrix of size a. The prior for g is specified as an inverse Chi-squared distribution with shape parameter $v = 1$ and scale parameter $\tau^2 = 0.25$, see Equation 4.B.6.

$$g \sim \text{Inverse-}X^2(1, 0.25) \tag{4.B.6}$$

In a balanced design ($n_1 = n_2 = n_3 = n_4$), the Bayes factor between $\mathcal{M}_1$ and $\mathcal{M}_0$ can be calculated using Equation 4.B.7 derived by Rouder et al. (2012). R code for computing this integral is given below.

$$BF_{10} = \int_0^\infty (1 + gn)^{-(a-1)/2} \times (1 - \frac{R^2}{(1 + gn)/(gn)})^{-(N-1)/2} \pi(g) dg = 33.92343 \tag{4.B.7}$$

```
#############################################
# Example 3 - Uncovering seasonal patterns #
#############################################

library(BayesFactor)
library(extraDistr)

# Data
N        <- 48                # Total number of observations
a        <- 4                 # Number of seasons
n        <- N / a             # Number of observations per season
rsquared <- 0.30429656366968499 # R^2 of the model M1
rscale   <- 0.5 # Sqrt of scale parameter for prior distribution on g

# Integral from Equation 12 in Rouder et al. (2012)
```

```
integrand <- function(g) {
  prior <- function(x, rscale) {
    extraDistr::dinvchisq(x, nu = 1, tau = rscale^2)
  }
  p <- (1 + g*n)^(-(a-1)/2) *
          (1 - (rsquared / ((1 + g*n)/(g*n))))^(-(N-1)/2)
  prior <- prior(g, rscale)
  return(p * prior)
}

# Compute the Bayes factor from the integral
BF10 <- integrate(integrand, lower = 0, upper = Inf)$value
BF10
# [1] 33.92343

# Alternatively, the "BayesFactor" package can be used
data <- read.csv("https://osf.io/we6hu/download",
                  stringsAsFactors = TRUE)
logBF10 <- anovaBF(formula = sales ~ Season, data = data,
                    rscaleFixed = rscale)@bayesFactor$bf
BF10   <- exp(logBF10)
BF10
# [1] 33.92343
```

4.B.3.2 Practical evaluations

This example can be reproduced in JASP via the Bayesian ANOVA analysis in the ANOVA module. The interface of the Bayesian ANOVA analysis in JASP and the options required to reproduce this example are shown in the figure at the bottom of this section. An example JASP file containing the data, analysis, and results is available in the online appendix at `https://osf.io/wtn9g/`.

After loading the data into JASP and opening this analysis, the auditor can drag the outcome variable 'sales' to the 'Dependent Variable' box. Next, the auditor can drag the grouping variable 'Season' to the 'Fixed Factors' box. The results of the Bayesian ANOVA are automatically computed. Under the section 'Post Hoc Tests', the auditor can drag the 'Season' variable to the input box to compute Bayes factors for a difference between each pair of seasons.

The first table in the output shows the comparison of the two models under consideration. The table displays the prior probability for each model $p(M)$, the posterior probability for each model $p(M \mid \text{data})$, and Bayes factor for the alternative model versus the null model $BF_{10} = 33.923$. The post hoc comparisons table shows the Bayes factors ($BF_{10,U}$) for the comparisons between each pair of seasons.

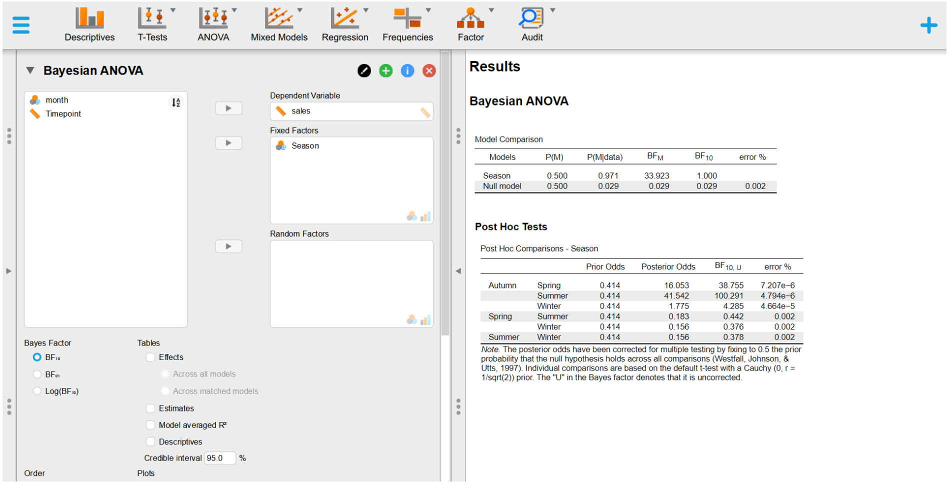


Figure 4.5: The interface of the Bayesian ANOVA in JASP and the options required to reproduce the seasonal patterns example.

4.B.4 Example 4: Determining algorithmic bias

4.B.4.1 Technical details

In this example, we follow a joint multinomial sampling scheme (Jamil et al., 2017). The cell counts, y_{ij}, for the i^{th} row and j^{th} column of the contingency table with dimensions $R \times C$ are assumed to be jointly multinomially distributed, see Equation 4.B.8.

$$(y_{ij}, \ldots, y_{RC}) \sim \text{Multinomial}(y_{ij}, p_{ij}, \ldots, p_{RC}) \tag{4.B.8}$$

The prior distribution on the multinomial parameters, p, is a conjugate Dirichlet distribution with $\alpha = 1$, see Equation 4.B.9.

$$p_{ij}, \ldots, p_{RC} \sim \text{Dirichlet}(\alpha_{ij}, \ldots, \alpha_{RC}) \tag{4.B.9}$$

The calculations for the Bayes factor in favor of independence under the joint multinomial sampling scheme are discussed in Jamil et al. (2017), see Equation 4.B.10. R code for computing this Bayes factor is given below.

$$BF_{01} = \frac{D(y_{*\cdot} + \xi_{*\cdot})}{D(\xi_{*\cdot})} \times \frac{D(y_{\cdot *} + \xi_{\cdot *})}{D(\xi_{\cdot *})} \times \frac{D(\alpha_{**})}{D(y_{**} + \alpha_{**})} = 11077.96 \tag{4.B.10}$$

```
############################################
# Example 4 - Determining algorithmic bias #
############################################
```

```
library(BayesFactor)
library(LearnBayes)

# Data
y <- matrix(c(97, 195, 167, 477,
              826, 1648, 1678, 4137), nrow = 2, byrow = TRUE)

# Logarithm of Equation 8 in Jamil et al. (2017)
C             <- ncol(y)
R             <- nrow(y)

ystardot      <- rowSums(y)
ydotstar      <- colSums(y)
ydotdot       <- sum(y)

alphastarstar <- matrix(1, nrow = 2, ncol = 4)
alphastardot  <- rowSums(alphastarstar)
alphadotstar  <- colSums(alphastarstar)
alphadotdot   <- sum(alphastarstar)

xistardot     <- alphastardot - (C - 1)
xidotstar     <- alphadotstar - (R - 1)
xidotdot      <- alphadotdot - (R - 1) * (C - 1)

ldirichlet <- function(a) {
  sum(lgamma(a)) - lgamma(sum(a))
}

part1 <- ldirichlet(ystardot + xistardot) - ldirichlet(xistardot)
part2 <- ldirichlet(ydotstar + xidotstar) - ldirichlet(xidotstar)
part3 <- ldirichlet(alphastarstar) - ldirichlet(y + alphastarstar)

# Compute the Bayes factor
logBF01 <- part1 + part2 + part3
BF01 <- exp(logBF01)
BF01
# [1] 11077.96

# Alternatively, the "BayesFactor" package can be used
logBF10 <- contingencyTableBF(y, sampleType = "jointMulti",
                              priorConcentration = 1)@bayesFactor$bf
BF01 <- 1 / exp(logBF10)
BF01
# [1] 11077.96

# Another alternative is the "LearnBayes" package
BF10 <- ctable(y, alphastarstar)
BF01 <- 1 / BF10
BF01
```

```
# [1] 11077.96
```

4.B.4.2 Practical evaluations

This example can be reproduced in JASP via the Bayesian contingency tables analysis in the Frequencies module. The interface of the Bayesian contingency tables analysis in JASP and the options required to reproduce this example are shown in the figure at the bottom of this section. An example JASP file containing the data, analysis, and results is available in the online appendix at `https://osf.io/wtn9g/`.

After having loaded the data into JASP, the auditor is able to filter the data such that all customers that are predicted to default (in this data set TARGET = 1) are removed from the data set. This way only the customers that are predicted to have paid their loan remain in the data. After all, only these customers are potentially subject to a false positive (incorrectly predicted to default). This filter can be set by clicking the filter icon in the top left corner of the data and entering the input shown in Figure 4.6.

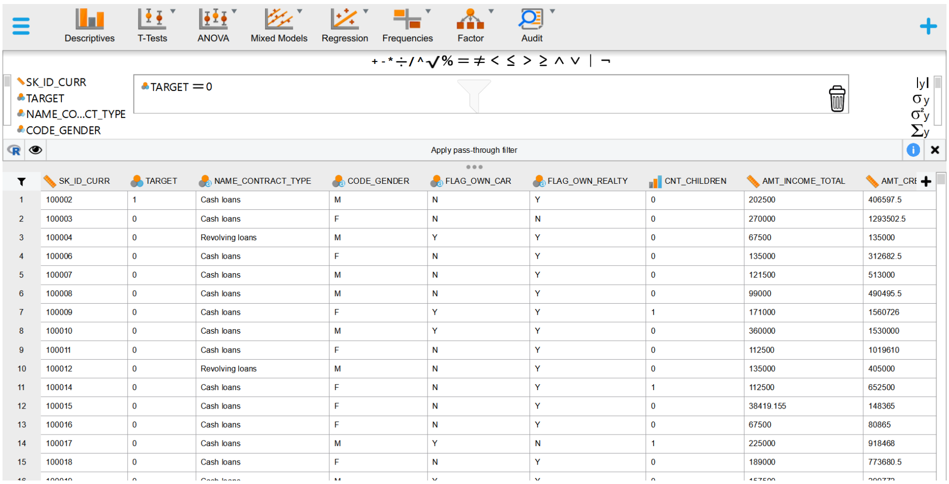


Figure 4.6: The required input filter to remove the cases with TARGET = 1 in this example.

After opening the Bayesian contingency tables analysis the auditor can drag the predicted values in the variable 'Predicted' to the box for 'Rows', and the variable 'Race' containing the person's ethnicity to the box for 'Columns'. Under the section 'Statistics' the auditor can set the 'Sample' option to 'Joint multinomial'. The results of the statistical analysis are automatically computed.

The first table in the output displays a two-by-four contingency table that contains the observed counts in the data. By default, the second table in the output shows the Bayes factor for a joint multinomial sampling plan, testing whether all categories displayed in the rows and columns are independent of each other. The

Bayes factor in favor of the null hypothesis $\mathcal{H}_0$ can be requested by changing the option ‘Bayes Factor’ in the interface from ‘BF_{10}’ to ‘BF_{01}’. This results in a Bayes factor in favor of $\mathcal{H}_0$ shown in the chapter ($BF_{01} = 11077.956$).

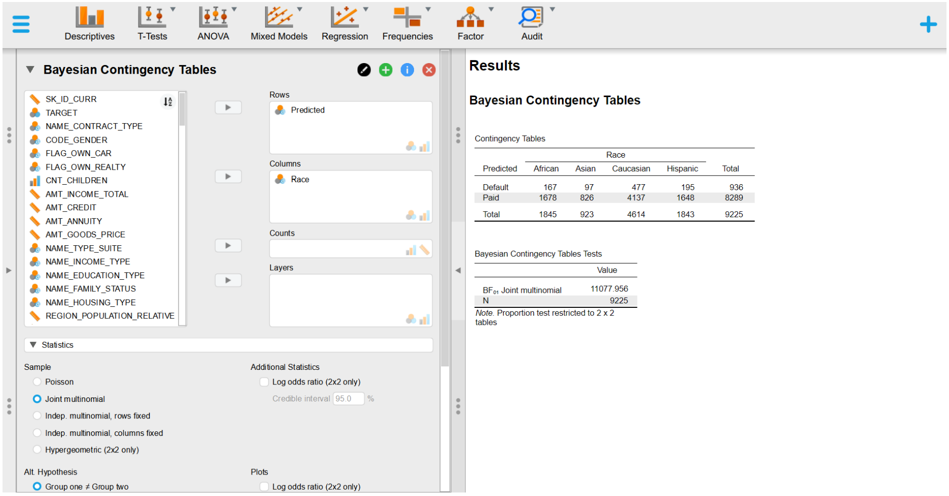


Figure 4.7: The interface of the Bayesian contingency tables analysis in JASP and the options required to reproduce the algorithmic bias example.

Chapter 5

An Impartial Bayesian Hypothesis Test for Audit Sampling

Abstract

Auditors who perform audit sampling are often interested in obtaining evidence for or against the hypothesis that the misstatement in a population of items is lower than a critical limit, the so-called performance materiality. Here, we propose to perform this hypothesis test using a Bayesian approach that involves the use of an impartial prior distribution, assigning equal prior probabilities to the competing interval hypotheses. Firstly, we argue that the impartial prior distribution is sensible for auditors because it is easy to justify, interpret, and explain. Secondly, we show that Bayes factors computed using this prior distribution have desirable statistical properties. Finally, we compare these Bayes factors with traditional p-values in an audit sampling context and elaborate on the merits of the impartial Bayesian hypothesis test.

Keywords: Audit sampling, Bayes factor, impartial, evidence, prior distribution.

5.1 Introduction

An audit is a final inspection of an organization's financial statements with the goal of providing stakeholders of the organization with reasonable assurance about the accuracy and completeness of those statements. In many countries, audits are mandatory for financial institutions, governments, and listed organizations. Since business decisions are taken every day based on the auditor's opinion, auditors need to formulate their opinions based on audit evidence. Hence, the objective of

This chapter is under review for publication as Derks, K., de Swart, J., Wagenmakers, E.–J., & Wetzels, R. (2022). An impartial Bayesian hypothesis test for audit sampling. doi: `https://doi.org/10.31234/osf.io/8nf3e`

an audit is to obtain evidence that can serve as a basis for an opinion about the financial situation of an organization.

Despite the increasing availability of data and the associated advent of data-driven methods, tests of details remain firmly entrenched as the dominant accounting practice for obtaining audit evidence. Traditionally, a large amount of audit evidence was obtained from tests of details (Power, 1992), but in recent decades the growing availability of data has shifted the focus towards obtaining audit evidence from other analytical procedures (e.g., Appelbaum et al., 2017; Boersma et al., 2020). However, adaptation of these ideas and technologies is only slowly permeating the audit practice (Gepp et al., 2018), and many audit firms still use some form of test of details to obtain audit evidence (Christensen et al., 2015). Because inspecting all available data is time consuming and expensive, tests of details often take place through audit sampling.

Audit sampling is defined as the application of audit procedures to a subset of a population of items with the goal of performing inference on an unknown characteristic of that population (ISA 530; International Auditing and Assurance Standards Board (IAASB), 2018). For example, in a financial audit an auditor may want to investigate a claim that a population of trade receivables contains misstatements that do not exceed a certain critical limit, the so-called performance materiality. To investigate this claim and make a statement about the true misstatement in the trade receivables account the auditor can inspect a representative sample from this population of items. When the sample is planned, selected, and evaluated using probability theory, this practice is referred to as statistical audit sampling (ISA 530; International Auditing and Assurance Standards Board (IAASB), 2018). In contrast to non–statistical sampling, statistical sampling is supposed to enable the auditor to obtain statistical evidence to support or contradict the claim that the misstatement in the population is below the performance materiality.

To obtain statistical evidence for or against this claim the auditor can engage in statistical hypothesis testing. Currently, the dominant framework of statistical hypothesis testing in the audit practice is p-value null-hypothesis significance testing (NHST). Whereas NHST is commonly used, it cannot provide the auditor with direct evidence to support their statements (Wagenmakers, 2007), nor does it provide a good measure of the strength of the evidence that the sample provides for their statements (Berger and Sellke, 1987; Berkson, 1942; Dyckman, 2016; Edwards et al., 1963; Johnstone, 1994; Wagenmakers, 2007). We refer the reader to Johnstone (2021) and the previous chapter for a comprehensible explanation of this problem in the fields of accounting research and practice, respectively.

An alternative to p-value NHST is Bayesian hypothesis testing. A Bayesian hypothesis test can provide the auditor with a direct assessment of the strength of evidence for or against their hypothesis via the Bayes factor (e.g., Jeffreys, 1935, 1939; Kass and Raftery, 1995). In a Bayesian audit sampling procedure, the auditor defines a so-called prior distribution which incorporates all existing information about the misstatement in the population (Johnstone, 1990; Nichols and Baker, 1977). Applying this prior distribution confers two advantages to the auditor. Firstly, incorporating existing information into the prior distribution means there is more available information at the start of substantive testing, which generally allows for a more efficient estimate of the population misstatement (Knoblett, 1970). Secondly, since the prior distribution partly determines the predictions from the auditor's hypothesis, the information in the prior distribution is incorporated directly into a hypothesis test. This allows for direct accumulation of audit evidence, and thus a more transparent audit (Derks et al., 2021a). From a Bayesian point of view, the degree to which the data in the sample support or contradict the auditor's hypothesis over an alternative hypothesis—quantified by the Bayes factor—measures the evidence in the sample for or against this hypothesis.

Even though incorporating existing information can increase efficiency and transparency in substantive testing, there are cases in which an argument can be made to include only a minimal amount of information in the statistical analysis. For example, suppose that the auditor does not have any a priori information about the misstatement in the population, or does not have the desire to incorporate pre-existing information into the statistical analysis. For such cases, default Bayesian hypothesis tests have been developed for many common statistical designs in which the auditor is unable or unwilling to incorporate pre-existing information into the analysis. For example, Rouder et al. (2009) developed a default Bayesian test for t-tests, Wetzels and Wagenmakers (2012) developed a default test for correlations and partial correlations, and Wetzels et al. (2012) developed a default Bayesian hypothesis test for ANOVA designs. Auditors might like to be more efficient and incorporate pre-existing information into the statistical analysis, but at the same time they might be fearful of being perceived as prejudiced. This means that many auditors may be reluctant to adopt informed Bayesian hypothesis tests and might prefer default Bayesian hypothesis tests instead. However, to the best of our knowledge no default Bayesian hypothesis test exists that is specifically tailored to the problems faced in an audit sampling context. In this chapter we propose a default Bayesian hypothesis test for audit sampling to fill this gap.

This chapter is structured as follows. Firstly, we introduce the reader to Bayesian statistics in an audit sampling context. Secondly, we illustrate the dependency of the Bayes factor on the specification of the prior distribution. Next, we propose a specific type of prior distribution as the basis for a default Bayesian hypothesis test for audit sampling. We argue that this Bayesian hypothesis test is an attractive statistical tool for auditors because it is consistent, easy to interpret, and easy to explain. Finally, we discuss the merits and demerits of the default Bayesian hypothesis test in an audit sampling context.

5.2 The Bayesian approach to statistical audit sampling

From a statistical point of view, the auditor wants to perform inference about an unknown characteristic, θ, that represents the misstatement in the population, and test this characteristic against the performance materiality θ_{max}. In a Bayesian setting, the auditor needs to specify a prior distribution for θ that incorporates their pre-existing information about the misstatement in the population. After seeing a data sample, y, the information about θ in the prior distribution is updated using Bayes' rule (Equation 5.2.1), in which the symbol $\propto$ means that the expression to the left of this symbol equals the expression to the right of this symbol multiplied by a scaling factor.

$$\underbrace{p(\theta \,|\, y)}_{\text{Posterior}} \propto \underbrace{l(y \,|\, \theta)}_{\text{Likelihood}} \times \underbrace{p(\theta)}_{\text{Prior}} \tag{5.2.1}$$

Bayes' rule stipulates that, after seeing the sample, y, the prior distribution $p(\theta)$ is updated to a posterior distribution $p(\theta \,|\, y)$ according to how well its candidate values are consistent with the observed sample outcomes: the likelihood $l(y \,|\, \theta)$. In the next subsections we will elaborate on these three building blocks of Bayesian inference: the prior distribution, the likelihood, and the posterior distribution. We also discuss how in an audit sampling context the prior and posterior distribution can be intuitively used to calculate a Bayes factor for or against the hypothesis of (in)tolerable misstatement and illustrate this procedure in the context of attributes sampling[1].

5.2.1 The prior distribution

The prior distribution is a probability distribution that assigns a relative probability to each of the candidate values of θ, according to how likely that value is before seeing any information from a sample. The prior distribution can be set freely, but it can be convenient to choose the family of the prior distribution such that the posterior distribution is in the same family (i.e., a conjugate prior). In the context of attributes sampling, where items are evaluated as correct or incorrect, the conjugate prior is the beta prior distribution because it remains a beta distribution when updated by this type of data[2]. The probability density $p(\theta; \alpha, \beta)$ for the beta distribution with parameters α, β is shown in Equation 5.2.2, in which $B(\alpha, \beta)$ is the beta function.

$$p(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \tag{5.2.2}$$

Using a prior distribution in tests of details can increase efficiency since it allows the auditor to incorporate various types of pre-existing information into the statistical analysis. For example, information obtained from tests of internal control systems can be incorporated into the prior distribution to reduce the required sample size, or to estimate the true misstatement in the population more accurately. However, because the prior distribution is based on pre-existing information it is crucial that this information can also be justified by the auditor. We refer the reader to Derks et al. (2021a) for a discussion of various methods to incorporate and justify pre-existing audit information in the prior distribution. However, in situations where there is no pre-existing information a noninformative prior distribution is often preferred by the auditor.

[1]We have included an example for monetary unit sampling (MUS) using a gamma prior distribution in Appendix 5.B.

[2]Note that the beta distribution can also be used for monetary unit sampling.

A noninformative (sometimes called negligible) prior distribution is often described as a prior distribution that incorporates no pre-existing information about its candidate values in order "to let the data speak for themselves" (Gelman et al., 2013, p. 51). In Bayesian literature there has been much debate about what it means to incorporate no information, see for example Kerman (2011b, pp. 1453–1455). A classic example of a noninformative prior distribution follows the principle of indifference, meaning that it assigns equal probability to all candidate values of θ (Gelman et al., 2013, p. 31). Suppose that the auditor chooses to specify the aforementioned Beta(α, β) prior distribution for θ. A common choice of parameters that reflect negligible information is $\alpha, \beta = 1$. Given these values of the parameters, the prior distribution is a uniform distribution and can therefore be considered indifferent with respect to the candidate values of the misstatement. Hence, the assumptions underlying a Beta(1, 1) prior distribution are easy to explain for the auditor. The Beta(1, 1) prior distribution is displayed in Figure 5.1 (dashed line).

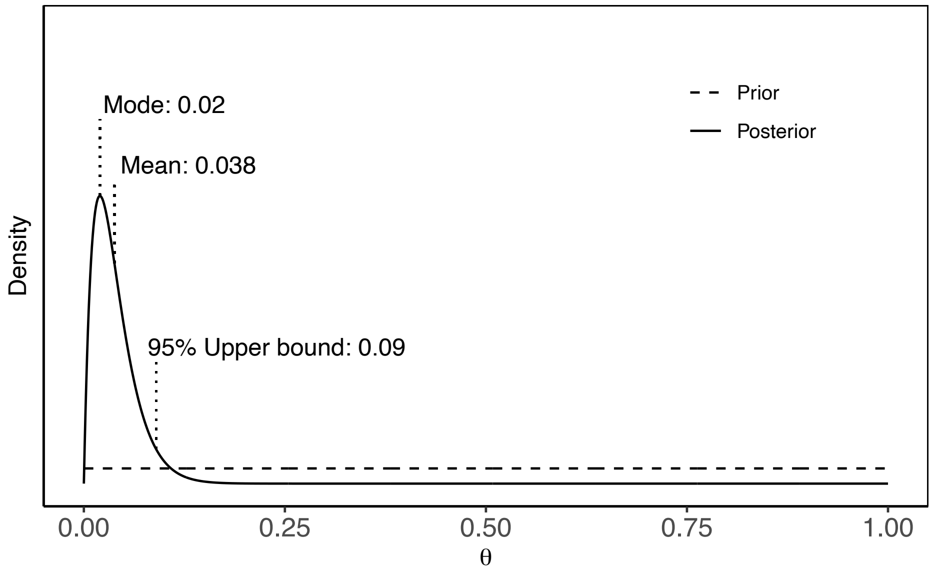


Figure 5.1: The Beta(1, 1) prior distribution for θ (dashed line) and Beta(2, 50) posterior distribution for θ (solid line). In the case of the Beta(1, 1) prior, the posterior distribution for θ is equal to the likelihood function of θ. The mode, mean, and 95 percent upper bound of the posterior distribution are indicated.

5.2.2 The likelihood

The likelihood function summarizes the evidence in the data about the parameters of the statistical model (Etz, 2018). More specifically, for each possible value of

θ, the likelihood function $l(y \mid \theta)$ measures the support provided by the data. In attributes sampling, the possible sample outcomes consist of k misstatements in n items. To explain how these possible sample outcomes relate to the population misstatement θ, the auditor must connect them using a probability distribution (Lehmann and Romano, 2006). The observed sample outcomes of k misstatements in n items are generally assumed to have been generated under a Binomial($k \mid n; \theta$) distribution, where the parameter θ represents the population misstatement rate (Stewart, 2012).

Suppose that the auditor inspects a sample, y, of $n = 50$ items and finds that $k = 1$ item contains a misstatement. To evaluate these sample outcomes in the context of the prior distribution the auditor can calculate the likelihood $l(y \mid \theta)$, which quantifies how likely the sample outcomes are under each candidate value of θ. For the above mentioned sample outcomes, the likelihood function is Binomial($k = 1 \mid n = 50; \theta$) and is depicted in Figure 5.1. Since the likelihood function is based solely on the sample outcomes k and n it has its maximum at $\theta = \frac{k}{n} = 0.02$, i.e., the observed misstatement proportion in the sample.

5.2.3 The posterior distribution

Using Bayes' rule, the information in the likelihood function can be used to update the prior distribution to the posterior distribution, which consequently quantifies the auditor's updated prior information about θ after seeing the sample. Since the posterior distribution is a probability distribution the auditor can formalize information about θ via its mathematical properties. For example, they can calculate location measures of the posterior distribution such as the mode that quantifies the most likely misstatement in the population, or the mean that quantifies the expected misstatement in the population. Moreover, they can calculate an x-percent upper bound for θ that expresses the value that exceeds, with x-percent probability, the misstatement in the population. Hence, the 95^{th} percentile of the posterior distribution $\theta_{.95}$ can be interpreted as a 95 percent upper bound for the population misstatement (Laws and O'Hagan, 2000; Neter and Godfrey, 1985).

Going back to the running example, the auditor updates their Beta($\alpha = 1$, $\beta = 1$) prior distribution with the Binomial($k = 1 \mid n = 50; \theta$) likelihood, resulting in a Beta($\alpha + k = 2$, $\beta + n - k = 50$) posterior distribution. The mode of the Beta(2, 50) posterior distribution is 0.02 and its mean is 0.038. The 95th percentile of the posterior distribution is 0.09 meaning that, given the pre-existing information and the information in the sample, there is a 95 percent probability that the misstatement in the population is lower than 9 percent. The Beta(2, 50) posterior distribution is shown in Figure 5.1 (solid line).

A Bayesian auditor will typically state that they are 95 percent certain a population does not contain misstatements that exceed the performance materiality when $\theta_{.95}$ is lower than θ_{max} (Stewart, 2013). This statement made using the posterior distribution is useful because it represents the auditor's aggregated knowledge about the misstatement θ over the course of the audit. However, from the posterior distribution alone the auditor is unable to quantify the evidence in the sample data for or against their claim. For this, the auditor needs to engage in an hypothesis test (as discussed in Chapter 4).

5.2.4 Bayesian hypothesis testing using the Bayes factor

A Bayesian hypothesis test involves a comparison of two competing hypotheses about the population misstatement θ. To test if θ is above or below the performance materiality θ_{max}, the auditor generally formulates the hypothesis of tolerable misstatement $\mathcal{H}_-$: $\theta < \theta_{max}$ and the hypothesis of intolerable misstatement $\mathcal{H}_+$: $\theta > \theta_{max}$. Note that these hypotheses correspond to a specific interval for θ and we have therefore denoted them as $\mathcal{H}_-$ and $\mathcal{H}_+$ instead of $\mathcal{H}_1$ and $\mathcal{H}_0$ to avoid confusion with a point null hypothesis. We leave the point null hypothesis out of this procedure because from an audit perspective, the case in which $\theta = \theta_{max}$ leads to an indecisive situation: On the one hand, θ_{max} can be seen as the maximum tolerable misstatement. On the other hand, θ_{max} can be seen as the minimum intolerable misstatement. In Section 5.4 we show that in this indecisive situation the Bayes factor will quantify this indecisiveness.

Formulating the two competing interval hypotheses $\mathcal{H}_+$ and $\mathcal{H}_-$ allows the auditor to evaluate each hypothesis individually in light of the observed data. Bayes' rule for hypothesis testing (Equation 5.2.3) prescribes that each hypothesis $\mathcal{H}_i$ is given a prior probability $p(\mathcal{H}_i)$ that represents the auditor's pre-existing information about the relative plausibility of the hypothesis before seeing the information from the sample. After observing the sample, y, the prior probabilities are updated to posterior probabilities $p(\mathcal{H}_i \mid y)$ according to how well the information in the sample accords with each hypothesis: the marginal likelihood $p(y \mid \mathcal{H}_i)$. The ratios of the prior and posterior probabilities are called the prior odds and posterior odds, respectively. The Bayes factor is the change from the prior to the posterior odds brought about by the information in the sample; hence the Bayes factor measures the relative evidence in the sample for the hypothesis $\mathcal{H}_-$ vis-a-vis the hypothesis $\mathcal{H}_+$:

$$\underbrace{\frac{p(\mathcal{H}_- \mid y)}{p(\mathcal{H}_+ \mid y)}}_{\text{Posterior odds}} = \underbrace{\frac{p(y \mid \mathcal{H}_-)}{p(y \mid \mathcal{H}_+)}}_{\text{Relative evidence}} \times \underbrace{\frac{p(\mathcal{H}_-)}{p(\mathcal{H}_+)}}_{\text{Prior odds}}. \tag{5.2.3}$$

In an audit sampling context, the Bayes factor for $\mathcal{H}_-$ or $\mathcal{H}_+$ can be intuitively obtained from the prior and posterior distributions directly. Since the hypotheses $\mathcal{H}_-$ and $\mathcal{H}_+$ cover mutually exclusive ranges of the parameter space of θ, it is allowed for the hypothesis $\mathcal{H}_-$: $\theta < \theta_{max}$ to be represented by the total probability in the range $[0, \theta_{max})$ of the prior and posterior distribution (Faulkenberry, 2019; Klugkist et al., 2005; Morey and Rouder, 2011). Vice versa, the hypothesis $\mathcal{H}_+$: $\theta > \theta_{max}$ can be represented by the total probability in the range $(\theta_{max}, 1]$ of the prior and posterior distribution. Hence, the prior and posterior distributions for θ encompass the hypotheses $\mathcal{H}_-$ and $\mathcal{H}_+$. Consequently, the Bayes factor can be calculated by comparing the probabilities of $\mathcal{H}_-$ and $\mathcal{H}_+$ under the prior and posterior distribution, see Equation 5.2.4. Equation 5.2.4 highlights that the Bayes factor is only dependent on the sample data y and the prior distribution $p(\theta)$.

$$BF_{-+}(y) = \frac{p(y \mid \mathcal{H}_-)}{p(y \mid \mathcal{H}_+)} = \frac{\int_0^{\theta_{max}} p(\theta \mid y)d\theta / \int_0^{\theta_{max}} p(\theta)d\theta}{\int_{\theta_{max}}^1 p(\theta \mid y)d\theta / \int_{\theta_{max}}^1 p(\theta)d\theta} \tag{5.2.4}$$

We continue the running example by calculating the Bayes factor for the sample of $n = 50$ items with $k = 1$ misstatement using the Beta(1, 1) prior distribution. Suppose the auditor has set the performance materiality θ_{max} to 3 percent. The areas corresponding to the hypotheses $\mathcal{H}_-$: $\theta < 0.03$ and $\mathcal{H}_+$: $\theta > 0.03$ are shaded for the Beta(1, 1) prior distribution in the left panel of Figure 5.2. The prior probability of the hypothesis $\mathcal{H}_-$ (light gray; 0.03) is the total probability in the range [0, 0.03) under the prior distribution, whereas the prior probability of $\mathcal{H}_+$ (dark gray; 0.97) is the total probability in the range (0.03, 1]. The prior odds in favor of $\mathcal{H}_-$ are therefore $\frac{p(\mathcal{H}_-)}{p(\mathcal{H}_+)} = \frac{0.03}{0.97} = 0.0309$, and $\frac{p(\mathcal{H}_+)}{p(\mathcal{H}_-)} = 32\frac{1}{3}$ in favor of $\mathcal{H}_+$.

After seeing the data from the sample, the posterior probability $p(\mathcal{H}_- \mid y) = 0.455$ is the total probability in the range [0, 0.03) under the posterior distribution, and the posterior probability $p(\mathcal{H}_+ \mid y) = 0.545$ is the total probability in the range (0.03, 1]. Hence, the posterior odds in favor of $\mathcal{H}_-$ are $\frac{0.455}{0.545} = 0.835$. The Bayes factor is the change from the prior odds to the posterior odds and is $BF_{-+} = \frac{0.835}{0.0309} = 27$. This Bayes factor indicates to the auditor that their sample of $n = 50$ items with $k = 1$ misstatement is 27 times more likely to be observed under $\mathcal{H}_-$ than under $\mathcal{H}_+$. According to guidelines proposed and reiterated in the Bayesian literature (van Doorn et al., 2021; Jeffreys, 1961; Wetzels et al., 2011), this Bayes factor implies strong evidence in favor of the hypothesis that the misstatement in the population is lower than 3 percent (see Table 5.1).

Table 5.1: Bayes factor labels motivated by Jeffreys (1961).

Bayes factor	Strength of evidence
1–3	Not worth more than a bare mention
3–10	Substantial
10–30	Strong
30–100	Very strong
> 100	Decisive

5.3 How the prior distribution influences the Bayes factor

Since the Bayes factor is heavily influenced by the prior distribution it is insightful to show the effect of the prior distribution on the Bayes factor. In this section we discuss three types of prior distributions: a uniform Beta(1, 1) prior distribution, an improper Beta(1, 0) prior distribution, and an impartial beta prior distribution. Based on its assumptions and the behavior of its Bayes factors, we suggest that the impartial prior distribution provides an attractive prior distribution for a default Bayesian hypothesis test.

5.3.1 Uniform prior: Beta($\alpha = 1$, $\beta = 1$)

The Beta(1, 1) prior distribution is a uniform prior distribution for θ in the range [0, 1] (see Figure 5.1). Given the range of the prior distribution and the fact that this prior distribution is flat, its median is $\frac{1}{2}$ and its mode is undefined.

The Beta(1, 1) prior distribution is a popular choice of noninformative prior distribution for audit sampling since it assigns equal probability to all candidate values of the misstatement θ (Steele, 1992; Stewart, 2013). This prior distribution exhibits indifference towards the individual values of θ since it assumes that every possible misstatement is equally likely before seeing the sample, which makes it easy to explain for the auditor.

When it comes to hypothesis testing in an audit sampling context, the Beta(1, 1) prior distribution is not impartial with respect to the two competing interval hypotheses $\mathcal{H}_-$ and $\mathcal{H}_+$. As discussed in the previous section, for a performance materiality $\theta_{max} = 0.03$, the Beta(1, 1) prior distribution contains the information that $\mathcal{H}_+$ is $32\frac{1}{3}$ times more plausible a priori than $\mathcal{H}_-$. Because in this case the prior distribution strongly prefers $\mathcal{H}_+$, even a small data sample from a population without material misstatement will greatly shift the support under the posterior distribution from $\mathcal{H}_+$ to $\mathcal{H}_-$. This behavior results in a Bayes factor that does not consistently quantify evidence for the hypothesis supported by the data. Instead, Bayes factors calculated on the basis of a Beta(1, 1) prior distribution can quantify evidence in favor of tolerable misstatement whereas the data show evidence in the opposite direction. More specifically the auditor may find that, although the sample contains misstatement (far) above the performance materiality, the Bayes factor quantifies evidence in favor of the hypothesis of tolerable misstatement.

We illustrate the relationship between the observed misstatement and the Bayes factor in the right panel of Figure 5.2, which shows the natural logarithmic Bayes factors from the Beta(1, 1) prior distribution as a function of the sample size, n, for various values of the observed misstatement $\hat{\theta} = \frac{k}{n}$. The panel shows that, even though the auditor finds misstatement in the sample equal to (or above) the performance materiality θ_{max}, under reasonable sample sizes the Bayes factor quantifies evidence in favor of the hypothesis $\mathcal{H}_-$ (i.e., $\ln(BF_{-+}) > 0$). However, in the large sample limit, the Bayes factor from a Beta(1, 1) prior distribution will quantify evidence in favor of the hypothesis supported by the data, except when the observed misstatement $\hat{\theta}$ is exactly equal to the performance materiality θ_{max}. In the case where $\hat{\theta} = \theta_{max}$, the median of the posterior distribution will go to θ_{max} as the sample size grows (see Appendix 5.A), which means the posterior odds will go to 1. Hence, the Bayes factor will be equal to $\frac{1}{p(\mathcal{H}_-)/p(\mathcal{H}_+)} = \frac{p(\mathcal{H}_+)}{p(\mathcal{H}_-)}$, a constant that depends on the prior distribution. For example, for the Beta(1, 1) prior distribution, the Bayes factor will go to $\frac{0.97}{0.03} = 32\frac{1}{3}$ (of which the natural logarithm is 3.48, see Figure 5.2) if $\hat{\theta} = \theta_{max} = 0.03$. This anchoring proofs that the Bayes factor obtained from a Beta(1, 1) prior distribution overquantifies the evidence in the data for $\mathcal{H}_-$, since in the large sample limit it indicates evidence in favor of $\mathcal{H}_-$ whereas the data contain inconclusive evidence.

In sum, even though the Beta(1, 1) prior distribution is indifferent with respect to the values of θ and therefore easy to explain and justify, it is arguably unsuited

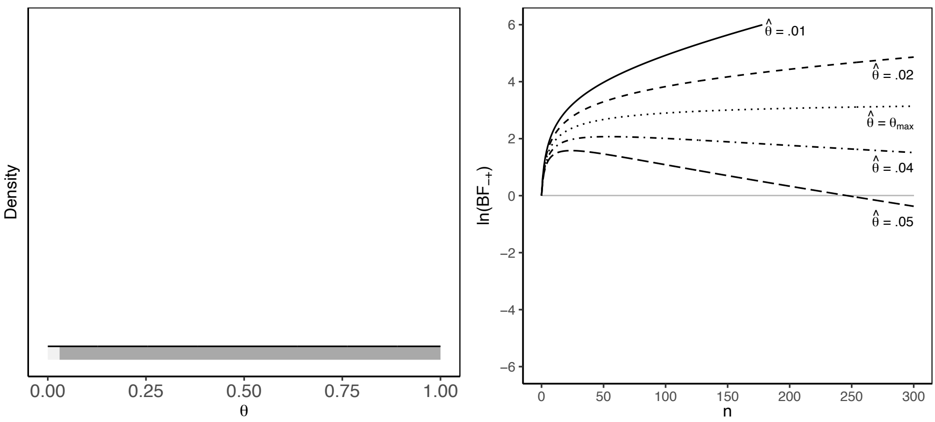


Figure 5.2: The hypotheses are $\mathcal{H}_-$: $\theta < \theta_{max}$ and $\mathcal{H}_+$: $\theta > \theta_{max}$, where θ_{max} has the value 0.03. The left panel displays the Beta(1, 1) prior distribution in which the area corresponding to $\mathcal{H}_-$ (light) and the area corresponding to $\mathcal{H}_+$ (dark) are indicated. The right panel shows natural logarithmic Bayes factors $\ln(BF_{-+})$ as a function of sample size for five observed misstatement proportions $\hat{\theta}$.

as a prior distribution for a default Bayesian hypothesis test for audit sampling since the resulting Bayes factor does not consistently quantify evidence for the hypothesis supported by the data. Clearly this is an undesirable situation for auditors because, on first sight, the conclusions following their statistical analysis may not match the data they use to substantiate these conclusions with. However, this paradox is resolved by noting that the Beta(1, 1) prior distribution expresses a strong preference for the hypotheses $\mathcal{H}_+$. It is therefore imperative that this prior is only used if the auditor can justify the strong assumption that the hypothesis of tolerable misstatement is highly likely before seeing a sample.

5.3.2 Improper prior: Beta($\alpha = 1$, $\beta = 0$)

The Beta(1, 0) prior distribution is a prior distribution for θ in the range [0, 1] that is improper, meaning its set of candidate values for θ is not well defined. Specifically, this prior distribution is the equivalent of an infinite point mass concentrated at $\theta = 1$. Given its shape, the median and mode of this prior distribution are both 1. Note that, even though the Beta(1, 0) prior distribution is improper, the resulting posterior distribution is proper only if $\beta > 0$. This implies $n - k > 0$, which means that there should be at least one non-misstated item in the sample. If $n - k = 0$, meaning that all items in the sample are misstated, the posterior distribution retains the shape of the prior distribution and is also improper.

The Beta(1, 0) prior distribution is a candidate for audit sampling because it is the implicit prior distribution in NHST. Concretely, in a frequentist binomial test the one-sided p-value (i.e., the probability of observing k misstatements or less in a sample of n items given that the null model $\theta = \theta_{max}$ is true) is equal to the probability that the misstatement θ exceeds the performance materiality θ_{max} under a Beta($1 + k$, $n - k$) distribution (Stewart, 2012, p. 8). Since the Beta($1 + k$, $n - k$) distribution is the posterior distribution for θ resulting from a Beta(1, 0) prior, the frequentist one-sided p-value can be interpreted as a Bayesian posterior probability for $\mathcal{H}_+$ given the improper Beta(1, 0) prior distribution (see Appendix 5.A). Hence, Bayesian planning and Bayesian evaluation of an audit sample using the Beta(1, 0) prior distribution yields the same results with respect to sample sizes and upper limits as NHST procedures (Mcbride and Ellis, 2001; Pratt, 1965; Thatcher, 1964).

We illustrate this similarity between the p-value and the posterior probability for $\mathcal{H}_+$ in the context of our running example, in which a sample of $n = 50$ and $k = 1$ was observed. Using a binomial test to assess the null hypothesis $\mathcal{H}_0$: $\theta \geq 0.03$, we obtain a one-sided p-value of 0.555. Using the improper Beta(1, 0) prior distribution, the posterior distribution is a Beta(2, 49) distribution. The probability for $\mathcal{H}_+$ under this posterior distribution is 0.555, which corresponds to the one-sided p-value. Note that this connection implies that an improper[3] prior distribution is (implicitly) the foundation of reference tables like those in American Institute of Certified Public Accountants (AICPA) (2019, Appendix A and Appendix C) and because of that it is commonly used by auditors in practice. However, because the Beta(1, 0) distribution concentrates an infinite amount of prior probability at $\theta = 1$, it is the most conservative prior distribution possible.

The Beta(1, 0) prior distribution is easy to justify for the auditor because it is the implicit prior distribution in NHST. However, there are three downsides due to the Beta(1, 0) prior distribution being improper. Firstly, since in an audit context it is crucial that the prior distribution can be explained by the auditor, the fact that the prior distribution assumes all items in the population to be fully misstated means that the auditor will have to justify this extremely conservative and unrealistic assumption. Secondly, because the posterior distribution is only proper if $n - k > 0$, this prior distribution can only be used if the sample contains at least one non-misstated item. Thirdly, because the prior probabilities of the competing interval hypotheses are undefined, this prior distribution leads to an infinite Bayes factor for every outcome in which k is lower than n. Naturally, this results in an undesirable situation for the auditor if they want to perform a Bayesian hypothesis test. Therefore, the Beta(1, 0) prior is arguably unsuited as a prior for a default Bayesian hypothesis test in an audit sampling context.

5.3.3 Impartial prior: Beta($\alpha \geq 1$, $\beta > 1$)

We propose an alternative prior distribution for θ that addresses the aforementioned problems of the Beta(1, 1) and Beta(1, 0) prior distributions with respect to the Bayes factor. The proposed impartial prior distribution for θ in the range [0, 1] has its median at the performance materiality θ_{max} and its mode at the ex-

[3]For American Institute of Certified Public Accountants (AICPA) (2019, Appendix A), the corresponding prior distribution is a Beta(1, 0) distribution and for American Institute of Certified Public Accountants (AICPA) (2019, Appendix C), the corresponding prior distribution is a Gamma(1, 0) distribution (Stewart, 2012, pp. 8–10).

pected most likely misstatement θ_{exp}. Therefore, the shape of the impartial prior distribution depends on these two parameters.

The impartial prior distribution is a candidate for audit sampling because it assumes the auditor expresses no prior preference for $\mathcal{H}_-$ or $\mathcal{H}_+$. Therefore, it can be considered impartial with respect to the competing interval hypotheses (Berger and Mortera, 1999; Casella and Berger, 1987). This makes the assumptions of the prior distribution easy to explain and justify for the auditor. Furthermore, the statistical motivation for assuming equal prior probabilities for $\mathcal{H}_-$ and $\mathcal{H}_+$ (i.e., $p(\mathcal{H}_-) = p(\mathcal{H}_+) = \frac{1}{2}$) is that, under this condition, the constant that the Bayes factor attains when $\theta = \theta_{max}$ equals $\frac{p(\mathcal{H}_+)}{p(\mathcal{H}_-)} = 1$. In other words, the Bayes factor does not differentiate between $\mathcal{H}_-$ and $\mathcal{H}_+$ if both are equally supported by the data.

We return to the running example by calculating the Bayes factor for the sample of $n = 50$ items with $k = 1$ misstatement using an impartial prior distribution. Using the Beta(1, 22.757) prior distribution, which has its mode at 0 and its median at 0.03, and the sample outcomes $n = 50$ and $k = 1$ the impartial Bayes factor in favor of $\mathcal{H}_-$ is $BF_{-+} = 1.822$ (see Equation 5.4.1), implying that the sample outcomes are about 2 times more likely to be observed under $\mathcal{H}_-$ than under $\mathcal{H}_+$ (anecdotal evidence). Note that the impartial Bayes factor is (much) lower than the Bayes factor of 27 resulting from the Beta(1, 1) prior distribution because the prior odds induced by the Beta(1, 1) prior distribution favored the hypothesis $\mathcal{H}_-$ much less to begin with. Also note that the Bayes factor resulting from the Beta(1, 0) prior distribution is even larger, since it diverts to infinity for every outcome in which k is lower than n.

In contrast to the Beta(1, 1) and Beta(1, 0) prior distributions, the impartial prior distribution is a suitable prior distribution for a default Bayesian hypothesis test for audit sampling because its assumptions are easy to explain and justify, and the Bayes factor can be calculated for all binomial data. In the next section we will explain how to calculate the impartial prior distribution and Bayes factor, and investigate the behavior of the resulting impartial Bayesian hypothesis test.

5.4 An impartial Bayesian hypothesis test

In the previous section we have outlined why the impartial prior distribution is a sensible prior that can be justified by the auditor. However, in order to propose this prior as a basis for a default Bayesian hypothesis test we further investigate the behavior of the impartial Bayes factor in this section. Firstly, we will describe how to define the impartial prior distribution and how to calculate the corresponding impartial Bayes factor. Secondly, we will discuss a desirable statistical property of the impartial Bayes factor that describes its behavior: model selection consistency.

5.4.1 Calculating the impartial prior and Bayes factor

The impartial prior distribution assigns equal prior probabilities to the interval hypotheses $\mathcal{H}_-$ and $\mathcal{H}_+$ and therefore the median of the prior lies at the performance materiality θ_{max}. Under this condition the mode of the prior distribution can be set freely to $\theta_{exp} \in [0, \theta_{max})$. Here, the parameter θ_{exp} can be interpreted as the auditor's assessment of the expected most likely misstatement in the intended sample; information that is generally available at the start of substantive testing. In the next subsections, we discuss how to determine the parameters of the impartial prior distribution and calculate the corresponding Bayes factor for two scenarios: one where the auditor does not expect any misstatements in the intended sample and one where the auditor does expect misstatements.

5.4.1.1 Zero expected most likely misstatement: Beta($\alpha = 1$, $\beta > 1$)

The assumption of no expected misstatements in the sample (i.e., $\theta_{exp} = 0$) implies that the mode of the prior distribution is zero. This condition restricts the α and β parameters of the beta prior distribution to $\alpha = 1$ and $\beta > 1$. Given these restrictions, the median of the beta distribution can be expressed as $1 - 2^{-\frac{1}{\beta}}$ (Kerman, 2011a). Therefore, setting the median of the prior distribution to the performance materiality θ_{max} implies that $\beta = \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}$. Since for this prior distribution the prior odds are 1, from Equation 5.2.4 it follows that the impartial Bayes factor can be calculated as the ratio of posterior odds (Equation 5.4.1), which is solely a function of the performance materiality θ_{max} and the sample outcomes n and k.

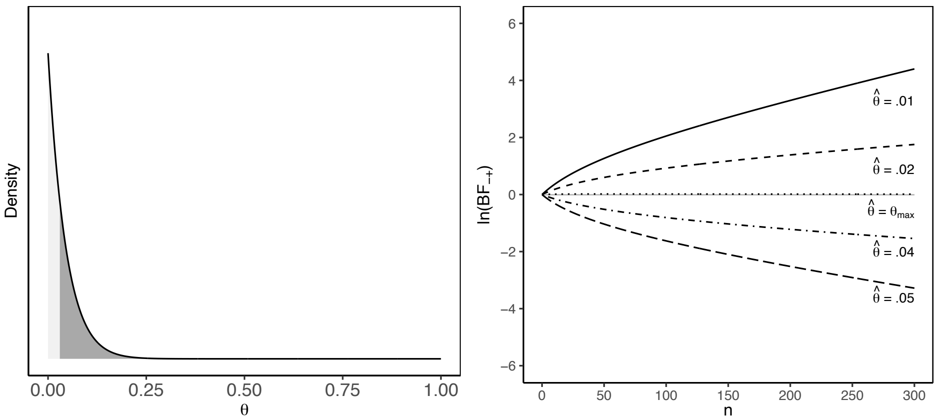


Figure 5.3: The hypotheses are $\mathcal{H}_{-}$: $\theta < \theta_{max}$ and $\mathcal{H}_{+}$: $\theta > \theta_{max}$, where θ_{max} has the value 0.03. The a priori most likely expected error θ_{exp} is 0. The left panel displays the Beta(1, 22.757) prior distribution in which the area corresponding to $\mathcal{H}_{-}$ (light) and the area corresponding to $\mathcal{H}_{+}$ (dark) are indicated. The right panel shows natural logarithmic Bayes factors $\ln(BF_{-+})$ as a function of sample size for five observed misstatement proportions $\hat{\theta}$.

$$BF_{-+}(\theta_{max}, n, k) = \frac{p(y \mid \mathcal{H}_-)}{p(y \mid \mathcal{H}_+)} = \frac{\int_0^{\theta_{max}} p(\theta; 1+k, \ln(\frac{1}{2})/\ln(1-\theta_{max}) + n - k) d\theta}{\int_{\theta_{max}}^1 p(\theta; 1+k, \ln(\frac{1}{2})/\ln(1-\theta_{max}) + n - k) d\theta} \tag{5.4.1}$$

To illustrate, for a performance materiality $\theta_{max} = 0.03$ the β parameter of the impartial beta prior distribution is $\beta = 22.757$. The left panel in Figure 5.3 shows the Beta(1, 22.757) prior distribution, which has its mode at 0 and its median at 0.03. Using the Beta(1, 22.757) prior distribution and the sample outcomes $n = 50$ and $k = 1$ the impartial Bayes factor in favor of $\mathcal{H}_-$ is $BF_{-+} = 1.822$. The right panel of Figure 5.3 illustrates the behavior of this impartial Bayes factor given various values of the observed misstatement $\hat{\theta} = \frac{k}{n}$. The panel shows that, for any value $\hat{\theta} \neq \theta_{max}$, the impartial Bayes factor consistently quantifies evidence for the hypothesis that accords best with the data. Moreover, this evidence will grow increasingly stronger as the sample size increases.

5.4.1.2 Non-zero expected most likely misstatement: Beta($\alpha > 1$, $\beta > 1$)

The assumption of non-zero expected misstatements (i.e., $\theta_{exp} > 0$) implies that the mode of the prior distribution is greater than zero. This condition restricts the α and β parameters of the prior distribution to $\alpha > 1$ and $\beta > 1$. Given these restrictions, the mode of the beta distribution can be expressed as $\theta_{exp} = \frac{\alpha-1}{\alpha+\beta-2}$ and the median can be approximated as $\theta_{max} \approx \frac{\alpha-\frac{1}{3}}{\alpha+\beta-\frac{2}{3}}$ (Kerman, 2011a). Thus, the α and β parameters of a beta distribution whose mode is equal to θ_{exp} and whose median is equal to θ_{max} can be approximated[4] by $\alpha \approx \frac{3\theta_{max}-\theta_{exp}(4\theta_{max}+1)}{3(\theta_{max}-\theta_{exp})}$ and $\beta \approx \frac{2+\theta_{max}(4\theta_{exp}-1)-5\theta_{exp}}{3(\theta_{max}-\theta_{exp})}$.

To illustrate, for a performance materiality $\theta_{max} = 0.03$ and an expected most likely misstatement in the sample $\theta_{exp} = 0.015$, the parameters of the impartial prior distribution are $\alpha = 1.641$ and $\beta = 43.105$. The left panel in Figure 5.4 shows the Beta(1.641, 43.105) prior distribution, which has its mode at 0.015 and its median at 0.03. After seeing the sample outcomes $n = 50$ and $k = 1$ the Bayes factor is $BF_{-+} = 1.668$, implying that the sample outcomes are 1.668 times more likely to have occurred under $\mathcal{H}_-$ than under $\mathcal{H}_+$. The right panel of Figure 5.4 illustrates the behavior of this impartial Bayes factor given various values of the observed misstatement $\hat{\theta}$. Again, this panel shows that for any value $\hat{\theta} \neq \theta_{max}$ the impartial Bayes factor consistently quantifies evidence for the hypothesis that accords best with the data.

[4] Please see Appendix 5.A to this chapter for the derivations of these formulas. Note that an iterative procedure, in which one determines $\beta = \frac{\alpha-1}{\theta_{exp}-\alpha+2}$ and gradually increases $\alpha > 1$ until the median of the beta distribution is below θ_{max}, can be more accurate in determining the parameters of the prior distribution under the assumption of expected misstatement. For this reason, an iterative procedure has been used to calculate the parameters of this prior distribution.

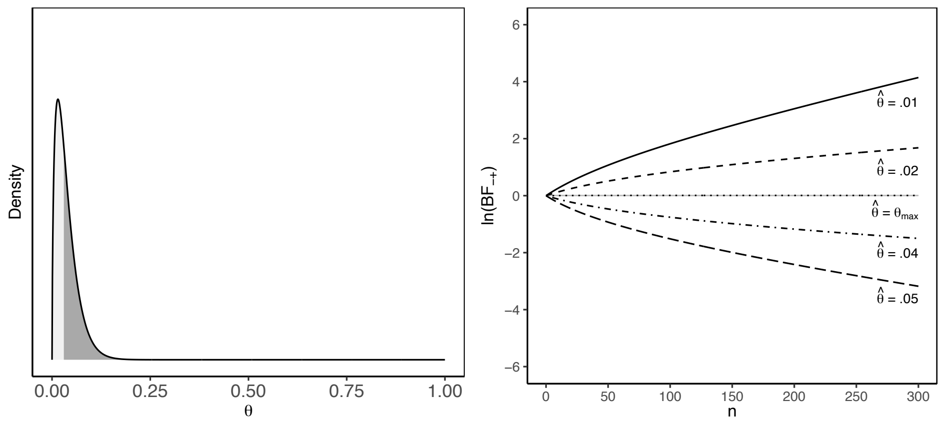


Figure 5.4: The hypotheses are $\mathcal{H}_-$: $\theta < \theta_{max}$ and $\mathcal{H}_+$: $\theta > \theta_{max}$, where θ_{max} has the value 0.03. The a priori most likely expected error θ_{exp} is 0.015. The left panel displays the Beta(1.641, 43.105) prior distribution in which the area corresponding to $\mathcal{H}_-$ (light) and the area corresponding to $\mathcal{H}_+$ (dark) are indicated. The right panel shows natural logarithmic Bayes factors $\ln(BF_{-+})$ as a function of sample size for five observed misstatement proportions $\hat{\theta}$.

5.4.2 Model selection consistency

Since one goal of this chapter is to develop a prior distribution that can be used widely and easily by auditors in practice, we will discuss a statistical property of the impartial Bayes factor that describes its consistent behavior: model selection consistency (Ly et al., 2016; van Ravenzwaaij and Etz, 2021).

The property of model selection consistency implies that, for a sample generated from a model, the Bayes factor supporting that model should go to infinity as the sample size goes to infinity. For example, for data sampled from a population without material misstatement, the Bayes factor in favor of $\mathcal{H}_-$ should grow without bound if the number of samples increases. Concretely, for any sample with $\hat{\theta} < \theta_{max}$, the numerator in Equation 5.4.1 will approach one as the sample size grows, its denominator will approach zero, which means that the impartial Bayes factor will tend to infinity (Equation 5.4.2). Vice versa, for any sample with $\hat{\theta} > \theta_{max}$, the numerator in Equation 5.4.1 will approach zero as the sample size grows, its denominator will approach one, which means that the impartial Bayes factor will go to zero (Equation 5.4.3). Hence, the impartial Bayes factor leads to the correct decision whenever the sample size is large enough.

$$\hat{\theta} < \theta_{max} \Rightarrow \lim_{n \to \infty} BF_{-+} = \infty \tag{5.4.2}$$

$$\hat{\theta} > \theta_{max} \Rightarrow \lim_{n \to \infty} BF_{-+} = 0 \tag{5.4.3}$$

Furthermore, when data from a sample is observed but does not show evidence to support either of the hypotheses the Bayes factor ought to be indifferent, i.e.,

$BF_{-+} = 1$. In the case of $\hat{\theta} = \theta_{max}$, the median of the posterior distribution remains close to θ_{max} which implies equal posterior probabilities and—given the equal prior probabilities—a Bayes factor of about 1 (Equation 5.4.4)[5].

$$\hat{\theta} = \theta_{max} \Rightarrow \lim_{n \to \infty} BF_{-+} \approx 1 \tag{5.4.4}$$

Considering that the impartial Bayes factor is model selection consistent and indifferent in case the data supports none of the hypothesis, we believe that the impartial prior is a sensible prior distribution for a default Bayesian hypothesis test for audit sampling.

5.5 Comparison with p-value methodology

In the previous section we have shown that the impartial Bayes factor consistently quantifies evidence for the hypothesis supported by the data. Moreover, we have argued that the impartial prior distribution is easy to justify and explain in an audit sampling context. With these properties in hand, auditors can obtain evidence for or against the hypothesis of (in)tolerable misstatement without having to spend time and effort to justify the (effect of the) prior distribution. However, even though the impartial Bayes factor is easy to use, the auditor must be aware of how it compares to the p-value NHST methodology that is currently often used in the audit practice (i.e., is prescribed in audit guides like American Institute of Certified Public Accountants (AICPA) (2019)). In this section we first discuss two practical advantages of the impartial prior distribution compared to NHST procedures. Next, we provide a detailed comparison of Bayes factors and p-values.

Firstly, since the impartial prior distribution is less conservative than the implicit Beta(1, 0) prior distribution in NHST, using the impartial prior distribution for planning an audit sample leads to a more efficient sampling procedure. In this case, more efficient means that the same audit result can be achieved with a smaller sample size. In Appendix 5.A we show that, if the auditor expects no misstatements in the sample, the relative sample size reduction compared to NHST equals $\frac{\ln(\frac{1}{2})}{\ln(\text{ACR})}$. Here, ACR is the Audit Control Risk, the risk of incorrectly rejecting the null hypothesis $\mathcal{H}_0$: $\theta \geq \theta_{max}$. To illustrate, at a confidence level of 95 percent, ACR equals 5 percent and, in the case of no expected misstatements, a reduction of $\frac{\ln(\frac{1}{2})}{\ln(0.05)} = 23$ percent is achieved by using the impartial prior distribution when compared to planning a sample using the NHST approach. In Table 5.2 we provide a more detailed comparison of sample sizes in cases where misstatements are expected. The table shows that, for larger values of the expected most likely misstatement θ_{exp}, the absolute reduction in sample size increases but the relative reduction in sample size decreases.

[5]In Appendix 5.A we show that the impartial Bayes factor gets increasingly close to 1 as the sample size grows without bound. Note that θ is a continuous variable and therefore, in practice $\hat{\theta}$ and θ_{max} will often not be exactly equal. As a consequence, while Equation 5.4.4 theoretically holds true, in individual cases the Bayes factor will always keep meandering randomly as the data come in (but on average the Bayes factor will be 1, see Wagenmakers et al. (2019) for more details).

Table 5.2: Comparison of required sample sizes for traditional NHST versus the impartial prior distribution for various values of the expected most likely misstatement θ_{exp}. The table shows outcomes for a performance materiality $\theta_{max} = 0.03$ and a confidence level of 95 percent (i.e., ACR = 0.05).

θ_{exp}	n_{NHST}	$n_{\text{impartial}}$	Δn	Relative reduction
0	99	76	23	23.23%
$\frac{1}{157}$	157	127	30	19.11%
$\frac{2}{208}$	208	174	34	16.35%
$\frac{3}{257}$	257	218	39	15.18%
$\frac{4}{303}$	303	262	41	13.53%

Secondly, using the impartial prior distribution for evaluating an audit sample leads to a more efficient use of information, meaning more audit evidence can be obtained from the same sample. In Appendix 5.A we show that, for $\hat{\theta} = 0$, the relationship between the one-sided p-value and the impartial Bayes factor simplifies to $BF_{-+} = \frac{2-p}{p}$. Therefore, if no misstatements are found in the sample, the evidence against $\mathcal{H}_+$ obtained using the impartial Bayes factor is more than the evidence against $\mathcal{H}_0$ indicated by the p-value. The relationship between the p-value and the impartial Bayes factor approaches $BF_{-+} = \frac{1-p}{p}$ as $\hat{\theta} = 1$ and n goes to infinity. From this, we conclude that the impartial Bayes factor can obtain at least as much evidence as the one-sided p-value when the misstatement is high, and more evidence when the misstatement is low.

To illustrate the relationship between the p-value of a one-sided null hypothesis test and Bayes factors more clearly, we describe the results of a simulation study. In this study we simulated 3000 audit samples with size $n \in \{100, \ldots, 500\}$ and observed misstatements, $k \in \{0, \ldots, n\}$, and compared the results with respect to the observed misstatement $\hat{\theta} = \frac{k}{n}$, the one-sided p-value, and two Bayes factors: one calculated using a Beta(1, 1) distribution and the other calculated using an impartial prior distribution. In this simulation study, the performance materiality was set to 3 percent, implying that p-values indicate evidence against the hypothesis $\mathcal{H}_0$: $\theta \geq 0.03$ and Bayes factors indicate evidence in favor of $\mathcal{H}_-$: $\theta < 0.03$ over $\mathcal{H}_+$: $\theta > 0.03$. In the subsections below we provide a detailed comparison between the observed misstatement in the sample, the one-sided p-value, and the two Bayes factors. In general, Bayes factors and p-values do not disagree about the direction of the evidence in the data. However, the two measures are known to disagree about the strength of the evidence in many cases (Wetzels et al., 2011).

5.5.1 Evaluating the relationship between the p-values and the observed misstatement $\hat{\theta}$

The relationship between the p-value and $\hat{\theta}$ is shown as a scatter plot in Figure 5.5. Each point corresponds to one of the 3000 simulated combinations of n and k. For interpretation, the figure is divided into panels that distinguish different evidence categories (Wetzels et al., 2011). The figure shows that the p-value is roughly in line with the observed misstatement, as higher values of $\hat{\theta}$ generally correspond to higher p-values. However, note that $p = 0.05$ for a sample with 10 items should provide more evidence against the null hypothesis than $p = 0.05$ for a sample with 100 items. Since identical p-values do not convey identical levels of evidence, the p-value is a poor measure of statistical evidence (Lindley, 1957). Moreover, the top left panel of Figure 5.5 points out that even for the larger simulated sample sizes where $\hat{\theta}$ is substantially lower than 0.03, $\mathcal{H}_0$ is still not rejected.

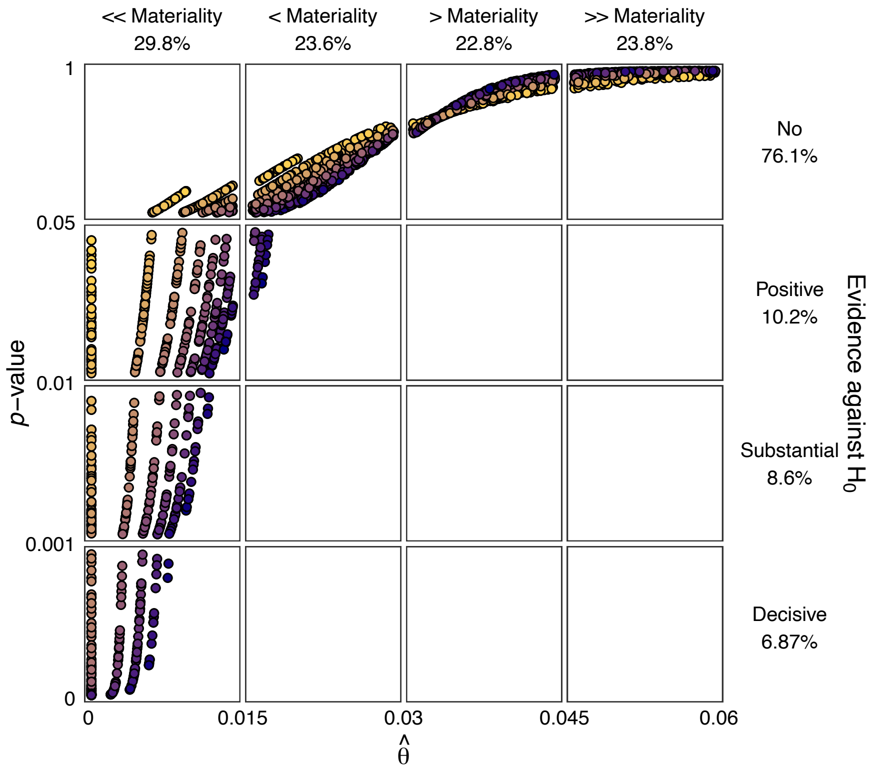


Figure 5.5: The relationship between the observed misstatement and p-values when testing the null hypothesis $\mathcal{H}_0$: $\theta \geq 0.03$. Darker (blue) points correspond to higher sample sizes. The scale of the axes is based on the decision categories given in Wetzels et al. (2011).

5.5.2 Evaluating the relationship between the observed misstatement $\hat{\theta}$ and the Bayes factor

The relationship between the observed misstatement $\hat{\theta}$, the Beta(1, 1) Bayes factor, and the impartial Bayes factor is shown in Figure 5.6. Like the comparison of p-values and observed misstatement, both Bayes factors are closely in line with the observed misstatement. For example, lower values of $\hat{\theta}$ correspond to higher values of BF_{-+}, and higher values of $\hat{\theta}$ corresponds to lower values of BF_{-+}. The difference between the two Bayes factor approaches is in what they consider to be inconclusive evidence. In the left panel it can be seen that Bayes factors calculated using a Beta(1, 1) prior distribution quantify inconclusive evidence at $\hat{\theta} > \theta_{max}$ (the exact value depends on n), whereas impartial Bayes factors do so at $\hat{\theta} = \theta_{max}$. This implies that, for data sampled from a population without material misstatement, a uniform prior will always produce a larger Bayes factor than an impartial prior.

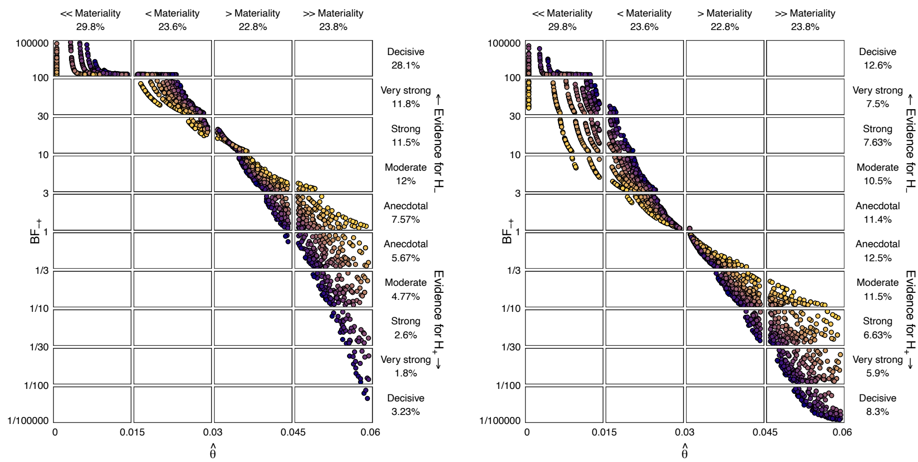


Figure 5.6: The relationship between Bayes factors and the observed misstatement when testing the hypothesis $\mathcal{H}_{-}$: $\theta < 0.03$ versus the hypothesis $\mathcal{H}_{+}$: $\theta > 0.03$. The left panel shows Bayes factors using the Beta(1, 1) prior distribution and the right panel shows Bayes factors using the Beta(1, 22.757) distribution. Darker (blue) points correspond to higher sample sizes. The scale of the axes is based on the decision categories given in Wetzels et al. (2011).

5.5.3 Evaluating the relationship between the p-value and the Bayes factor

The frequentist NHST approach is still the dominant hypothesis testing framework in audit sampling. Consequently, it is of interest to auditors who would like to apply this Bayes factor in practice to see which values of the impartial Bayes factor compare to certain p-values.

To illustrate this relationship, we show the p-values, Beta(1, 1) Bayes factors, and impartial Bayes factors calculated from the simulated data in Figure 5.7. Note that the relationship between the p-value and the Bayes factor is much more exact than the relationship between θ_{max} and the p-value and Bayes factors in the previous comparisons. As discussed in the beginning of this section, if no misstatements are expected and observed in the sample the relationship between the p-value and the impartial Bayes factor can be expressed as $BF_{-+} = \frac{2-p}{p}$. In this case, a p-value of 0.1 corresponds to an impartial Bayes factor of 19 and a p-value of 0.05 corresponds to an impartial Bayes factor of 39, whereas a p-value of 0.01 corresponds to an impartial Bayes factor of 199. Note that, if the number of (expected) misstatements in the sample increases, the relationship between the one-sided p-value and the impartial Bayes factor approaches $BF_{-+} = \frac{1-p}{p}$. In these cases, p-values of 0.1 correspond to an impartial Bayes factor of 9 or 19, and a p-value of 0.05 corresponds to an impartial Bayes factor of 19 or 39, whereas a p-value of 0.01 corresponds to an impartial Bayes factor of 99 or 199.

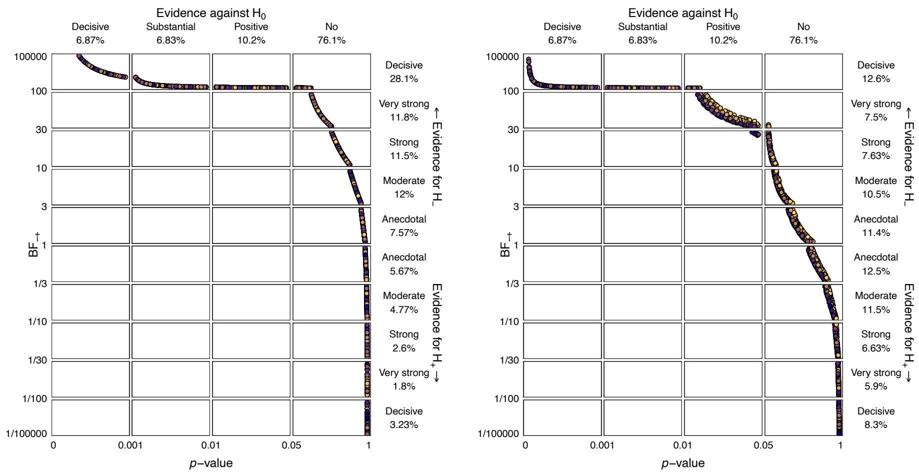


Figure 5.7: The relationship between Bayes factors testing the hypothesis $\mathcal{H}_{-}$: $\theta < 0.03$ versus the hypothesis $\mathcal{H}_{+}$: $\theta > 0.03$ and p-values testing the null hypothesis $\mathcal{H}_0$: $\theta \geq 0.03$. The left panel shows Bayes factors using the Beta(1, 1) prior distribution and the right panel shows Bayes factors using the Beta(1, 22.757) distribution. Darker (blue) points correspond to higher sample sizes. The scale of the axes is based on the decision categories given in Wetzels et al. (2011).

This has practical implications for auditors who perform audit sampling. Compared to the p-value, the impartial Bayes factor more quickly indicates that the misstatement is below the performance materiality. For example, if in the frequentist framework the null hypothesis would be rejected at $p < .1$, the same conclusion can be drawn with the impartial Bayes factor approach, but the decision bound would be set to either 9 or 19 depending on the expectation of the auditor. Likewise, an impartial $BF_{-+} = 10$ under the assumption of no expected misstatement corresponds to $p = 0.18$. From this we conclude that the impartial Bayes factor can yield similar decisions in planning and evaluation as the p-value methodology, but only if the decision bounds are set to specific values.

5.6 Concluding comments

In this chapter we have shown how to calculate the Bayes factor in an audit sampling context using the prior and posterior distribution directly. In sum, the auditor can specify a prior distribution for the population misstatement, θ, and obtain the Bayes factor by calculating the ratio of prior-to-posterior odds induced by the prior and posterior distribution. In this chapter we have also discussed that auditors need to be aware of what a prior distribution implies about the prior odds, and what its effect is on the Bayes factor. We have proposed a default Bayesian hypothesis test based on a prior distribution that is impartial with respect to the hypotheses being tested and whose Bayes factors have desirable statistical properties. Moreover, the assumptions underlying these prior distributions are easy to explain and justify for auditors. Our main conclusion is therefore that the proposed default Bayesian hypothesis test based on the impartial prior distribution provides more suitable answers in audit sampling than the p-value or Bayes factors computed with other noninformative prior distributions.

Hence, the impartial Bayes factor is an attractive statistical measure that can be easily calculated given the performance materiality, the sample size, and the observed misstatements. Its ease of calculation and justification makes the impartial Bayes factor easy to apply for any auditor, for example to re-formulate the outcomes of a frequentist sample from a Bayesian point of view. Another advantage of the impartial prior distribution is that it increases the auditor's efficiency. We have shown that using the impartial prior distribution allows the auditor to work with smaller sample sizes when compared to commonly used alternatives such as NHST. Moreover, because the impartial prior distribution requires input from the auditor about the expected most likely misstatement θ_{exp}, it is also suitable as a prior for a sensitivity analysis (Martel-Escobar et al., 2005). By computing the Bayes factor under many plausible values of θ_{exp}, the auditor can assess the robustness of the impartial Bayes factor to the choice of θ_{exp}. Finally, given that in the Bayesian framework the auditor can sequentially accumulate and monitor evidence over time, the impartial prior can be an efficient starting point for audit sampling procedures in which the auditor can decide whether to stop sampling or continue collecting more evidence at any point.

Nevertheless, we are aware that the auditor can select a variety of different prior distributions to be impartial with respect to the hypotheses $\mathcal{H}_-$ and $\mathcal{H}_+$. For example, they might specify a Beta(1, 1) prior distribution truncated to the range $[0, \theta_{max} \times 2]$ (symmetric on the real scale), or they might specify a localized beta prior (Ly, 2018, Chapter 11) (symmetric on the log-odds scale), to assign equal prior probabilities to the hypotheses. While there exists a plethora of viable options, we propose the impartial prior distribution because it is easy to use and to justify and because it can incorporate relevant audit information. That is, the impartial prior can be readily applied by any auditor without spending too much time or effort. This enables auditors to shift their focus away from quantifying evidence and toward judging evidence, an activity that is arguably much more in line with their area of expertise.

To make the proposed impartial prior distributions and Bayes factors easily accessible for auditors, we provide benchmark tables like those in American Institute of Certified Public Accountants (AICPA) (2019, Appendix A and Appendix C) in Appendix 5.C. These tables contain impartial Bayes factors for the binomial likelihood for some possible combinations of the sample size, n, and the observed misstatements, k, for a hypothesized maximum misstatement $\theta_{max} = 0.01$, 0.05, and 0.1, and zero expected misstatements. We propose that such tables can accompany the standard sample size and upper limit tables to give the auditor an indication of the strength of evidence in their samples. Note that the provided tables contain only a subset of the possible Bayes factors and only those for the binomial likelihood, but that they can be easily extended using the formulas provided in this chapter and the appendix. Moreover, we have made these Bayes factor calculations available in the R package 'jfa' (Derks, 2022) and in the Audit module (Derks et al., 2021b) of the open-source statistical software program JASP (JASP Team, 2022; Love et al., 2019). Appendix 5.D demonstrates how to obtain the impartial Bayes factor in these software implementations.

As a final note we like to emphasize that the purpose of the proposed prior distributions is to be impartial with respect to the hypotheses of (in)tolerable misstatement, and consequently obtain Bayes factors that exhibit desirable traits. However, we encourage the auditor to use informed prior distributions when performing Bayesian audit sampling, and to use the impartial Bayes factor mostly as a benchmark procedure in the case no information about the hypotheses is available, if it is costly or difficult to come up with an informed prior distribution, or if the auditor genuinely has pre-existing information that corresponds to impartiality with respect to the hypotheses. Given this application, the impartial Bayes factor can potentially serve as an entry-level Bayesian measure of evidence for auditors unfamiliar to Bayesian inference, and as a more advanced tool to assess the quality and robustness of evidence for auditors who are familiar with it.

5.A Derivations

5.A.1 Equivalence p-value and posterior probability $\mathcal{H}_+$ under the Beta$(1+k,\ n-k)$ posterior

In a binomial test, the one-sided p-value for the hypothesis $\mathcal{H}_0$: $\theta \geq \theta_{max}$ is equal to the posterior probability for $\mathcal{H}_+$: $\theta > \theta_{max}$ given a Beta(1, 0) prior distribution. The one-sided p-value for the null hypothesis $\mathcal{H}_0$: $\theta \geq \theta_{max}$ is the probability of observing k misstatements or less in a sample of n items, given a binomial distribution with misstatement (i.e., error rate) parameter $\theta = \theta_{max}$:

$$p(X \leq k) = \sum_{i=0}^{k} \binom{n}{k} \theta_{max}^{k} (1 - \theta_{max})^{n-k}. \tag{5.A.1}$$

The posterior probability for $\mathcal{H}_+$ given a Beta(1, 0) prior distribution is the probability that the misstatement parameter exceeds θ_{max} given a Beta$(1+k,\ 0+n-k)$ posterior distribution:

$$p(\mathcal{H}_+ \mid n, k) = \int_{\theta_{max}}^{1} \frac{\theta^k (1-\theta)^{n-k-1}}{B(\alpha, \beta)}. \tag{5.A.2}$$

Since the cumulative distribution function for θ_{max} under a Beta(α, β) distribution (i.e., the posterior probability for $\mathcal{H}_-$) is the regularized incomplete beta function $I_{\theta_{max}}(\alpha,\ \beta)$, the posterior probability for $\mathcal{H}_+$ given a Beta(1, 0) prior distribution can also be expressed as

$$p(\mathcal{H}_+ \mid n, k) = 1 - I_{\theta_{max}}(1+k, n-k). \tag{5.A.3}$$

The relationship between the two quantities in Equation 5.A.1 and Equation 5.A.3 is well known and provided in Equation 5.A.4, see Pearson (1934, p. 24), Hartley and Fitch (1951), Raiffa and Schlaifer (1961, p. 271) and Dutka (1981).

$$\sum_{i=0}^{k} \binom{n}{k} \theta_{max}^{k} (1 - \theta_{max})^{n-k} = 1 - I_{\theta_{max}}(1+k, n-k) \tag{5.A.4}$$

5.A.2 The impartial Bayes factor approaches 1 if $\hat{\theta} = \theta_{max}$

For the impartial prior distribution, the prior odds are 1, and, consequently, the Bayes factor is equal to the posterior odds. This means that if the posterior odds are 1, then the Bayes factor is also 1. Because the hypotheses are represented by the range above and below θ_{max}, posterior odds of 1 correspond to a posterior median of θ_{max}. We prove that, in the case of $\hat{\theta} = \theta_{max}$, the limit of the median of any beta posterior distribution is equal to θ_{max}, regardless of the prior parameters α and β. Hence, if $\hat{\theta} = \theta_{max}$ the posterior odds and the Bayes factor approach 1 for all impartial prior distributions.

The median $\theta_{.5}$ of the Beta(α, β) distribution can be approximated as (Kerman, 2011a)

$$\theta_{.5} \approx \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}}. \tag{5.A.5}$$

Following Bayes' rule, a Beta(α, β) prior distribution is updated to a posterior distribution by the data k and n. The posterior distribution is then defined as Beta($\alpha + k$, $\beta + n - k$). However, $\frac{k}{n} = \hat{\theta} = \theta_{max}$ implies $k = \theta_{max} n$ and therefore the posterior distribution can also be written as Beta($\alpha + \theta_{max} n$, $\beta + n - \theta_{max} n$). Plugging these parameters of the posterior distribution into Equation 5.A.5 for the median yields

$$\theta_{.5} = \frac{\alpha + \theta_{max} n - \frac{1}{3}}{\alpha + \theta_{max} n + \beta + n - \theta_{max} n - \frac{2}{3}}. \tag{5.A.6}$$

Simplification of Equation 5.A.6 leads to the following approximation of the posterior median in the case where $\hat{\theta} = \theta_{max}$:

$$\theta_{.5} = \frac{-\frac{1}{3} + \alpha + \theta_{max} n}{-\frac{2}{3} + \alpha + \beta + n}. \tag{5.A.7}$$

Dividing the numerator and the denominator in Equation 5.A.7 by the sample size n yields

$$\theta_{.5} = \frac{-\frac{1}{3n} + \frac{\alpha}{n} + \theta_{max}}{-\frac{2}{3n} + \frac{\alpha}{n} + \frac{\beta}{n} + 1}. \tag{5.A.8}$$

The expressions $-\frac{1}{3n}$, $\frac{\alpha}{n}$, $-\frac{2}{3n}$, and $\frac{\beta}{n}$ in Equation 5.A.8 all tend to zero as n tends to infinity. Hence, in the case of $\hat{\theta} = \theta_{max}$ the limit of the posterior median of any Beta(α, β) posterior distribution is θ_{max} regardless of the prior parameters α and β, see Equation 5.A.9.

$$\lim_{n \to \infty} \theta_{.5} = \theta_{max} \tag{5.A.9}$$

5.A.3 Relative reduction in sample size compared to NHST when $k = 0$

The impartial prior distribution can result in a relative reduction of $\frac{\ln(\frac{1}{2})}{\ln(\text{ACR})}$ compared to NHST procedures. When no errors are expected in the sample, the NHST procedure for planning the sample size comes down to finding the smallest integer n for which $(1 - \theta_{max})^n \leq \text{ACR}$. This implies that $n = \frac{\ln(\text{ACR})}{\ln(1 - \theta_{max})}$. By using the impartial prior distribution, the posterior distribution is Beta(1,

$\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k$) instead of Beta(1, $n - k$), since Beta(1, 0) is the implicit prior in case of NHST. This means that the sample size is reduced by $\Delta n = \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}$ in absolute terms. Dividing this absolute reduction by the NHST sample size n results in the relative sample size reduction of $\frac{\ln(\frac{1}{2})}{\ln(\text{ACR})}$.

5.A.4 Relationship impartial Bayes factor and p-value

For $\theta_{exp} = 0$ and $\hat{\theta} = 0$ ($k = 0$), the relationship between the one-sided p-value and the impartial Bayes factor is $BF_{-+} = \frac{2-p}{p}$ for all n. Moreover, for $\hat{\theta} = 1$ ($k = n$) the relationship between the p-value and the impartial Bayes factor goes to $BF_{-+} = \frac{1-p}{p}$ as n tends to infinity.

To prove this, we first determine the exact relationship between the impartial Bayes factor and the one-sided p-value. For the impartial prior distribution, the prior odds are 1, and, consequently, the impartial Bayes factor is the ratio of posterior odds:

$$BF_{+-} = \frac{p(\mathcal{H}_+ \mid y)}{1 - p(\mathcal{H}_+ \mid y)}. \tag{5.A.10}$$

To determine the relationship between the impartial Bayes factor and the p-value, we can introduce $p(X \leq k)$ (i.e., the one-sided p-value) into the right side of Equation 5.A.10, resulting in Equation 5.A.11.

$$\begin{aligned}
BF_{+-} &= \frac{p(\mathcal{H}_+ \mid y)p(X \leq k)}{(1 - p(\mathcal{H}_+ \mid y))p(X \leq k)} \\
BF_{+-} &= \frac{p(\mathcal{H}_+ \mid y)p(X \leq k)}{p(X \leq k) - p(\mathcal{H}_+ \mid y)p(X \leq k)} \\
BF_{+-} &= \frac{p(X \leq k)}{\frac{p(X \leq k) - p(\mathcal{H}_+ \mid y)p(X \leq k)}{p(\mathcal{H}_+ \mid y)}} \\
BF_{+-} &= \frac{p(X \leq k)}{\frac{p(X \leq k)}{p(\mathcal{H}_+ \mid y)} - p(X \leq k)}
\end{aligned} \tag{5.A.11}$$

Note that $p(X \leq k) > 0$ and $p(\mathcal{H}_+ \mid y) > 0$. Thus, the exact relationship between the one-sided p-value and the impartial Bayes factor reads $BF_{+-} = \frac{p(X \leq k)}{\frac{p(X \leq k)}{p(\mathcal{H}_+ \mid y)} - p(X \leq k)}$ or $BF_{-+} = \frac{\frac{p(X \leq k)}{p(\mathcal{H}_+ \mid y)} - p(X \leq k)}{p(X \leq k)}$.

Since $p(X \leq k)$ is equal to the posterior probability for $\mathcal{H}_+$ using a Beta(1, 0) prior distribution (Equation 5.A.4), the factor $\frac{p(X \leq k)}{p(\mathcal{H}_+ \mid y)}$ is a weighting factor that weighs the probability of $\mathcal{H}_+$ under a Beta($1 + k$, $0 + n - k$) posterior distribution with the posterior probability of $\mathcal{H}_+$ given the specified impartial prior distribution. To determine the bounds of the relationship in Equation 5.A.11, we calculate the weighting factor for the two most extreme cases: $k = 0$ ($\hat{\theta} = 0$) and $k = n$ ($\hat{\theta} = 1$).

We first determine the lower bound of Equation 5.A.11. Concretely, for $\theta_{exp} = 0$ (the impartial Beta($1, \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}$) prior with a mode at zero) the weighting factor for a given sample size n and number of misstatements k is equal to

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{\sum_{i=0}^{k} \binom{n}{k} \theta^k (1-\theta)^{n-k}}{\int_{\theta_{max}}^{1} \frac{\theta^k (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k - 1}}{B(1+k, \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k)} d\theta}. \tag{5.A.12}$$

Filling in $k = 0$ and simplifying the result leads to a weighting factor of 2.

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\int_{\theta_{max}}^{1} \frac{\theta^k (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k - 1}}{B(1+k, \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k)} d\theta}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\int_{\theta_{max}}^{1} \frac{1}{B(1+k, \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k)} (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} d\theta}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{1}{B(1+k, \frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - k)} \int_{\theta_{max}}^{1} (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} d\theta}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{1}{\int_0^1 (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}} d\theta} \int_{\theta_{max}}^{1} (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} d\theta}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{\theta_{max}}^{1} (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} d\theta}{\int_0^1 (1-\theta)^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} d\theta}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} \frac{d}{d\theta}(1-\theta) d\theta}{\int_1^0 u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} \frac{d}{d\theta}(1-\theta) d\theta}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} (\frac{d}{d\theta}(1) - \frac{d}{d\theta}(\theta)) d\theta}{\int_1^0 u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} (\frac{d}{d\theta}(1) - \frac{d}{d\theta}(\theta)) d\theta}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} (0 - \frac{d}{d\theta}(\theta)) d\theta}{\int_1^0 u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})} + n - 1} (0 - \frac{d}{d\theta}(\theta)) d\theta}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}(-1)d\theta}{\int_{1}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}(-1)d\theta}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}(-1)du}{\int_{1}^{0} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}(-1)du}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}{\int_{1}^{0} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}{-\int_{0}^{1} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}{-(\int_{0}^{1} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du)}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{\int_{1-\theta_{max}}^{0} -u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}{\int_{0}^{1} u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n-1}du}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{-[\frac{u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}}{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}]_{1-\theta_{max}}^{0}+c}{[\frac{u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}}{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}]_{0}^{1}+c}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{-[\frac{1}{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}]_{1-\theta_{max}}^{0}}{[\frac{1}{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}u^{\frac{\ln(\frac{1}{2})}{\ln(1-\theta_{max})}+n}]_{0}^{1}}}$$

$$\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} = \frac{(1-\theta_{max})^n}{\frac{-[\frac{1}{\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}}u^{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}+n}]_{1-\theta_{max}}^{0}}{[\frac{1}{\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}+n}u^{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}+n}]_{0}^{1}}}$$

$$
\begin{aligned}
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= \frac{(1-\theta_{max})^n}{\frac{-[\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)} u^{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}+n}]_{1-\theta_{max}}^{0}}{[\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)} u^{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}+n}]_{0}^{1}}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= \frac{(1-\theta_{max})^n}{\frac{-[\ln(1-\theta_{max})\frac{1}{n\ln(1-\theta_{max})-\ln(2)} u^{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}+n}]_{1-\theta_{max}}^{0}}{[\ln(1-\theta_{max})\frac{1}{n\ln(1-\theta_{max})-\ln(2)} u^{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}+n}]_{0}^{1}}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= \frac{(1-\theta_{max})^n}{\frac{\frac{\ln(1-\theta_{max})(1-\theta_{max})^{\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}}}{n\ln(1-\theta_{max})-\ln(2)}}{\frac{n\ln(1-\theta_{max})-\ln(2)}{\frac{\ln(1-\theta_{max})}{n\ln(1-\theta_{max})-\ln(2)}}}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= \frac{(1-\theta_{max})^n}{\frac{(\ln(1-\theta_{max})(1-\theta_{max})^{\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}})(n\ln(1-\theta_{max})-\ln(2))}{(n\ln(1-\theta_{max})-\ln(2))\ln(1-\theta_{max})}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= \frac{(1-\theta_{max})^n}{\frac{\ln(1-\theta_{max})(1-\theta_{max})^{\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}}}{\ln(1-\theta_{max})}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= \frac{(1-\theta_{max})^n}{(1-\theta_{max})^{\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= (1-\theta_{max})^{n-\frac{n\ln(1-\theta_{max})-\ln(2)}{\ln(1-\theta_{max})}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= (1-\theta_{max})^{\frac{n\ln(1-\theta_{max})-(n\ln(1-\theta_{max})-\ln(2))}{\ln(1-\theta_{max})}} \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= (1-\theta_{max})^{\frac{\ln(2)}{\ln(1-\theta_{max})}} \\
\ln(\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)}) &= \frac{\ln(2)}{\ln(1-\theta_{max})}\ln(1-\theta_{max}) \\
\ln(\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)}) &= \ln(2) \\
\frac{p(X \leq k)}{p(\mathcal{H}_+ \,|\, y)} &= 2
\end{aligned}
\tag{5.A.13}
$$

Plugging Equation 5.A.13 into Equation 5.A.12 shows that if $k = 0$, for all n, the relationship between the one-sided p-value and the impartial Bayes factor is $BF_{+-} = \frac{p(X \leq k)}{2-p(X \leq k)}$ or $BF_{-+} = \frac{2-p(X \leq k)}{p(X \leq k)}$.

On the other hand, the probability of observing $k = n$ misstatements or fewer in a sample of n items (i.e., the one-sided p-value) is equal to 1. Moreover, the posterior probability for $\mathcal{H}_+$ given an impartial prior distribution approaches 1 as the sample size goes to infinity (an implication of Equation 5.4.3). In this case, the limit of the weighting factor is 1.

$$\lim_{n \to \infty} \frac{p(X \leq k)}{p(\mathcal{H}_+ \mid y)} = 1 \tag{5.A.14}$$

This implies that, in this case, the relationship between the p-value and the impartial Bayes factor goes to $BF_{+-} = \frac{p(X \leq k)}{1 - p(X \leq k)}$ or $BF_{-+} = \frac{1 - p(X \leq k)}{p(X \leq k)}$ as the sample size increases.

Consequently, all possible impartial Bayes factors BF_{-+} lie on or within the upper bound of $\frac{2-p}{p}$ and the lower bound of $\frac{1-p}{p}$, see Figure 5.8.

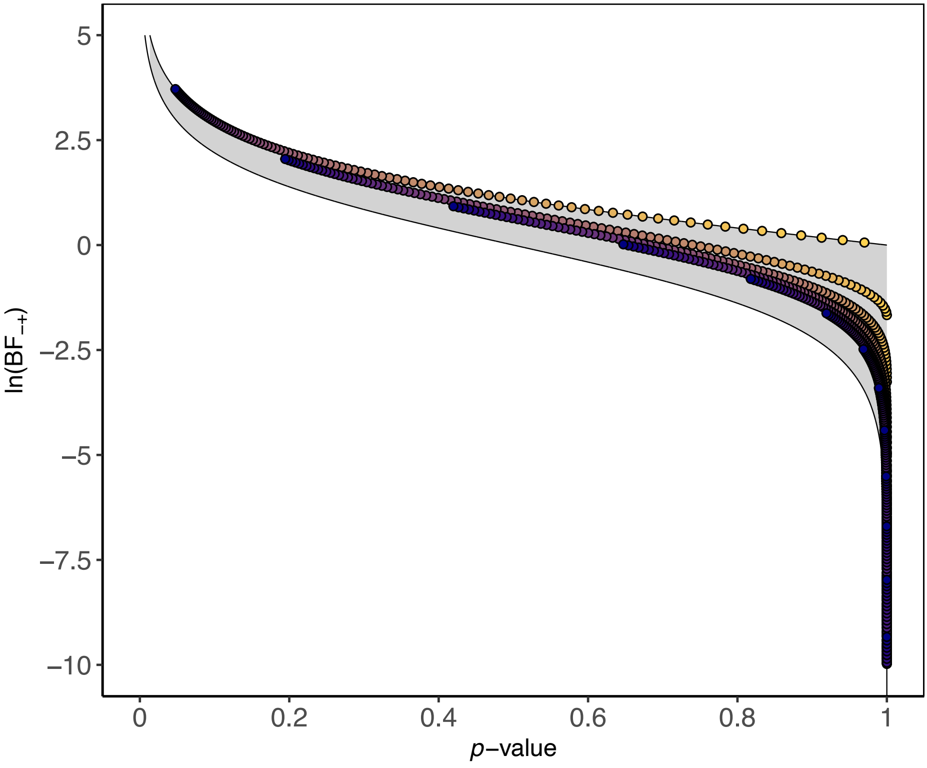


Figure 5.8: Comparison of one-sided p-values and natural logarithmic impartial Bayes factors for a performance materiality $\theta_{max} = 0.03$. Darker (blue) points correspond to higher sample sizes. The black lines represent the upper bound $\ln(\frac{2-p}{p})$ and the lower bound $\ln(\frac{1-p}{p})$ of the relationship between the one-sided p-value and the impartial Bayes factor.

5.A.5 Parameters of the impartial prior distribution when $\alpha = 1$ and $\beta > 1$

The mode θ_{exp} of the impartial beta prior distribution, given $\alpha = 1$ and $\beta > 1$, is zero. Furthermore, the median θ_{max} of the impartial beta prior distribution, given $\alpha = 1$ and $\beta > 1$, is (Kerman, 2011a)

$$\theta_{max} = 1 - 2^{-\frac{1}{\beta}}. \tag{5.A.15}$$

Hence, to define the β parameter of the impartial prior distribution given a specific value of θ_{max}, we solve Equation 5.A.15 for β.

$$\begin{aligned}
\theta_{max} + 2^{-\frac{1}{\beta}} &= 1 \\
2^{-\frac{1}{\beta}} &= 1 - \theta_{max} \\
-\frac{1}{\beta}\ln(2) &= \ln(1 - \theta_{max}) \\
-\frac{\ln(2)}{\beta} &= \ln(1 - \theta_{max}) \\
-\ln(2) &= \ln(1 - \theta_{max})\beta \\
\beta &= \frac{-\ln(2)}{\ln(1 - \theta_{max})} \qquad (5.A.16)
\end{aligned}$$

Note that $-\ln(2) = \ln(\frac{1}{2})$, and thus a final simplification of Equation 5.A.16 leads to the following closed formula for β.

$$\beta = \frac{\ln(\frac{1}{2})}{\ln(1 - \theta_{max})} \tag{5.A.17}$$

5.A.6 Parameters of the impartial prior distribution when $\alpha > 1$ and $\beta > 1$

The mode θ_{exp} of the impartial beta prior distribution, given $\alpha > 1$ and $\beta > 1$, is given by

$$\theta_{exp} = \frac{\alpha - 1}{\alpha + \beta - 2}. \tag{5.A.18}$$

Furthermore, the median θ_{max} of the impartial beta prior distribution, given $\alpha > 1$ and $\beta > 1$, is approximated by (Kerman, 2011a)

$$\theta_{max} \approx \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}}. \tag{5.A.19}$$

Hence, to define the α and β parameters of the impartial prior distribution for any combination of θ_{max} and θ_{exp}, we can solve the system of equations comprised of Equation 5.A.18 and Equation 5.A.19 for α and β. First, we rewrite Equation 5.A.19 for the median to Equation 5.A.20 by isolating α.

$$\begin{aligned}
\theta_{max} &\approx \frac{\frac{3\alpha-1}{3}}{\frac{3\alpha+3\beta-2}{3}} \\
\theta_{max} &\approx \frac{3\alpha-1}{3\alpha+3\beta-2} \\
\theta_{max}(3\alpha+3\beta-2) &\approx 3\alpha-1 \\
3\theta_{max}\alpha+3\theta_{max}\beta-2\theta_{max} &\approx 3\alpha-1 \\
3\theta_{max}\alpha-3\alpha &\approx 2\theta_{max}-3\theta_{max}\beta-1 \\
3\alpha(\theta_{max}-1) &\approx 2\theta_{max}-3\theta_{max}\beta-1 \\
\alpha &\approx \frac{2\theta_{max}-3\theta_{max}\beta-1}{3(\theta_{max}-1)} \qquad (5.A.20)
\end{aligned}$$

Equation 5.A.20 can then be used to substitute α in Equation 5.A.18.

$$\theta_{exp} \approx \frac{\frac{2\theta_{max}-3\theta_{max}\beta-1}{3(\theta_{max}-1)}}{\frac{2\theta_{max}-3\theta_{max}\beta-1}{3(\theta_{max}-1)}+\beta-2} \qquad (5.A.21)$$

We then rewrite Equation 5.A.21 for the mode to Equation 5.A.22 by isolating β.

$$\begin{aligned}
\theta_{exp} &\approx \frac{\frac{2\theta_{max}-3\theta_{max}\beta-1-3(\theta_{max}-1)}{3(\theta_{max}-1)}}{\frac{2\theta_{max}-3\theta_{max}\beta-1+3\beta(\theta_{max}-1)-6(\theta_{max}-1)}{3(\theta_{max}-1)}} \\
\theta_{exp} &\approx \frac{\frac{2-3\theta_{max}\beta-\theta_{max}}{3(\theta_{max}-1)}}{\frac{5-3\beta-4\theta_{max}}{3(\theta_{max}-1)}} \\
\theta_{exp} &\approx \frac{2-3\theta_{max}\beta-\theta_{max}}{5-3\beta-4\theta_{max}} \\
\theta_{exp}(5-3\beta-4\theta_{max}) &\approx 2-3\theta_{max}\beta-\theta_{max} \\
5\theta_{exp}-3\theta_{exp}\beta-4\theta_{exp}\theta_{max} &\approx 2-3\theta_{max}\beta-\theta_{max} \\
3\theta_{max}\beta-3\theta_{exp}\beta &\approx 2-\theta_{max}+4\theta_{exp}\theta_{max}-5\theta_{exp} \\
3\beta(\theta_{max}-\theta_{exp}) &\approx 2-\theta_{max}+4\theta_{exp}\theta_{max}-5\theta_{exp} \\
\beta &\approx \frac{2-\theta_{max}+4\theta_{exp}\theta_{max}-5\theta_{exp}}{3(\theta_{max}-\theta_{exp})} \qquad (5.A.22)
\end{aligned}$$

Thus, a final simplification of Equation 5.A.22 leads to a definitive expression for β.

$$\beta \approx \frac{2+\theta_{max}(4\theta_{exp}-1)-5\theta_{exp}}{3(\theta_{max}-\theta_{exp})} \qquad (5.A.23)$$

Since β is solely determined by θ_{max} and θ_{exp}, we can plug Equation 5.A.22 into Equation 5.A.20, yielding an expression for α that is also solely determined by θ_{max} and θ_{exp}.

$$\alpha \approx \frac{2\theta_{max} - 3\theta_{max}\frac{2-\theta_{max}+4\theta_{exp}\theta_{max}-5\theta_{exp}}{3(\theta_{max}-\theta_{exp})} - 1}{3(\theta_{max} - 1)} \tag{5.A.24}$$

Hence, a final simplification of Equation 5.A.24 leads to a definitive expression for α.

$$\begin{aligned}
\alpha &\approx \frac{\theta_{exp} - 4\theta_{max}^2\theta_{exp} + 5\theta_{max}\theta_{exp} + 3\theta_{max}^2 - 2\theta_{exp}\theta_{max} - 3\theta_{max}}{3(\theta_{max} - 1)(\theta_{max} - \theta_{exp})} \\
\alpha &\approx \frac{-(\theta_{max} - 1)(4\theta_{max}\theta_{exp} - 3\theta_{max} + \theta_{exp})}{3(\theta_{max} - 1)(\theta_{max} - \theta_{exp})} \\
\alpha &\approx \frac{-(4\theta_{max}\theta_{exp} - 3\theta_{max} + \theta_{exp})}{3(\theta_{max} - \theta_{exp})} \\
\alpha &\approx \frac{3\theta_{max} - 4\theta_{max}\theta_{exp} - \theta_{exp}}{3(\theta_{max} - \theta_{exp})} \\
\alpha &\approx \frac{3\theta_{max} - \theta_{exp}(4\theta_{max} + 1)}{3(\theta_{max} - \theta_{exp})}
\end{aligned} \tag{5.A.25}$$

5.B Bayes factors using the gamma distribution

The procedure discussed in this chapter requires that the auditor specifies a beta prior distribution for θ in the range [0, 1]. The beta distribution is generally used for attributes sampling, where an item can be judged as correct or incorrect and the likelihood is therefore considered binomial[6]. However, it has been argued that for monetary unit sampling (MUS), where items can be judged as partially correct, a gamma prior distribution with shape parameter α and rate (inverse scale) parameter β combined with a Poisson likelihood is a more fitting setup (Stewart, 2013). The probability density for the gamma distribution with parameters α, β is defined in Equation 5.B.1, in which $\Gamma(\alpha)$ is the gamma function. Using a gamma prior distribution in a MUS context, the procedure for calculating the Bayes factor is the same but the calculations for the α and β parameters of the collection of impartial prior distributions are different.

$$p(\theta; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \tag{5.B.1}$$

As an example, suppose that for a population of $m = 1,000,000$ monetary units a performance materiality of $\theta_{max} = 30,000$ monetary units applies. The hypotheses that the auditor wants to test are the hypothesis that the amount of

[6]Note that the beta distribution can also be used for monetary unit sampling.

misstatement in the population is lower than the performance materiality $\mathcal{H}_{-}$: $\theta < 30,000$ versus the hypothesis that the amount of misstatement in the population is higher than the performance materiality $\mathcal{H}_{+}$: $\theta > 30,000$.

To test these hypotheses, the auditor can specify a Gamma(α, β) prior distribution for θ in the range [0; m]. The prior probability of the hypothesis $\mathcal{H}_{-}$: $\theta < \theta_{max}$ is represented by the total probability in the range [0; θ_{max}) of the prior distribution. Vice versa, the prior probability of the hypothesis $\mathcal{H}_{+}$: $\theta > \theta_{max}$ is represented by the total probability in the range (θ_{max}; m] of the prior distribution. For the prior distribution to be impartial with respect to the hypotheses $\mathcal{H}_{-}$ and $\mathcal{H}_{+}$, the total probability in the range [0; θ_{max}) must be equal to the total probability in the range (θ_{max}; m].

5.B.1 Zero expected most likely misstatement

A gamma prior distribution that has this property, under the assumption of zero expected misstatements (i.e., $\alpha = 1$), is the Gamma($\alpha = 1$, $\beta = -\frac{\ln(\frac{1}{2})}{\theta_{max}}$) prior distribution. Calculating the α and β parameters of this prior distribution for a performance materiality $\theta_{max} = 30,000$ results in $\alpha = 1$ and $\beta = 0.00002310491$.

In MUS, the sample is generally evaluated using the sum of the taints of the items (American Institute of Certified Public Accountants (AICPA), 2019, Appendix C). The taint of an item is the proportional error of that item $t_i = \frac{y_i - x_i}{y_i}$. Suppose that $n = 50$ items are selected, and a single item is found to be overstated. This item has a recorded value of $y_i = 5,000$ and a true value $x_i = 2,000$, yielding a taint $t_i = 0.6$. Thus, a total of $k = \sum_{i=1}^{n} t_i = 0.6$ taints were found in the sample. The resulting posterior distribution is of the form Gamma($\alpha = 1 + k$, $\beta = -\frac{\ln(\frac{1}{2})}{\theta_{max}} + \frac{1}{m/n}$), and the posterior parameters are $\alpha = 1.6$ and $\beta = 0.00007310491$.

The Bayes factor in this example is $BF_{-+} = 3.03$, implying that these sample outcomes are 3.03 times more likely to occur under $\mathcal{H}_{-}$ than under $\mathcal{H}_{+}$. As is the case for attributes sampling, under the assumption of no expected misstatements the impartial Bayes factor is easy to calculate using summary statistics from the sample as it is solely a function of the performance materiality, the total value of the population, and the sample outcomes, see Equation 5.B.2.

$$BF_{-+}(\theta_{max}, m, n, k) = \frac{\int_{0}^{\theta_{max}} p(\theta; 1 + k, -\frac{\ln(\frac{1}{2})}{\theta_{max}} + \frac{1}{m/n}) d\theta}{\int_{\theta_{max}}^{m} p(\theta; 1 + k, -\frac{\ln(\frac{1}{2})}{\theta_{max}} + \frac{1}{m/n}) d\theta} \tag{5.B.2}$$

5.B.2 Non-zero expected most likely misstatement

Incorporating the expected most likely misstatement θ_{exp} into the impartial prior distribution can only be achieved using an iterative procedure since, in the case of $\alpha > 1$, the gamma distribution does not have an approximate or closed-form expression for the median. However, its mode can be expressed as $\theta_{exp} = \frac{\alpha - 1}{\beta}$, so that β can be expressed in terms of α as $\beta = \frac{\alpha - 1}{\theta_{exp}}$. Therefore, to approximate the values of α and β one can increase the value of α (starting at $\alpha = 1$) until the median of the prior distribution is close to—or equal to—θ_{max}.

Suppose that the auditor expects a most likely error of $\theta_{exp} = 15,000$. Using the procedure described above, the parameters for the gamma prior distribution can be determined as $\alpha = 1.68095$ and $\beta = 0.00004539667$. The mode of this prior distribution is equal to 15,000 and its median is equal to 30,000. After seeing the sample outcomes of $n = 50$ and $k = 0.6$, the posterior parameters are $\alpha = 2.28095$ and $\beta = 0.00009539667$. The Bayes factor is $BF_{-+} = 2.536$, implying that the data are 2.536 times more likely to occur under the hypothesis $\mathcal{H}_-$ than under the hypothesis $\mathcal{H}_+$.

5.C Bayes factor tables

This appendix presents impartial Bayes factors (BF_{-+}) using the beta distribution for combinations of the sample size, n, and the observed misstatements, k, against a performance materiality of θ_{max}.

5.C.1 Impartial Bayes factors for a performance materiality of 10 percent

This table presents Bayes factors in favor of tolerable misstatement based on equal prior probabilities and zero expected errors for a performance materiality of 10 percent.

Table 5.3: Impartial Bayes factors in favor of tolerable misstatement for a performance materiality of 10 percent.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$n = 20$	15.45	3.16	1.02	0.37	0.14	0.05
$n = 25$	26.86	5.18	1.66	0.64	0.26	0.1
$n = 30$	46.18	8.32	2.6	1.02	0.43	0.18
$n = 35$	78.9	13.22	3.98	1.55	0.68	0.3
$n = 40$	134.31	20.91	6.02	2.31	1.02	0.47
$n = 45$	228.15	33.04	9.04	3.37	1.48	0.7
$n = 50$	387.07	52.26	13.54	4.89	2.12	1.02
$n = 55$	656.19	82.8	20.29	7.05	3	1.43
$n = 60$	1111.96	131.53	30.48	10.15	4.21	1.99
$n = 65$	1883.81	209.52	45.96	14.64	5.89	2.74
$n = 70$	3190.94	334.68	69.56	21.18	8.24	3.76
$n = 75$	5404.57	536.1	105.74	30.75	11.53	5.13
$n = 80$	9153.39	861.01	161.44	44.86	16.2	7.01
$n = 85$	15502.03	1386.24	247.53	65.75	22.84	9.59
$n = 90$	26253.52	2237.05	381.1	96.85	32.35	13.16
$n = 95$	44461.27	3617.78	589.06	143.38	46.06	18.12
$n = 100$	75296.24	5862.32	913.89	213.31	65.91	25.06
$n = 125$	1048856	67147.91	8618.95	1662.07	427.35	136.96
$n = 150$	1.46e+07	794128.5	86567.11	14188.07	3105.98	850.65
$n = 200$	2.83e+09	1.18e+08	9908457	1247503	209910.8	44244.4

5.C.2 Impartial Bayes factors for a performance materiality of 5 percent

This table presents Bayes factors in favor of tolerable misstatement based on equal prior probabilities and zero expected errors for a performance materiality of 5 percent.

Table 5.4: Impartial Bayes factors in favor of tolerable misstatement for a performance materiality of 5 percent.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$n = 20$	4.58	1.02	0.31	0.09	0.03	0.01
$n = 25$	6.21	1.38	0.43	0.14	0.04	0.01
$n = 30$	8.32	1.83	0.59	0.21	0.07	0.02
$n = 35$	11.04	2.39	0.79	0.29	0.1	0.03
$n = 40$	14.56	3.08	1.02	0.39	0.15	0.05
$n = 45$	19.11	3.93	1.3	0.51	0.2	0.08
$n = 50$	24.99	4.99	1.64	0.65	0.27	0.11
$n = 55$	32.59	6.29	2.05	0.82	0.35	0.15
$n = 60$	42.41	7.92	2.54	1.02	0.44	0.19
$n = 65$	55.1	9.93	3.14	1.26	0.55	0.25
$n = 70$	71.51	12.44	3.85	1.54	0.68	0.31
$n = 75$	92.7	15.56	4.72	1.87	0.84	0.39
$n = 80$	120.1	19.45	5.76	2.26	1.02	0.48
$n = 85$	155.5	24.3	7.03	2.72	1.23	0.59
$n = 90$	201.26	30.37	8.57	3.27	1.47	0.71
$n = 95$	260.39	37.95	10.43	3.93	1.75	0.85
$n = 100$	336.81	47.44	12.7	4.7	2.08	1.02
$n = 125$	1216.81	145.9	34.12	11.45	4.78	2.29
$n = 150$	4389.25	456.03	93.6	28.13	10.83	4.93
$n = 200$	57056	4661.45	758.92	184.08	58.79	22.99

5.C.3 Impartial Bayes factors for a performance materiality of 1 percent

This table presents Bayes factors in favor of tolerable misstatement based on equal prior probabilities and zero expected errors for a performance materiality of 1 percent.

Table 5.5: Impartial Bayes factors in favor of tolerable misstatement for a performance materiality of 1 percent.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$n = 20$	1.45	0.29	0.06	0.01	0	0
$n = 25$	1.57	0.32	0.07	0.02	0	0
$n = 30$	1.7	0.35	0.08	0.02	0	0
$n = 35$	1.84	0.39	0.1	0.02	0	0
$n = 40$	1.99	0.42	0.11	0.02	0	0
$n = 45$	2.14	0.46	0.12	0.03	0.01	0
$n = 50$	2.31	0.5	0.13	0.03	0.01	0
$n = 55$	2.48	0.54	0.15	0.04	0.01	0
$n = 60$	2.66	0.59	0.16	0.04	0.01	0
$n = 65$	2.84	0.63	0.18	0.05	0.01	0
$n = 70$	3.04	0.68	0.2	0.05	0.01	0
$n = 75$	3.25	0.73	0.21	0.06	0.02	0
$n = 80$	3.47	0.78	0.23	0.07	0.02	0
$n = 85$	3.7	0.84	0.25	0.07	0.02	0
$n = 90$	3.94	0.9	0.27	0.08	0.02	0.01
$n = 95$	4.2	0.96	0.29	0.09	0.03	0.01
$n = 100$	4.46	1.02	0.32	0.1	0.03	0.01
$n = 125$	6.02	1.37	0.44	0.15	0.05	0.01
$n = 150$	8.03	1.81	0.6	0.22	0.08	0.02
$n = 200$	13.93	3.02	1.02	0.4	0.16	0.06

5.D Software implementation

To promote the use of impartial Bayes factors in the audit practice, we discuss two free and open-source software packages that can be used to obtain the impartial Bayes factor from summary statistics of an audit sample. First, we show how to obtain the impartial Bayes factor in R through the package 'jfa' (Derks, 2022). Next, we show how to obtain the impartial Bayes factor using the statistical software program JASP (JASP Team, 2022; Love et al., 2019) through its Audit module (Derks et al., 2021b).

5.D.1 Implementation in R

In R, the impartial Bayes factor can be obtained via the code below. The auditor can use the 'evaluation' function from the 'jfa' package together with the value of the performance materiality, $\theta_{max} = 0.03$, the sample size, $n = 50$, and the observed misstatements, $k = 1$, to reproduce the impartial Bayes factor for the example discussed in the chapter.

```
library(jfa) # Load the jfa package

m <- 0.03 # Materiality
n <- 50   # Sample size
k <- 1    # Observed misstatements

# Create the impartial prior distribution using the auditPrior() function
```

```
prior <- auditPrior(method = "impartial", materiality = m,
                    likelihood = "binomial")

# Compute the impartial Bayes factor using the evaluation() function
evaluation(materiality = m, x = k, n = n, prior = prior)

#   Bayesian Audit Sample Evaluation
#
# data: 1 and 50
# number of errors = 1, number of samples = 50, taint = 1, BF10 = 1.8218
# alternative hypothesis: true misstatement rate is less than 0.03
# 95 percent credible interval:
#  0.00000000 0.06354809
# estimate:
#  0.01393601
# estimates obtained via method "binomial" + "prior"
```

5.D.2 Implementation in JASP

This example can be reproduced in JASP via the Bayesian evaluation analysis in the Audit module. The interface of the Bayesian evaluation analysis in JASP and the options required to reproduce this example are displayed in Figure 5.9.

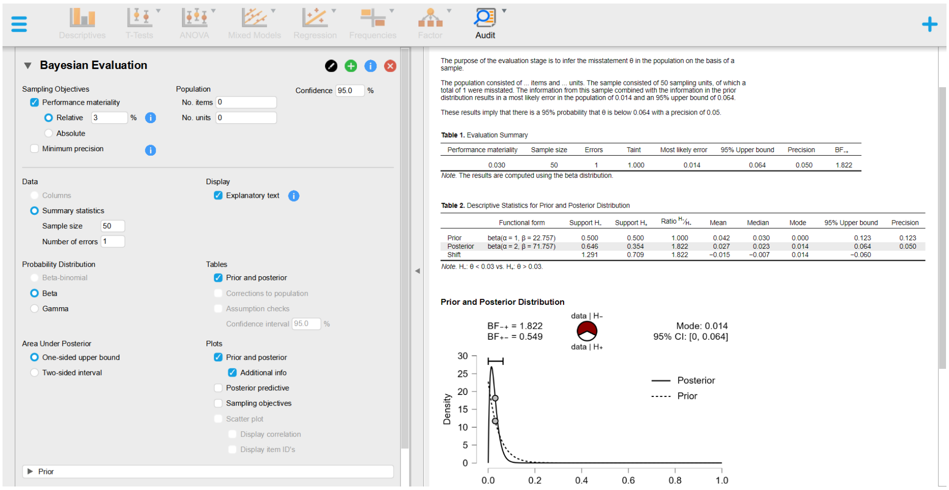


Figure 5.9: Snapshot of the interface of the Bayesian evaluation analysis in JASP and the options required to compute the impartial Bayes factor for $\theta_{max} = 0.03$, $n = 50$ and $k = 1$.

The Bayesian evaluation analysis allows for evaluation of audit samples on the basis of summary statistics only, and it is therefore not required to load a data set. After opening the analysis, the auditor can check the option 'Performance materiality' and specify θ_{max} as 3 percent. Next, the auditor can fill in the sample outcomes $n = 50$ and $k = 1$ as the 'Sample size' and the 'Number of errors' respectively. To use the impartial beta prior distribution, the auditor can navigate to the 'Prior' section and select the option 'Impartial'. The results of the statistical analysis are automatically computed.

The table at the top of the output shows the sample size, n, and number of found misstatements, k, in the sample. By default, the total tainting (i.e., proportional error) in the sample is shown alongside the most likely error in the population (i.e., the mode of the posterior distribution). The last column in the table displays the 95 percent upper credible bound (i.e., the 95^{th} percentile of the posterior distribution), in this case 0.064.

The Bayes factor in favor of the hypothesis of tolerable misstatement (BF_{-+}) is displayed by default. Given the sample outcomes and the performance materiality in this example, the impartial Bayes factor in favor of tolerable misstatement is 1.822. Under the 'Tables' and 'Plots' sections the auditor is also able to request a table and a figure containing descriptive information of the prior and posterior distribution by clicking 'Prior and posterior'.

Part III

Software Implementation

Chapter 6

JASP for Audit: Statistical Tools for the Auditing Practice

Abstract

Properly setting up, conducting and documenting a statistical audit sample is not an easy task. This chapter introduces JASP for Audit, open-source and user-friendly software specifically designed for auditors to make the statistical components of an audit easier. Firstly, an overview of the functionality of JASP for Audit is provided. Secondly, four distinguishing features of the software are discussed and their benefits for the auditing practice are explained. Thirdly, the software is demonstrated by means of three examples concerning, respectively, testing internal controls, tests of details and tax auditing. The chapter concludes with recommendations for the use of JASP for Audit in the auditing practice.

Keywords: audit, Bayesian, open-source, sampling, software.

6.1 Introduction

Audit sampling has long been part of the audit process, both for testing internal controls and for substantive testing (Power, 1992; de Swart et al., 2013). Although current developments within data analysis offer auditors the possibility to integrally audit some populations using data (Brown-Liburd et al., 2015; Salijeni et al., 2019), it often remains more effective and efficient to use a sample for populations for which the correct values are not readily available or for which it can be wrongly assumed that the correct values are available (van Batenburg, 2018b,a). After all, using a statistical sample auditors can make a probability

Parts of this chapter are published as Derks, K., de Swart, J., van Batenburg, P., Wagenmakers, E.–J., & Wetzels, R. (2021). JASP for Audit: Bayesian tools for the auditing practice. *Journal of Open Source Software*, *6*(68), 2733. doi: `https://doi.org/10.21105/joss.02733` and under review for publication (in dutch) as Derks, K., de Swart, J., & Wetzels, R. (2022) Open-source software als brug tussen de auditor en de statisticus.

statement about an entire population on the basis of only a small observed part of this population. In addition, developments in the field of statistics offer more opportunities to deal more effectively and efficiently with available prior information and (sample) data than before (Stewart, 2013; de Swart et al., 2013; Derks et al., 2021a, 2022b). As a result, audit sampling is becoming more targeted and less expensive (van Batenburg, 2018b); an example of how analytical procedures and audit sampling are not mutually exclusive, but rather complementary (Yoon and Pearce, 2021). Therefore, audit sampling as a tool to contribute to sufficient and appropriate audit evidence will not disappear from audit practice any time soon. On the contrary, the combination of analytical procedures and audit sampling is expected to become increasingly important in the field (van Batenburg, 2018a; Kogan et al., 2019).

Despite the major role of audit sampling in the auditing practice (Van Der Nest et al., 2015; Christensen et al., 2015), properly designing, implementing and documenting a statistical sample is not an easy task (see for example ISA 530 about sampling (International Auditing and Assurance Standards Board (IAASB), 2018)). Three obstacles can be identified herein that pose a hurdle for auditors. First, auditors must possess sufficient knowledge of statistical theory to set up and conduct a statistical sample in a responsible manner (e.g., Stewart, 2012), while they are not always adequately trained to possess, let alone maintain, the expert knowledge of a statistician (Symon, 1974). Second, auditors need access to software that is easy to use and performs analyses in accordance with international auditing standards (Schouten, 2007; Binck, 2012; Ahmi and Kent, 2013; Bradford et al., 2020). Last, the documentation and interpretation of statistical results is not always easy. For example, it can be difficult to formulate the statistical statement about the misstatement in a population associated with a frequentist confidence interval (Hoekstra et al., 2014), or to ascertain the amount of evidence for acceptance or rejection of a financial population from a p-value (see Chapter 4). All three obstacles lead to the fact that planning and evaluating a statistical sample is often outsourced to auditors with more specialized knowledge of statistics (Brazel and Agoglia, 2007).

With the current developments in the field of data analysis, external auditors are increasingly encountering the possibilities of using statistical methods and automated processes to gather audit evidence (e.g., Brown-Liburd and Vasarhelyi, 2015; Appelbaum et al., 2017; Boersma et al., 2020). For internal auditors, a similar impact of these developments applies (de Swart et al., 2016). In line with this trend, auditors are therefore increasingly expected to reduce their dependence on statisticians by developing their own data analytic skills (Li, 2022). In addition, the growing use of analytical procedures will also increase the complexity of the statistical sample (Appelbaum et al., 2017). These developments have the effect that auditors today are expected to be qualified in multiple areas of statistics, which are not necessarily part of an auditor's expertise (Commissie Eindtermen Accountantsopleiding (CEA), 2020). The open-source software JASP for Audit helps auditors to correctly set up, perform and document a statistical sample, thus bridging the gap between the auditor and the statistician.

In this chapter we introduce JASP for Audit, software specifically designed to support auditors in the statistical components of an audit. The software has four features that are distinctive from other statistical software, and advantageous in the context of auditing practice. First, the interface of the software was developed with the auditor in mind. This means that the interface is user-friendly (e.g., the graphical user interface supports both Dutch and English), and that it aligns with international auditing standards. Second, the design of JASP for Audit ensures that sampling activities are documented and thus can be reproduced in a statistically sound manner. For example, the software automatically generates a report that provides insight into the statistical choices made in planning, selection and evaluating the sample and the interpretation of these choices. This is relevant to the auditor because conducting and reporting a statistical sample still proves to be relatively difficult (International Forum of Independent Audit Regulators (IFIAR), 2020, 2021, 2022). JASP for Audit also minimizes the chance of computational errors, as it is previously published in Derks et al. (2021b), which means that the technical and functional integrity of the software has been independently determined. Third, in addition to the standard frequentist statistics JASP for Audit also offers Bayesian statistics, which allows auditors to make optimal use of existing information from earlier stages of the audit or, for example, prior knowledge about the auditee or industry. In other words, JASP for Audit helps auditors understand, properly perform, document, and justify their—increasingly complex—audit samples. Finally, JASP for Audit is open-source software (Von Krogh and Spaeth, 2007). This means that the software can be downloaded for free and the source code is freely available to anyone.

This chapter is organized as follows. Section 6.2 discusses the aforementioned four distinguishing features of JASP for Audit. Subsequently, Section 6.3 demonstrates the software by means of three examples from the audit practice. Finally, Section 6.4 contains conclusions and recommendations for the use of JASP for Audit in practice.

6.2 Open-source software for auditors: JASP for Audit

Open-source software is software whose source code is publicly available without restriction and without cost. A characteristic of open source is that the source code can be adapted to create derivative software. This derivative software must be distributed under the same conditions as the original software (for the exact definition of open-source, see `https://opensource.org/osd`). However, in addition to the technical side—who is allowed to view, modify and distribute the code—there is another characteristic of open source software that is important to mention. Namely, unlike closed-source software, open-source software is developed collaboratively in a community with software developers and users. Thus, open-source development is sometimes compared to the crowded Bazaar and closed-source development to the Cathedral (Raymond, 1999). This has a number of advantages, for example that the end user always comes first and development of the software is normally faster than with closed-source (AlMarzouq et al., 2005; Morgan and Finnegan, 2007). An example of open source software aimed at the auditor is JASP.

JASP[1] (JASP Team, 2022; Love et al., 2019) is free and open-source statistical software with a point-and-click graphical interface which runs on the R statistical programming language (R Core Team, 2022). The software is a widely used tool for statistical analysis within several scientific fields (e.g., Wagenmakers et al., 2018a; Brydges and Gaeta, 2019; Faulkenberry et al., 2020; Kelter, 2020). For example, in 2022 JASP is downloaded around 33,000 times per month and is included in the curriculum of 226 universities from 59 different countries. The software is available in 10 different languages, including English and Dutch. JASP includes a large assortment of standard statistical analyses—such as *t*-tests (Gronau et al., 2019), linear regression (van den Bergh et al., 2021) and ANOVA (van den Bergh et al., 2020)—in both frequentist and Bayesian forms. In addition, JASP currently offers 19 add-on modules that add functionality to the software (e.g., Ly et al., 2018).

JASP for Audit (Derks et al., 2021b) is an add-on module for JASP that enriches the software with functionality for statistical sampling in the audit. The purpose of JASP for Audit is to provide auditors with guidance on the statistical aspects of designing, implementing and reporting a statistical audit sample. This is done by implementing the most commonly used techniques for audit sampling in a user-friendly interface that directly relates to familiar concepts from international auditing standards. In addition, JASP for Audit automatically and on-the-fly produces an audit report with the statistical results and the interpretation of these results. Finally, JASP gives auditors access to the latest academically developed statistical techniques in a user-friendly way, as academic development of auditor-focused statistical methods goes hand in hand with development of the software. In short, by taking over the statistical heavy lifting from auditors and guiding them through planning, selection, performing and evaluating an audit sample, JASP for Audit reduces audit complexity.

The sections below discuss four distinguishing features of JASP for Audit for the auditing practice: the interface is user-friendly and aligns with international auditing standards, the audit report is automatically generated, there is a choice between frequentist and Bayesian statistics, and the source code is open-source.

6.2.1 Feature 1: The user interface is designed for auditing and therefore user-friendly

JASP for Audit has a point-and-click interface, which means that results are available without delay once the user has clicked an option. The interface is also available in Dutch and relates to familiar concepts from auditing standards. For example, the Audit Risk Model is included in JASP for Audit, both record sampling and monetary unit sampling are supported, as are various selection methods including cell sampling. To increase accessibility for novice users, the most common values have been set as default for all options and advanced options have been hidden to keep the interface simple.

The statistical stages from the sampling process are implemented in a workflow and each in a separate analysis. The workflow automatically goes through the

[1] JASP can be freely downloaded for Windows, Mac, and Linux at `www.jasp-stats.org`.

standard stages for statistical audit sampling, selects the appropriate statistical technique, interprets the results, and a produces a readable audit report. This approach reduces the risk of statistical errors and increases auditors' understanding of the statistical outcomes. In addition, it is possible to complete each stage in the workflow individually.

Each stage of the sampling workflow is documented with a help file that provides additional explanation about statistical concepts.

6.2.2 Feature 2: An audit report is automatically generated

JASP for Audit automatically generates the statistical results in both text, tables and figures, including an audit report in the language of both the auditor and the statistician. The report describes all statistical characteristics of the sample so that it can be reproduced at all times. In addition, the report spells out the statistical results, which makes the interpretation of the outcomes easier. The structure of the report follows the three statistical stages in the sampling workflow: planning, selection and evaluation. This is convenient for auditors, as it allows the report to directly connect to the options entered and the work performed. Examples of elements of the report include a table with the sample size, a table with the selection results, and a table with the results of the statistical evaluation.

The report can be attached in its entirety to a general audit report. In addition, the data file, the selected options and the results can be saved as a *.jasp* file. For example, a statistician can set up the sample and save the file, which can then be read by an auditor to then perform and/or reproduce the sample.

6.2.3 Feature 3: Both frequentist and Bayesian statistics are facilitated

JASP for Audit provides both a frequentist and a Bayesian version of the sampling workflow and the individual stages in the workflow. Consequently, JASP for Audit allows auditors to take advantage of the benefits of Bayesian statistics, which previously were not easy to use within an audit. To our knowledge, there is no internationally and widely accessible, externally validated software available that implements Bayesian methods specifically for auditors.

Bayesian statistics is a way to increase efficiency and transparency in sampling by allowing available information to be optimally utilized (Derks et al., 2021a). For example, Bayesian statistics allows experts' knowledge to be incorporated into statistical analyses directly or through analytical procedures, which can reduce sample size and increase efficiency. In addition, within Bayesian statistics it is allowed to continuously monitor evidence over time (Wagenmakers et al., 2008). This is in contrast to the usual frequentist statistical analyses at a given sample size (Touw and Hoogduin, 2012).

6.2.4 Feature 4: The source code is open-source

JASP for Audit is open-source software, which means that the source code is freely available to anyone[2]. This has three advantages for the audit practice.

First, anyone can download the software for free without having to purchase a license for it. This eliminates the need for organizations to spend money on software licenses (e.g., IDEA (CaseWare Analytics, 2022), ACL (Dilligent, 2022)), resulting in a reduction in audit costs. Among other things, this makes JASP for Audit suitable as software for students, or to use when teaching courses on audit sampling to practitioners.

A second advantage of the open-source structure of JASP for Audit is that the source code is easy to adopt and modify. This means that users can propose or add features[3] to JASP for Audit themselves. In addition, accounting firms can implement their own preferences in the user interface, or include references to the guidelines and regulations of their professional offices in the software's documentation.

A third advantage is that it is easy to verify that the software is doing the correct calculations under the hood. The calculations in JASP for Audit are done in R by the package 'jfa' (Derks, 2022), whose source code is also open-source[4] and reviewed (Derks et al., 2021b). In this way, it is easy to check the source code of the software or to reproduce results independently in, for example, R (R Core Team, 2022), which guarantees full transparency to the users of the software. Of course, this feature does not take away from the fact that an accounting firm may have additional requirements for the use of open-source software.

6.3 Practical examples

In this section, we illustrate three practical examples from an audit context using JASP for Audit. In the first example, it is shown how the sample planning tables for testing internal controls contained in the AICPA's audit guide (American Institute of Certified Public Accountants (AICPA), 2019), which we consider relevant because it has been issued by the Association of International Certified Professional Accountants and is publicly available, and that of an international audit firm can be easily reproduced, justified, and expanded using JASP for Audit. While it is true that this example is primarily aimed at the external auditor, it is also relevant for internal auditors to use these sample sizes so that the external auditor can rely on the work of the internal auditor (de Swart et al., 2013). The second example shows how to reproduce an audit of internal declarations from Kloosterman (2018) using JASP for Audit. Finally, the third example shows how JASP for Audit can be used in the audit approach of the Dutch Tax Authorities using the planning, drawing and evaluation of a tax sample that was recently at the center of an appeal to the Court of Appeal of 's-Hertogenbosch (Buitenhuis,

[2] The source code for JASP for Audit can be found at `https://github.com/jasp-stats/jaspAudit`.

[3] Additional functionality for JASP for Audit can be nominated at `https://github.com/jasp-stats/jasp-issues/issues`.

[4] The source code for the R package jfa can be found at `https://github.com/koenderks/jfa`.

2022). These three examples are designed differently to show a diverse application of JASP for Audit. More examples including data and analyses are available in the internal data library of JASP (Wagenmakers and Kucharský, 2020). See Appendix 6.B for an introduction to the 'jfa' package and how to reproduce these examples in R.

6.3.1 Testing the operation of internal control measures

The first example demonstrates the functionality of JASP for Audit through a case familiar to the internal auditor: obtaining a minimum sample size for testing internal controls. American Institute of Certified Public Accountants (AICPA) (2019, Appendix A) provides guidance for this in the form of tables of minimum sample sizes, which depend on three parameters: the desired level of assurance (i.e., confidence) in the statistical statement after seeing the sample, the maximum allowable percentage of errors in the internal control measure, and the expected percentage of errors in the internal control measure. Other factors may also play a role, such as the frequency or nature of the control, but we limit ourselves here to the three parameters listed in American Institute of Certified Public Accountants (AICPA) (2019, Appendix A). A portion of American Institute of Certified Public Accountants (AICPA) (2019, Appendix A) is highlighted in Table 6.1. The following paragraphs describe how this table can be easily reproduced and nuanced using JASP for Audit.

Table 6.1: Sample sizes for testing internal controls (American Institute of Certified Public Accountants (AICPA), 2019, Appendix A).

	Control 1	Control 2	Control 3	Control 4
Confidence	90%	95%	95%	95%
Materiality	5%	4%	3%	3%
Expected error	2%	1.5%	0.5%	0.75%
Sample size	132	192	157	208

In order to reproduce this example, it is required to activate the audit module within JASP. This is done by clicking on the '+' icon in the upper right hand corner and then ticking the checkbox next to the 'Audit' module in the menu on the right hand side of the screen. The various analyses in JASP for Audit can then be viewed by clicking on the blue 'Audit' icon in the menu at the top of the screen. This example uses the planning stage of the audit workflow.

Figure 6.1 is a snapshot of the planning stage in JASP for Audit. The left side of the screen shows the graphical user interface, and the right side shows the audit report with the statistical results. In the user interface, the known parameters for the sample can be entered. To reproduce the sample size for Control 1 in Table 6.1, a 90 percent confidence level (this confidence level is partly related to the presumed audit evidence the auditor has already obtained from other audit procedures) and a performance materiality of 5 percent are specified. Furthermore, the expected errors in the sample are set to 2 percent.

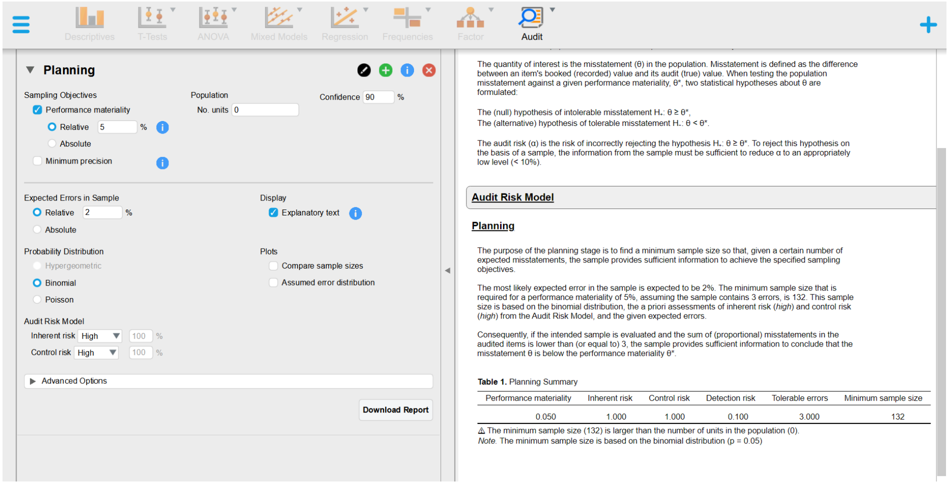


Figure 6.1: Snapshot of the planning stage in JASP for Audit for reproducing the example in the AICPA audit guide.

After specifying the parameters, the correct calculations are automatically performed under the hood, after which the audit report is completed with information about (the establishment of) the minimum sample size. The table in the audit report shows that for this specific combination of parameters a minimum sample size of 132 applies. Changing the parameters so that they reflect the requirements for Controls 2, 3 and 4 yields minimum sample sizes of 192, 157 and 208 respectively.

Another example of guidance for planning samples for testing internal controls can be found in the audit guide of an international audit firm. This guide prescribes that to be able to say with 90 percent certainty that no more than 8.8 percent of controls are not functioning properly, a sample size of 25 is needed if no errors are found in the sample. With JASP for Audit it is easy to calculate and check this (see Figure 6.2). Of course, the purpose of this last example is not to demonstrate that JASP for Audit can reproduce the sample planning prescribed by one specific accounting firm. It is merely an illustration that JASP for Audit speaks the same language used within an accounting firm.

An advantage of using JASP for Audit over existing sample size tables is that the auditor can easily calculate the corresponding sample size for any combination of parameters. They are therefore not bound by the limited number of possibilities in existing sample size tables and can be more flexible when planning an audit sample. For example, in addition to the audit guide, JASP for Audit also allows the auditor to calculate the upper bound on the population misstatement for any other sample size than the 25 mentioned here, or for other confidence levels than 90 percent, or in case errors do get found. In addition, JASP for Audit provides a Bayesian version of this analysis, in which prior knowledge of the controls can be responsibly included in the planning of the sample.

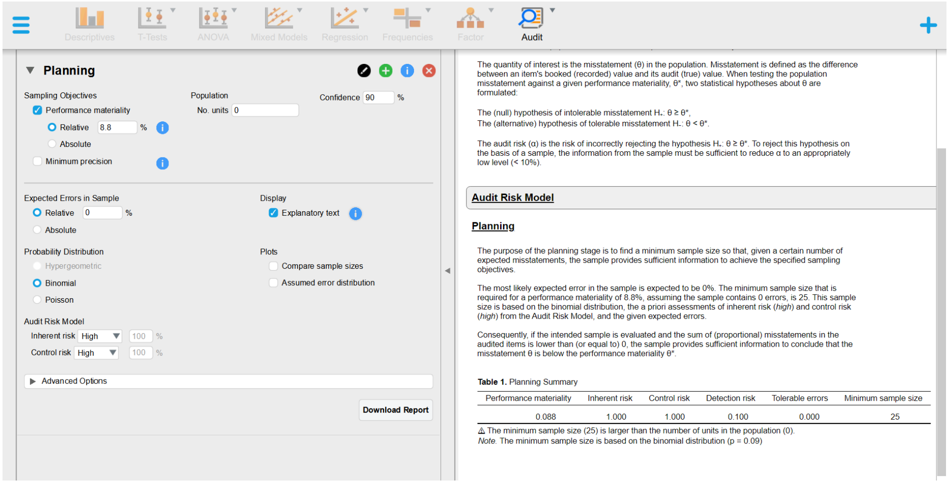


Figure 6.2: Snapshot of the planning stage for reproducing example in the audit guide of an international audit firm.

6.3.2 Testing the correctness of internal declarations

This second example demonstrates the functionality of JASP for Audit through a case in which an auditor conducts tests of details as part of an internal audit. The scenario is as follows: An internal auditor wants to obtain assurance on employees' claimed expenses for the entire year. These expenses are in a population of declarations consisting of 769 items with a total value of €2,333,333. For this example, a performance materiality of €105,000 (4.5 percent of the total value of the population) and an audit risk of 5 percent applies[5]. In other words, in order to state that the population does not contain a material misstatement, the auditor wants to be able to state with 95 percent certainty that the total amount of misstatement in the population is less than €105,000. In this example, there is no pre-existing information present from earlier stages of the audit: No risk analysis has been applied and thus the sample size cannot be reduced based on previous work. Furthermore, in this example the auditor works with an expected error of 0, meaning they do not allow for misstatements to occur in the sample.

The data for this example represent the 769 internal declarations of the organization, each of which has a corresponding booked value and a true value. The true value has been added for illustrative purposes, as it is normally determined by the auditor when the sample is audited. The following sections discuss step-by-step the planning, selection, and evaluation of the audit sample using JASP for Audit.

This example uses the sampling workflow within the Audit module. However, the stages of the sampling workflow described below (with the exception of the execution stage) are also separately available in JASP for Audit.

[5]This numerical example is partially taken from Kloosterman (2018).

6.3.2.1 Planning stage

In the planning stage, the auditor calculates a minimum sample size given the purpose of the sample. At this stage, the auditor also reviews relevant information collected during earlier stages of the audit. Note that information is relevant to the auditor if it can be used to adjust the amount of audit information needed to assess the misstatement in the population, for example, information about the quality of the organization's internal control systems. In a frequentist approach, this information is typically incorporated into the analysis using the Audit Risk Model (Voorhoeve, 2018). In the Bayesian approach, this information enters the analysis via the prior distribution (de Swart et al., 2013; van Batenburg, 2018c; Derks et al., 2021a).

The auditor begins the sampling workflow in the planning stage by calculating the minimum number of units from the population to be audited, given their expectation about the number of errors in the sample.

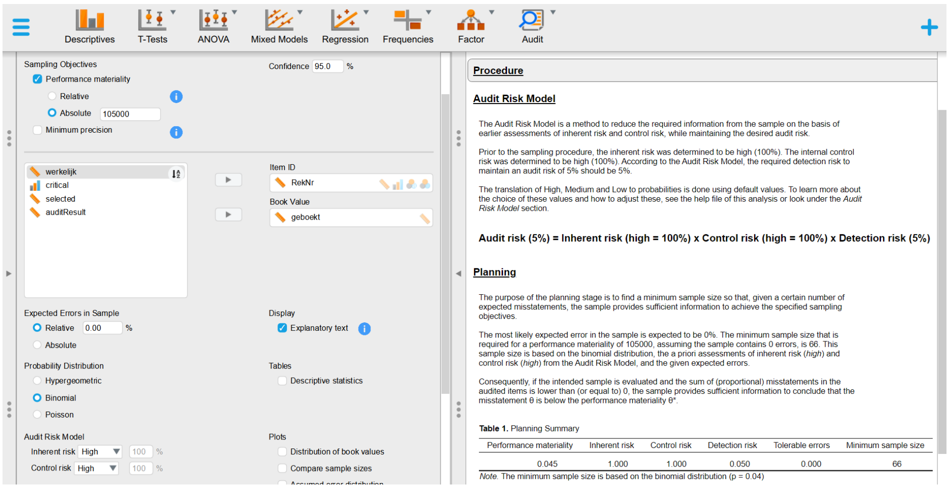


Figure 6.3: Snapshot of the planning stage in JASP for Audit for the audit of internal declarations.

Figure 6.3 is a snapshot of the planning stage from the sampling workflow in JASP for Audit for this example. On the left side of the screen, the graphical interface is again displayed. After entering the performance materiality in absolute (monetary) units in the top row, dragging the *RekNr* and *geboekt* variables into the 'ID' and 'Book Value' boxes, and specifying the expected errors in the sample, the minimum sample size is automatically calculated. On the right side of the screen, as the auditor fills in the relevant options, the audit report automatically appears. The standard table in the audit report shows that in this case the auditor must select 66 items—none of which may contain a misstatement—to approve the population with the required 95 percent certainty. In the interface, the auditor can choose to expand the audit report with figures and tables to clarify the statistical results. For example, a figure or table with descriptive statistics of the book values in the population can be added to the audit report, or the auditor can compare the minimum sample size under different probability distributions. With the calculated sample size in hand, the auditor continues the sampling workflow to the selection stage.

6.3.2.2 Selection stage

The auditor uses the minimum sample size calculated in the planning stage as input for the selection stage. In this second stage of the sampling workflow, a number of representative units from the population are selected in a statistically sound manner to be audited.

When selecting a statistical sample, an auditor generally needs to make three choices. First, it is important to specify which units are to be drawn from the population, the so-called sampling units. If the auditor decides to select at the item level it is called record sampling, at the euro level this is called monetary unit sampling (Leslie et al., 1979). Second, the auditor must choose the method by which the sampling units are drawn from the population. Drawing sampling units is typically done through one of three standard methods: fixed interval selection, cell selection, or random selection (van Batenburg, 2018d), see Figure 6.4. A third choice the auditor needs to make is whether there is a particular order to the items in the population, which lowers the representativeness of the sample. To ensure that there is no pattern present in the items, the auditor may choose to randomize the order of the items in the population prior to the selection procedure.

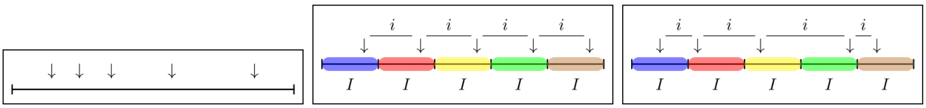


Figure 6.4: Illustration of the three common audit sampling algorithms implemented in JASP for Audit. The box on the left displays a random sampling algorithm in which each sampling unit is selected at random. The box in the middle displays a fixed interval algorithm with interval I in which the interval i between the selected units remains constant. The box on the right displays a cell sampling algorithm in which the interval I remains the same, but the interval i varies.

Figure 6.5 is a snapshot of the selection stage from the sampling workflow in JASP for Audit. An appropriate sample is automatically drawn without the auditor having to adjust options in the interface. Under the sampling units option, monetary units is checked in this example. This choice is made under the hood by JASP for Audit based on the options provided in the planning stage. In the interface, the auditor can specify the method for selecting the sampling units (i.e., euros). By default, this method is set to fixed interval selection where the first sampling unit from each interval is selected. In this example, the auditor chooses to randomize the order of the items for safety. The standard table in the audit report indicates that, with a fixed interval of €2,333,333 / 66 = €35,353.54, 66 euros distributed among 66 items were selected. The same table shows that with

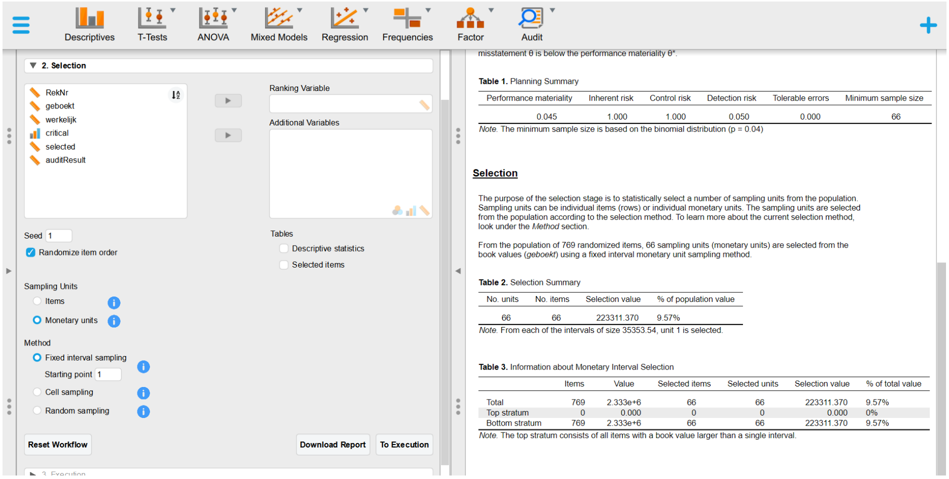


Figure 6.5: Snapshot of the selection stage in JASP for Audit for the audit of internal declarations.

this sample, 9.57 percent of the total value of the population is audited. Again, the auditor can use the options in the interface to expand the audit report. For example, the auditor can choose to include a table in the audit report with all selected items and any relevant characteristics of these items. With the selection of items to be audited in hand, the auditor continues the workflow to the execution stage.

6.3.2.3 Execution stage

After the auditor has selected the required items from the population, the audit is performed. In this stage of the audit, the auditor inspects the selected items and determines whether they contain a misstatement. Subsequently, the auditor enriches the sample data with the results of their audit. Here, it is important to make explicit in what way a misstatement is defined: Either an item is entirely misstated as soon as it contains a misstatement (regardless of the size of the misstatement), or an item is partially misstated if it contains a misstatement.

Figure 6.6 is a snapshot of the execution stage from the sampling workflow in JASP for Audit. In the interface, the auditor can choose to annotate the selected items in two ways: as correct or incorrect (full misstatements) or with an audited (true) value (partial misstatements). In the example, the auditor chooses to annotate the items with their true values. After selecting this option, the auditor can adjust the two fields with column names. The first column will contain the result of the selection procedure. In the second column, the auditor can enter the true values of the items in the sample. After adding these columns to the data set by means of the 'Insert Variables' button, a data editing window opens in the interface containing the items in the sample. In this window, the auditor can annotate the items by filling in their true values. In the example, the

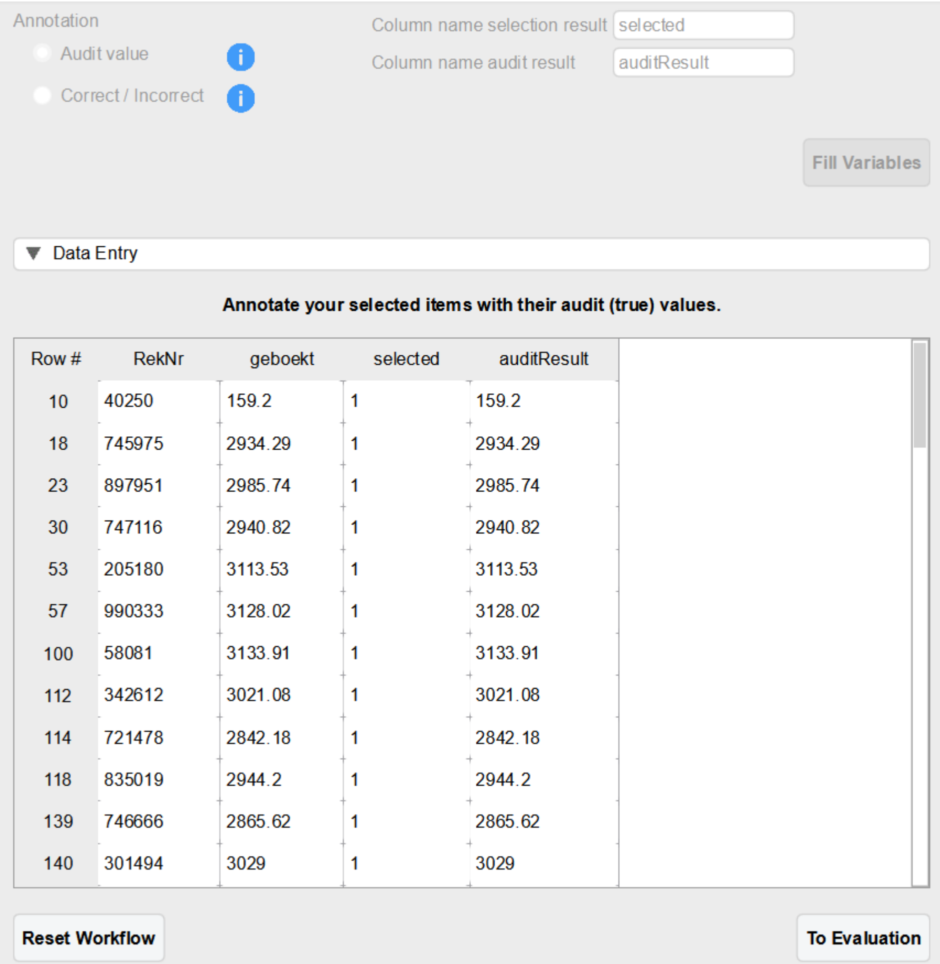


Figure 6.6: Snapshot of the execution stage in JASP for Audit for the audit of internal declarations.

auditor finds four misstatements in the sample. Table 6.2 is a summary of these four observed misstatements and their properties (Kloosterman, 2018, p. 38).

With the annotated items in hand, the auditor continues the workflow to the evaluation stage.

6.3.2.4 Evaluation stage

The auditor uses the annotated items that follow from the execution stage as input for the evaluation stage. In this stage of the sampling workflow, the auditor extrapolates the (possibly) found misstatements in the sample to the population while taking into account the audit risk. Here the auditor faces the choice of how to project the sample results to the population.

Table 6.2: Summary of the four misstatements observed in the sample. The table is taken from Kloosterman (2018). The true values of the items have been added for clarification.

RekNr	Geboekt	Misstatement	True value	Conclusion
47500	€1,364.52	€87.50	€1,277.02	€87.50 private expense, no cost
41400	€32,971.81	€9,354.10	€23,617.71	Asset Maintenance, $\frac{1}{3}$ to be capitalized
40250	€159.20	€19.11	€140.09	€19.11 private expense, no cost
15230	€3,250.00	€3,250.00	€0.00	Incorrectly not activated

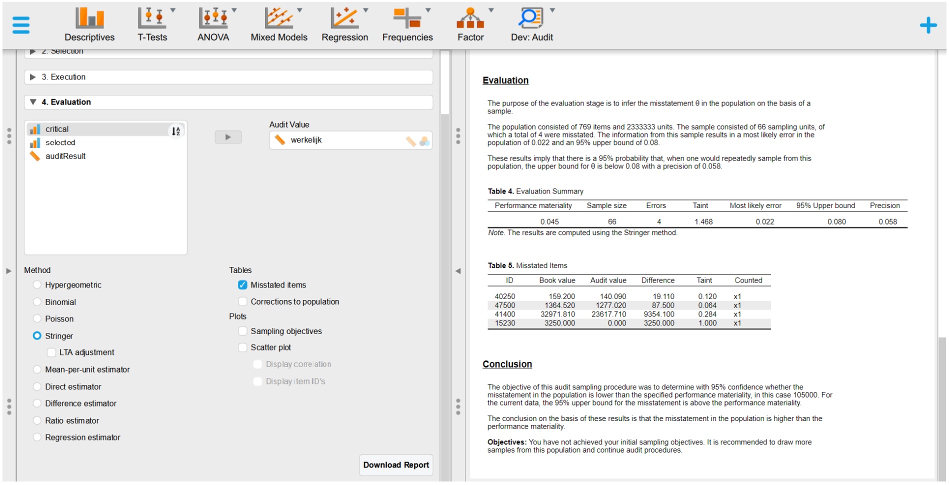


Figure 6.7: Snapshot of the evaluation stage in JASP for Audit for the audit of internal declarations.

Figure 6.7 is a snapshot of the evaluation stage from the sampling workflow in JASP for Audit. In the interface, an appropriate evaluation method for monetary statements is automatically selected, in this case the commonly used Stringer bound (Bickel, 1992). The auditor only needs to drag the variable containing the true values to the 'Audit Value' box, after which the statistical evaluation of the sample takes place automatically. The standard table in the audit report shows that the most likely misstatement amount in the population is equal to €51,894.14. The upper limit of the one-sided 95 percent confidence interval for this misstatement amount is €186,772.40[6]. The audit report contains the statistical interpretation of these results and the conclusion in relation to the purpose of the sample. The auditor can select options in the interface to expand the audit report, for example, a figure comparing the performance materiality, the most likely error, and the 95 percent upper bound on the population misstatement. In this case, the auditor has obtained insufficient and inappropriate audit evidence to conclude that the misstatement in the population is below €105,000 with 95 percent certainty and must therefore conclude that the declarations contain a

[6]See Appendix 6.A for the calculations underlying these results.

material misstatement.

To conclude the sampling workflow in JASP for Audit, the audit report can be exported to different formats (e.g., *.pdf*, *.html*) and the analysis can be saved locally as a *.jasp* file. The auditor can then share this file with the statistician for verification of the data and options, who can then send it to management to communicate the sample results in an understandable way via the audit report.

6.3.3 Testing the correctness of taxes

As a final example, we will reproduce the results of a sample that was recently at the center of an appeal case at the Court of Appeal of 's-Hertogenbosch (Buitenhuis, 2022). An auditor of the Tax Administration initiated an investigation into corporate income tax, sales tax and payroll taxes, using, among other things, a monetary unit sample in accordance with the Tax Administration's audit approach (Belastingdienst, 2021). The stakeholder had claimed an amount of €2,812,567 in total expenses. With a materiality of €120,000 and an expectation of zero errors, this leads to a sample size of 71[7] in the planning stage of JASP for Audit. In the selection stage of JASP for Audit, the required sample can be selected by loading the population file with the expenditures, indicating the amount of the expenditures as the book value, and checking the cell sampling method preferred by the Tax Administration (Belastingdienst, 2021). During the execution stage, three misstatements were identified with an accumulated misstatement fraction of 1.89. After the true values are added to the sample list during the execution stage, JASP for Audit generates the report of the entire workflow in the evaluation stage. See Figure 6.8 for the evaluation section from this report. It can be seen that the upper limit at €229,536 is greater than the materiality of €120,000, that the expenses cannot be approved and that the most likely misstatement (i.e., the consequential loss of taxes expressed in base euros) is €73,830.

6.4 Conclusions and recommendations

This chapter introduced JASP for Audit, open-source software developed to support auditors in responsibly planning, selecting and evaluating their statistical samples. It has been argued that this software can provide a bridge between auditors and statisticians by providing four features that are beneficial to the audit practice. In addition, JASP for Audit provides a unique opportunity for collaboration for auditors, accounting firms, universities and business schools in the field of statistical auditing. Because the source code is open-source, the latest scientific developments and recommendations from practitioners can be quickly incorporated into the software. This makes state-of-the-art sampling techniques available

[7]Note that there are small differences between values in Buitenhuis (2022) and the values generated with JASP for Audit. The Tax Administration assumes a sample size of 72 and a rounded sample interval of €40,000. JASP for Audit assumes an exact sample interval of €39,063. This causes a small difference in the calculated correction. For the evaluation, both the Tax Administration and JASP for Audit use the Stringer bound method, which depends on the individual misstatement fractions. These are not given in Buitenhuis (2022) and are simulated as 0.5, 0.5, and 0.89 for convenience. Because of this, the resulting upper bounds are also slightly different.

The population consisted of 75 items and 2812567 units. The sample consisted of 72 sampling units, of which a total of 3 were misstated. The information from this sample results in a most likely error in the population of 73829.884 and an 95% upper bound of 229535.117.

These results imply that there is a 95% probability that, when one would repeatedly sample from this population, the upper bound for θ is below 229535.117 with a precision of 155705.234.

Table 1. Evaluation Summary

Performance materiality	Sample size	Errors	Taint	Most likely error	95% Upper bound	Precision
120000.000	72	3	1.890	73829.884	229535.117	155705.234

Note. The results are computed using the Stringer method.

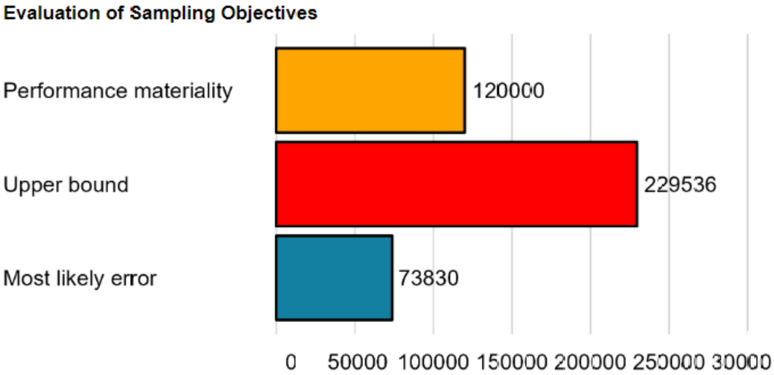


Figure 1. Evaluation information for the current annotated selection. The materiality is compared with the maximum misstatement and the most likely error. The most likely error (MLE) is an estimate of the true misstatement in the population. The upper bound is an estimate of the maximum error in the population.

Conclusion

The objective of this audit sampling procedure was to determine with 95% confidence whether the misstatement in the population is lower than the specified performance materiality, in this case 120000. For the current data, the 95% upper bound for the misstatement is above the performance materiality.

The conclusion on the basis of these results is that the misstatement in the population is higher than the performance materiality.

Figure 6.8: Evaluation report in JASP for Audit for the tax audit example.

to every auditor, which will improve the quality of audits overall. JASP for Audit streamlines the sampling process for auditors so they can focus on the core tasks in an audit.

Since new software is not easily adopted, we mention here some recommendations for the use of JASP for Audit in the auditing practice. First, reference can be made to the validation of statistical outcomes. For example, this thesis contains several chapters and appendices using JASP for Audit that can be used to substantiate adoption of the software. In addition, the statistical results from the software are regularly validated against other audit software and the results of these procedures are also open-source[8].

[8]These so-called unit tests can be found at `https://github.com/koenderks/jfa/actions/`

In short, JASP for Audit can be a valuable addition to the audit software curriculum. The hope is that JASP for Audit will help increase the quality of audits and reduce the time and money spent per audit.

workflows/r.yml and https://github.com/jasp-stats/jaspAudit/actions/workflows/unittests.yml.

6.A Calculations Stringer bound

In the example, a sample of $n = 66$ euros was drawn from a population of $N = 2,333,333.00$ euros. The auditor found $k = 4$ misstatements (see Table 6.2). The audit risk is set at $\alpha = 0.05$. In the example, the auditor is interested in the most likely misstatement amount in the population and the 95 percent upper limit for this misstatement amount.

Using the booked value y_i and the actual value x_i, the auditor first calculates the proportional misstatement (taint) per euro, t_i, in the sample, see Equation 6.A.1.

$$t_i = \frac{y_i - x_i}{y_i} \tag{6.A.1}$$

These taints are respectively $t_1 = 0.06412511$, $t_2 = 0.28369992$, $t_3 = 0.12003769$, and $t_4 = 1$ for the misstatements in Table 6.2. The probability of misstatement per euro in the sample can then be calculated using Equation 6.A.2.

$$\theta_{\text{mle}} = \frac{\sum_{i=0}^{n} t_i}{n} = \frac{1.467863}{66} = 0.02224034 \tag{6.A.2}$$

The most likely amount of misstatement in the population is calculated by multiplying the probability of misstatement per euro θ_{mle} by the total value of the population N, see Equation 6.A.3.

$$\theta_{\text{mle}} \times N = 0.02224034 \times 2.333.333,00 = 51,894.14 \tag{6.A.3}$$

The upper bound for the misstatement amount is determined using the Stringer bound (Touw and Hoogduin, 2012, pp. 190–194)[9]. The Stringer bound is a statistical upper bound for the probability of misstatement per euro and is calculated using Equation 6.A.4 (Bickel, 1992).

$$p(0; 1 - \alpha) + \sum_{i=0}^{k} [p(i; 1 - \alpha) - p(i - 1; 1 - \alpha)] \times t_i \tag{6.A.4}$$

In Equation 6.A.4, $p(i; 1-\alpha)$ is the $1-\alpha$ upper bound for at most i misstatements in n items according to the binomial distribution (Clopper and Pearson, 1934). It is equivalent to the $1 - \alpha$ percentile of a Beta$(1 + i,\, n - i)$ distribution (Pearson, 1934). For an error-free sample, the upper bound for θ therefore simply equals $1 - \alpha^{\frac{1}{n}}$. When calculating the Stringer bound in the case of found misstatements,

[9] Note that the Stringer bound can be computed using both the Poisson distribution and the binomial distribution. Touw and Hoogduin (2012) assume Poisson. Because especially for larger error fractions the binomial distribution gives sharper upper bounds, JASP for Audit assumes the binomial distribution as described in Bickel (1992).

the taints are placed in descending order in the formula. The complete formula for the Stringer bound in the example is given in Equation 6.A.5.

$$\begin{aligned} p(0;0.95) &+ [p(1;0.95) - p(0;0.95)] \times t_4 \\ &+ [p(2;0.95) - p(1;0.95)] \times t_2 \\ &+ [p(3;0.95) - p(2;0.95)] \times t_3 \\ &+ [p(4;0.95) - p(3;0.95)] \times t_1 \end{aligned} \tag{6.A.5}$$

Filling in this formula (without any rounding) gives a 95 percent upper bound θ_{95} for the probability of misstatement per euro θ of 0.08004531, rounded 8 percent of the total value of the population. The 95 percent upper bound on the amount of misstatement in the population is calculated by multiplying the 95 percent upper bound on the probability of misstatement per euro by the total value of the population N, see Equation 6.A.6.

$$\theta_{.95} \times N = 0.08004531 \times 2.333.333,00 = 186,772.40 \tag{6.A.6}$$

6.B The ‘jfa’ R package

The ‘jfa’ package is open-source software for the programming language R (R Core Team, 2022) that provides the statistical audit sampling methods implemented in JASP for Audit (Derks et al., 2021b). The package provides an intuitive workflow for planning, performing, evaluating, and reporting a statistical audit sample compliant with international auditing standards (International Auditing and Assurance Standards Board (IAASB), 2018; Public Company Accounting Oversight Boards (PCAOB), 2020; American Institute of Certified Public Accountants (AICPA), 2021). Next to the frequentist audit sampling methodology, the package implements Bayesian equivalents of these methods whose statistical underpinnings are described throughout this thesis. The ‘jfa’ package is the first R package that implements every step in the statistical audit sampling workflow[10]. This appendix illustrates the functionality of the ‘jfa’ package using the three practical examples described in this chapter.

6.B.1 Installation

Before the reader can start to reproduce the examples in this chapter, it is required to first install the ‘jfa’ package from the CRAN repository[11] and subsequently load the package in the current R session. The following R code downloads the package directly from CRAN via a call to `install.packages` and loads the package via a call to `library`.

[10]The R package ‘audit’ implements evaluation of audit samples and the r package ‘MUS’ implements planning and evaluation of audit samples.

[11]See `https://cran.r-project.org/package=jfa` for the CRAN repository of the ‘jfa’ package.

```
install.packages("jfa") # Download the package from CRAN
library(jfa)            # Load the package in the R session
```

6.B.2 Intended workflow

Auditors generally have a four-stage approach to audit sampling. This approach consists of (1) planning a minimum required sample size, (2) selecting the planned number of sampling units from the population, (3) executing the tests of details, and (4) evaluating the misstatement in the sample to perform inference about the population misstatement. The 'jfa' package provides an intuitive workflow that fits seamlessly into this four-stage approach to audit sampling. Concretely, the package provides four main functions to the auditor; `planning`, `selection`, `evaluation`, and `report`. Additionally, it provides the function `auditPrior` to create a prior distribution for use in Bayesian audit sampling. Since providing an efficient workflow is one of the goals of the package, the output obtained from functions in the workflow can be used directly as input arguments for subsequent functions in the workflow. Figure 6.9 shows the intended workflow of the package, and highlights how it facilitates auditors' four-stage approach to audit sampling.

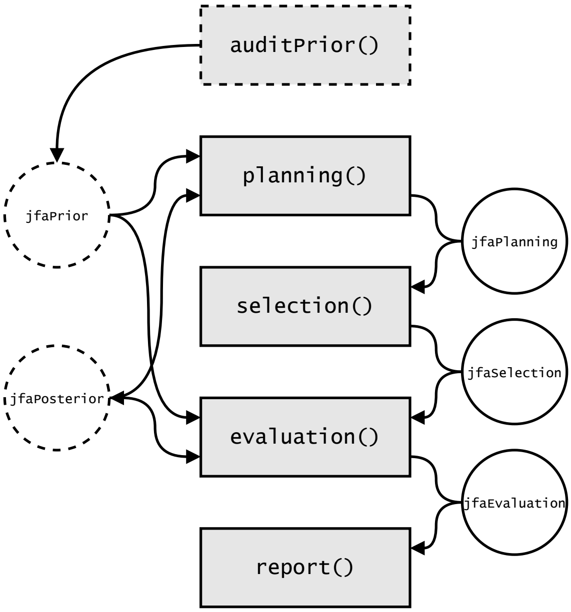


Figure 6.9: The figure displays the intended workflow in the 'jfa' package. Rectangles represent functions and circles indicate objects returned from these functions. Functions and objects with dotted lines are available when performing a Bayesian analysis.

6.B.3 Example 1: Sample sizes for internal controls testing

Calculating a minimum sample size for internal controls testing can be achieved via the `planning` function. In general, the auditor needs to supply to the `planning` function four parameters that are readily available: the performance materiality (argument `materiality`), the expected errors in the sample (argument `expected`), the likelihood of the data (argument `likelihood`), and the required confidence level (argument `conf.level`). Optionally, a prior distribution created using the `auditPrior` function can be specified (argument `prior`) to perform Bayesian planning using the specified prior distribution. Given the inputs, the `planning` function returns an object of class `jfaPlanning` containing the input options and the statistical results, including the minimum required sample size.

6.B.3.1 Frequentist approach

We first reproduce the first example in the chapter using the frequentist approach implemented in ‘jfa’. For the first control in Table 6.1, the auditor wants to calculate a sample size such that, when at most 2 percent errors are found in the sample, they can be 90 percent confident that the misstatement in the population is lower than the performance materiality of 5 percent. The corresponding sample size is 132, see American Institute of Certified Public Accountants (AICPA) (2019, Appendix A). The R code below computes this sample size.

```
# Control 1
planning(materiality = 0.05, expected = 0.02, conf.level = 0.90,
         likelihood = "binomial")

#        Classical Audit Sample Planning
#
# minimum sample size = 132
# sample size obtained in 133 iteration(s) via method "binomial"
```

Second, for the second control in Table 6.1, the auditor wants to calculate a sample size such that, when at most 1.5 percent errors are found in the sample, they can be 95 percent confident that the misstatement in the population is lower than the performance materiality of 4 percent. The corresponding sample size is 192, see American Institute of Certified Public Accountants (AICPA) (2019, Appendix A). The R code below computes this sample size.

```
# Control 2
planning(materiality = 0.04, expected = 0.015, conf.level = 0.95,
         likelihood = "binomial")

#        Classical Audit Sample Planning
#
# minimum sample size = 192
# sample size obtained in 193 iteration(s) via method "binomial"
```

Third, for the third control in Table 6.1, the auditor wants to calculate a sample size such that, when at most 0.5 percent errors are found in the sample, they can be 95 percent confident that the misstatement in the population is lower than the performance materiality of 3 percent. The corresponding sample size is 157, see American Institute of Certified Public Accountants (AICPA) (2019, Appendix A). The R code below computes this sample size.

```
# Control 3
planning(materiality = 0.03, expected = 0.005, conf.level = 0.95,
         likelihood = "binomial")

#        Classical Audit Sample Planning
#
# minimum sample size = 157
# sample size obtained in 158 iteration(s) via method "binomial"
```

Finally, for the last control in Table 6.1, the auditor wants to calculate a sample size such that, when at most 0.75 percent errors are found in the sample, they can be 95 percent confident that the misstatement in the population is lower than the performance materiality of 3 percent. The corresponding sample size is 208, see American Institute of Certified Public Accountants (AICPA) (2019, Appendix A). The R code below computes this sample size.

```
# Control 4
planning(materiality = 0.03, expected = 0.0075, conf.level = 0.95,
         likelihood = "binomial")

#        Classical Audit Sample Planning
#
# minimum sample size = 208
# sample size obtained in 209 iteration(s) via method "binomial"
```

6.B.3.2 Bayesian approach

In this subsection, a Bayesian approach is used to carry out the examples in the chapter. In general, the input for the `planning` function does not differ between the frequentist and Bayesian approaches, except that in the Bayesian approach the auditor specifies a prior distribution via the `prior` argument.

The purpose of the `auditPrior` function is to create a prior distribution that can be used as an input for the `planning` and `evaluation` functions. The default option for these functions is `prior = FALSE`, which means that they will perform a frequentist analysis. However, if the input for `prior` is an object created by the `auditPrior` function, Bayesian analyses are performed using the specified prior distribution. On the other hand, if `prior = TRUE` a uniform prior distribution is specified under the hood. To illustrate the versatility of the package when it comes to Bayesian inference for auditing, an informed prior distribution will be used throughout the following reconstructions. Keep in mind that this means that the sample sizes will be different from those obtained using the frequentist approach. Suppose that the auditor has information that leads to a Beta(1, 19) prior distribution. The following R code specifies this prior distribution.

```
# Create a custom Beta(1, 19) prior distribution
prior <- auditPrior(method = "param", likelihood = "binomial",
                    alpha = 1, beta = 19)
```

First, the required sample size for the first control in Table 6.1 is calculated using a Beta(1, 19) prior distribution. In contrast to the frequentist function call to `planning`, this time the `prior` argument in the function is also specified. The corresponding sample size is 68. The R code below computes this sample size.

```
# Control 1
planning(materiality = 0.05, expected = 0.02, conf.level = 0.90,
         likelihood = "binomial", prior = prior)

#         Bayesian Audit Sample Planning
#
# minimum sample size = 68
# sample size obtained in 69 iteration(s) via method "binomial" + "prior"
```

Second, the required sample size for the second control in Table 6.1 is calculated using a Beta(1, 19) prior distribution. The corresponding sample size is 140. The R code below computes this sample size.

```
# Control 2
planning(materiality = 0.04, expected = 0.015, conf.level = 0.95,
         likelihood = "binomial", prior = prior)

#         Bayesian Audit Sample Planning
#
# minimum sample size = 140
# sample size obtained in 141 iteration(s) via method "binomial" + "prior"
```

Third, the required sample size for the third control in Table 6.1 is calculated using a Beta(1, 19) prior distribution. The corresponding sample size is 114. The R code below computes this sample size.

```
# Control 3
planning(materiality = 0.03, expected = 0.005, conf.level = 0.95,
         likelihood = "binomial", prior = prior)

#         Bayesian Audit Sample Planning
#
# minimum sample size = 114
# sample size obtained in 115 iteration(s) via method "binomial" + "prior"
```

Finally, the required sample size for the last control in Table 6.1 is calculated using a Beta(1, 19) prior distribution. The corresponding sample size is 140. The R code below computes this sample size.

```
# Control 3
planning(materiality = 0.03, expected = 0.0075, conf.level = 0.95,
         likelihood = "binomial", prior = prior)
```

```
#        Bayesian Audit Sample Planning
#
# minimum sample size = 140
# sample size obtained in 141 iteration(s) via method "binomial" + "prior"
```

6.B.4 Example 2: Auditing internal declarations

This second example utilizes the full audit sampling workflow in 'jfa'. First, the population data (available in the online appendix at `https://osf.io/2v7mw/`) is loaded into the R session. Next, the total number of euros in the population is stored as a separate variable, since it is required later in the analysis. The R code below performs these two actions.

```
# Read in population and total value of the population
population <- read.csv("https://osf.io/r47ap/download")
N.units <- sum(population$geboekt)

N.units # Total number of euros in the population = 2333333
```

6.B.4.1 Frequentist approach

First, using the `planning` function the minimum sample size for a performance materiality of 105000 / 2333333 = 0.045 percent is calculated for zero tolerable errors. This results in a sample size of 66 euros to inspect. The R code below computes this sample size.

```
planning <- planning(materiality = 105000 / N.units,
                     likelihood = "binomial")
summary(planning)

#        Classical Audit Sample Planning Summary
#
# Options:
#   Confidence level:            0.95
#   Materiality:                 0.045
#   Hypotheses:                  H0: θ >= 0.045 vs. H1: θ < 0.045
#   Expected:                    0
#   Likelihood:                  binomial
#
# Results:
#   Minimum sample size:         66
#   Tolerable errors:            0
#   Expected most likely error:  0
#   Expected upper bound:        0.044375
#   Expected precision:          0.044375
#   Expected p-value:            < 2.22e-16
```

Second, the 66 required euros are selected from the population using the `selection` function. The R code below selects these 66 units from the population.

```
# IMPORTANT: JASP uses the default rng in R versions prior to 3.6.0
suppressWarnings(RNGkind(sample.kind = "Rounding"))
set.seed(1)
selection <- selection(data = population, size = planning,
                       units = "values", values = "geboekt",
                       randomize = TRUE)
summary(selection)

#         Audit Sample Selection Summary
#
# Options:
#   Requested sample size:      66
#   Sampling units:             monetary units
#   Method:                     fixed interval sampling
#   Starting point:             1
#
# Data:
#   Population size:            769
#   Population value:           2333333
#   Selection interval:         35354
#
# Results:
#   Selected sampling units:    66
#   Proportion of value:        0.095705
#   Selected items:             66
#   Proportion of size:         0.085826

sample <- selection$sample # Save the sample
```

Last, the annotated sample is evaluated using the Stringer bound (Bickel, 1992) via the `evaluation` function. The R code below computes the statistical results.

```
evaluation <- evaluation(materiality = 105000 / N.units,
                         method = "stringer",
                         data = sample, values = "geboekt",
                         values.audit = "werkelijk")
summary(evaluation)

#         Classical Audit Sample Evaluation Summary
#
# Options:
#   Confidence level:            0.95
#   Materiality:                 0.045
#   Method:                      stringer
#
# Data:
#   Sample size:                 66
#   Number of errors:            4
#   Sum of taints:               1.4678627
```

```
#
# Results:
#   Most likely error:           0.02224
#   95 percent confidence interval: [0, 0.080045]
#   Precision:                   0.057805

evaluation$mle * N.units # Most likely error in euros = 51894.14
evaluation$ub * N.units  # 95% Upper bound in euros = 186772.4
```

6.B.4.2 Bayesian approach

In a Bayesian approach, the auditor must first specify a prior distribution using the `auditPrior` function. For illustrative purposes, an improper Beta(1, 0) prior distribution will be specified in this example because it yields an equivalent sample size (and thus the same sample and found misstatements) as in the frequentist approach, see Chapter 5. The R code below sets up the improper Beta(1, 0) distribution.

```
# Create an improper Beta(1, 0) prior distribution
prior <- auditPrior(method = "strict", likelihood = "binomial")
```

Second, the minimum sample size is calculated by specifying the `prior` argument in the `planning` function. This results in the familiar 66 euros (like in the frequentist example) to inspect. The R code below computes this sample size.

```
planning <- planning(materiality = 105000 / N.units,
                     likelihood = "binomial", prior = prior)
summary(planning)

#        Bayesian Audit Sample Planning Summary
#
# Options:
#   Confidence level:           0.95
#   Materiality:                0.045
#   Hypotheses:                 H0: θ >= 0.045 vs. H1: θ < 0.045
#   Expected:                   0
#   Likelihood:                 binomial
#   Prior distribution:         beta(α = 1, β = 0)
#
# Results:
#   Minimum sample size:        66
#   Tolerable errors:           0
#   Posterior distribution:     beta(α = 1, β = 66)
#   Expected most likely error: 0
#   Expected upper bound:       0.044375
#   Expected precision:         0.044375
#   Expected BF10:              Inf
```

Third, the 66 required units are selected from the population using the `selection` function. This R code is the same as in the frequentist example and therefore it is omitted here. As a result, the resulting sample is also the same. Finally, the sample is evaluated using the beta distribution by specifying the `prior` argument in the `evaluation` function. The R code below computes the statistical results.

```
evaluation <- evaluation(materiality = 105000 / N.units,
                         method = "binomial", prior = prior,
                         data = sample, values = "geboekt",
                         values.audit = "werkelijk")
summary(evaluation)

#        Bayesian Audit Sample Evaluation Summary
#
# Options:
#   Confidence level:             0.95
#   Materiality:                  0.045
#   Hypotheses:                   H0: θ >= 0.045 vs. H1: θ < 0.045
#   Method:                       binomial
#   Prior distribution:           beta(α = 1, β = 0)
#
# Data:
#   Sample size:                  66
#   Number of errors:             4
#   Sum of taints:                1.4678627
#
# Results:
#   Posterior distribution:       beta(α = 2.468, β = 64.532)
#   Most likely error:            0.022582
#   95 percent confidence interval: [0, 0.080622]
#   Precision:                    0.058039
#   BF10:                         Inf

evaluation$mle * N.units # Most likely error in euros = 52692.51
evaluation$ub * N.units  # 95% Upper bound in euros = 188117.9
```

6.B.5 Example 3: Tax audit

As a final example, we reproduce the tax audit example from the chapter. First, the population data (available in the online appendix at `https://osf.io/2v7mw/`) is loaded into the R session. Next, the total number of euros in the population is calculated and stored in a separate variable, since it is required at a later stage. The R code below performs these two actions.

```
# Read in population and total booked value
population <- read.csv("https://osf.io/7sdn4/download",
                       sep = ";", dec = ",")
N.units <- sum(population$book)

N.units # Total number of euros in the population = 2812567
```

6.B.5.1 Frequentist approach

In the frequentist approach, the sample is evaluated using the Stringer bound (Bickel, 1992). The R code below computes the statistical results.

```
evaluation <- evaluation(materiality = 120000 / N.units,
                         method = "stringer",
                         data = population, values = "book",
                         values.audit = "audit")
summary(evaluation)

#         Classical Audit Sample Evaluation Summary
#
# Options:
#   Confidence level:              0.95
#   Materiality:                   0.042666
#   Method:                        stringer
#
# Data:
#   Sample size:                   72
#   Number of errors:              3
#   Sum of taints:                 1.89
#
# Results:
#   Most likely error:             0.02625
#   95 percent confidence interval: [0, 0.081611]
#   Precision:                     0.05536

evaluation$mle * N.units # Most likely error in euros = 73829.88
evaluation$ub * N.units  # 95% Upper bound in euros = 229535.1
```

6.B.5.2 Bayesian approach

In the Bayesian approach, the sample is evaluated using the beta distribution. For this example, an impartial prior distribution is specified, see Chapter 5. The R code below sets up the impartial beta prior distribution for a performance materiality of 0.042666.

```
# Create an impartial beta prior distribution
prior <- auditPrior(method = "impartial", likelihood = "binomial",
                   materiality = 120000 / N.units)
```

Next, the sample is evaluated using the impartial prior distribution by specifying the `prior` argument in the `evaluation` function. The R code below computes the statistical results.

```
evaluation <- evaluation(materiality = 120000 / N.units,
                         method = "binomial", prior = prior,
                         data = population, values = "book",
                         values.audit = "audit")
summary(evaluation)
```

```
#        Bayesian Audit Sample Evaluation Summary
#
# Options:
#   Confidence level:           0.95
#   Materiality:                0.042666
#   Hypotheses:                 H0: θ >= 0.042666 vs. H1: θ < 0.042666
#   Method:                     binomial
#   Prior distribution:         beta(α = 1, β = 15.897)
#
# Data:
#   Sample size:                72
#   Number of errors:           3
#   Sum of taints:              1.89
#
# Results:
#   Posterior distribution:     beta(α = 2.89, β = 86.007)
#   Most likely error:          0.02175
#   95 percent confidence interval: [0, 0.068085]
#   Precision:                  0.046335
#   BF10:                       3.0084

evaluation$mle * N.units # Most likely error in euros = 61173.06
evaluation$ub * N.units  # 95% Upper bound in euros = 191493.3
```

Part IV

Conclusion

Chapter 7

Discussion and Future Directions

In this thesis, I have argued that auditors should adopt Bayesian inference into their statistical toolbox. When compared to the currently dominant frequentist methodology, Bayesian methods can provide auditors with many practical benefits. In particular, the preceding chapters have outlined the arguments in favor of the Bayesian approach in a modern auditing context, they have discussed the development of new Bayesian statistical methods, and they have described how the accessibility of Bayesian methods for auditing has been improved via open-source software. In sum, this thesis aimed to develop an innovative approach to statistical auditing that elevates the field to the Bayesian methodological standards currently upheld in other scientific disciplines (van de Schoot et al., 2021).

The first part of this thesis focused on Bayesian parameter estimation, demonstrating how pre-existing information can be incorporated into the prior distribution and the statistical model. The central message of this part is that integrating pre-existing information into the statistical analysis can help auditors to set up an efficient procedure and provide a transparent opinion because it is specifically tailored to the audit and the auditee. The second part of this thesis covered Bayesian hypothesis testing and advocated the use of the Bayes factor as a gauge for audit evidence. The main takeaway from this part is that Bayesian hypothesis testing using the Bayes factor offers practical advantages over the standard frequentist approach, and that the interpretation of the Bayes factor closely resembles the conclusions that auditors want to draw from their samples. The final part discussed JASP for Audit, an open-source software implementation of the discussed audit sampling methodology, and demonstrated how it can aid the auditor in planning, performing, evaluating, and reporting a (Bayesian) audit sample.

In the remaining parts of this discussion, I will go over the main contributions of this thesis to the field of auditing and suggest possible directions for further research or development.

7.1 Contributions to auditing

This thesis makes three main contributions to audit theory and practice:

- It explains and demonstrates the practical benefits of Bayesian inference in a modern auditing context;
- It develops innovative Bayesian techniques that can be tailored to real-world audit situations;
- It improves the accessibility of Bayesian methodology to auditors via open-source software.

These three contributions may be useful to a wide range of stakeholders of the audit, including those responsible for developing auditing standards and methodologies; practitioners interested in conducting effective and efficient audits; and academics involved in auditing research.

7.1.1 Benefits of Bayesian inference

The first goal of this thesis was to promote the use of Bayesian statistics in the auditing profession. It has long been argued that auditors can benefit practically from using Bayesian statistics to analyze their samples. However, these Bayesian benefits are seldom known to auditors, or are difficult to make use of in practice. Furthermore, over the years, the body of fundamental literature outlining the Bayesian benefits for auditing has unfortunately become outdated and therefore difficult to relate to for auditors in practice. For this reason, the chapters in this thesis modernize, and expand upon, the arguments in favor of Bayesian inference made by scholars in the mid-to-late twentieth century.

Chapter 2, Chapter 3, and Chapter 4 have outlined the advantages of Bayesian parameter estimation and hypothesis testing in a modern auditing context. Four main advantages have been put forward. To sum up, Bayesian inference provides auditors with a relatively simple interpretation of statistical results; it enables statistical conclusions to be easily extended to any level of complexity; it increases transparency towards stakeholders of the audit; and, finally, it can help auditors work more efficiently. Particularly in today's audits that are subject to close scrutiny and time pressure, Bayesian inference can make it easier for auditors to accomplish their goals.

Of course, there is a trade-off that auditors need to make. While a Bayesian analysis can improve transparency and efficiency, the justification of a Bayesian analysis takes time and effort on the part of the auditor. Information that is incorporated into the prior distribution or the statistical model should be properly justified, which means that the auditor must carefully consider the pros and cons of the Bayesian approach to determine if its benefits will outweigh the time and effort required to set up and justify a Bayesian analysis. Chapter 2 and Chapter 3 have discussed how to weigh these pros and cons in detail.

I hope that by discussing the benefits and drawbacks of Bayesian inference in a modern auditing context, I have made a convincing case for the use of Bayesian statistics in auditing theory and practice.

7.1.2 Development of Bayesian techniques

The second goal of this thesis was to develop innovative Bayesian statistical techniques for audit sampling. Auditors need easy-to-use statistical tools in order to use Bayesian statistics in their fieldwork. Chapter 5 describes the development of a default Bayesian hypothesis test for audit sampling which uses an impartial prior distribution. The impartial prior distribution is useful for auditors because it is simple to justify (i.e., it corresponds to the assumption that intolerable misstatement is equally likely as tolerable misstatement before seeing data) and because it can take into account pre-existing information that is familiar in an audit sampling context (e.g., the most likely expected misstatement). The statistical motivation for the impartial prior distribution is that the corresponding Bayes factor has desirable properties, that is, the Bayes factor will always quantify evidence in the direction supported by the data. Since it is also easy to calculate, the impartial Bayes factor can be a useful benchmark method or entry-level statistical tool for auditors to quantify audit evidence.

I hope that by creating an easy-to-use Bayesian hypothesis test, I have made a convincing case for the use of the Bayes factor when it comes to hypothesis testing in auditing theory and practice.

7.1.3 Accessibility of Bayesian methodology

The third goal of this thesis was to improve the accessibility of Bayesian methods to auditors via open-source software. In order for auditors to use Bayesian techniques in practice, they must have access to these tools at any time. A viable solution for improving auditors' access to Bayesian techniques is to implement these techniques in open-source software. Chapter 6 has discussed the development of JASP for Audit, an open-source software implementation of the Bayesian auditing framework for parameter estimation and hypothesis testing discussed in the preceding chapters. JASP for Audit has a number of design principles that are relevant for auditors in practice. For example, it offers both frequentist and Bayesian techniques; the user interface is designed for auditing and therefore user-friendly; and the software automatically makes the correct statistical choices and generates an audit report containing an explanation of the statistical results. This approach aims to minimize the dependency on an auditor's statistical knowledge, while increasing their understanding of the statistical theory underlying the audit process. By programming the fundamental Bayesian techniques for audit sampling into freely accessible software, this thesis ensures that any auditor is able to use Bayesian techniques in their audit at all times.

JASP for Audit has already proven to be a welcome addition to the audit software curriculum, as it is used in academic textbooks (e.g., Strang, 2022, Chapter 2), university courses and workshops. Additionally, we are working to expand JASP for Audit in partnership with departments of the Dutch government (Auditdienst Rijk, Uitvoering van Beleid SZW). While it is impossible to say how many times the implementation in JASP has been downloaded so far, the implementation in R has been downloaded over 21,000 times, further demonstrating that the output of this project has been useful for many auditors in practice.

I hope that by making the Bayesian techniques covered in this thesis accessible to all auditors, I have made a convincing case for the use of open-source software, particularly JASP for Audit, in auditing theory and practice.

7.2 Opportunities for further research and development

Because this thesis mainly covers the foundational concepts of Bayesian inference in auditing, there is a wide range of areas for future research and potential development. In this section I will focus on three research directions that are worth following up on.

7.2.1 Further development of Bayesian methodology

This thesis lays the groundwork for the improvement of auditing theory as well as the development of more sophisticated Bayesian statistical methodology. Research into this area can potentially lead to a further increase of transparency and efficiency in the audit.

First, further research could enhance the methodological toolbox for Bayesian parameter estimation in an audit context. With regard to parameter estimation, this thesis concentrated on how to incorporate pre-existing information into the prior distribution and the statistical model. However, it does not fully discuss how to take advantage of the other benefits offered by Bayesian inference, such as sequential sampling (Rouder, 2014). Sequential sampling allows auditors to modify their sampling procedures as new information becomes available, however, how to select the best samples has not been looked into. For instance, the auditor could use sequential sampling——choosing and auditing the item that will provide them with the most information (e.g., the largest reduction in uncertainty) each time. Further research into this idea is warranted as it can lead to an increase in efficiency. Additionally, the thesis makes no mention of the practice of estimating parameters when multiple statistical models are viable. Through Bayesian model averaging (Hinne et al., 2020), Bayesian inference can take into account the uncertainty about the composition of the model. In this approach, the auditor does not need to explicitly choose which model to use because estimates from various models can be averaged according to how well the models describe the data. The viability of this strategy in the context of audit sampling is worth investigating in more detail.

Second, additional research may enhance the statistical methodology available for Bayesian hypothesis testing in an audit context. In terms of testing hypotheses, this thesis has covered the fundamental concept underlying the Bayes factor and demonstrated how to compute it for typical auditing scenarios. The Bayes factor, however, has a wide variety of other applications that this thesis has not covered. For instance, it does not discuss how to compute the Bayes factor for the hypothesis of (in)tolerable misstatement in more intricate (e.g., multilevel) models. Future research could outline the properties of these Bayes factors, explain how to calculate them in complex models, and discuss the considerations that come with using them in an audit setting. Needless to say, these Bayes factor calculations also need to be made available in open-source software. Furthermore, by carrying out a Bayes factor design analysis (Schönbrodt and Wagenmakers, 2018), auditors can also use the Bayes factor to create a batched sample that provides as much evidence as possible. Finally, the Bayes factor is not limited to supporting a specific claim about the misstatement but can also be used to determine the relative strength of evidence for two statistical models used in estimating the misstatement in a population (Fragoso et al., 2018). Because of this, the Bayes factor can aid the auditor in choosing which model to apply in the sample evaluation. Although such Bayes factor computations are already available in JASP (van den Bergh et al., 2021), there has not been any discussion of them in the auditing literature.

Third, this thesis mainly takes a bottom-up approach to stratified sampling in auditing. It describes, for instance, how stratum estimates can be averaged to arrive at a population estimate. When an auditor needs to move from data to a conclusion, this bottom-up strategy is helpful. However, auditors often use a top-down strategy to allocate performance materiality when planning a stratified sample because it can help to minimize the required samples in each stratum. Prior research has suggested a method for allocating performance materiality to each stratum by decomposing the pursued population posterior distribution into required stratum posterior distributions. However, this research mainly focuses on the use of gamma distributions to represent independent strata (Sellke, 1983; Stewart et al., 2007; Stewart, 2013). In this thesis, we mostly use beta distributions as prior distributions because we believe they are easier to set up and interpret for auditors. Furthermore, we introduce a multilevel structure into the model to explicitly relate the strata to each other. The effect of both of these changes on the top-down approach to optimally allocate performance materiality has not been investigated in the auditing literature and requires further study.

7.2.2 Further investigating the role of the prior distribution

This thesis offers statistical techniques for using Bayesian statistics in an auditing setting, but it does not look into the actual application of these techniques by auditors and the consistency of statistical results across auditors. The nature of Bayesian inference makes it likely that prior distributions and statistical results will differ among auditors. Investigating what gives rise to these differences and mapping them is important because it can give an impression of the variation in the process of obtaining audit evidence among auditors and audit firms. Having this insight can assist audit standard-setters in offering recommendations to practitioners, for example regarding the minimal evidential strength of audit evidence.

First, future research could aim to map the variation in prior distributions and resulting Bayes factors for (in)tolerable misstatement. Chapter 2 introduced five methods to construct a prior distribution on the basis of pre-existing information. However, based on how each auditor interprets the information available to them, prior distributions (or prior probabilities) will likely differ amongst auditors. Further research could quantify the variation in Bayes factors between auditors for a given audit, giving an indication of the difference in audit evidence between individual auditors or audit firms. Such research has been conducted in the social sciences (Stefan et al., 2022b), where it was examined whether or not various prior distributions alter the direction of the Bayes factor, whether or not they result in a change in the category of evidence strength, and how much of an impact they have on the value of the Bayes factor. In auditing, it is important to rule out that such changes in professional judgment (and thus in the prior distribution) affect the decisions that are made based on an audit sampling procedure. Furthermore, since the Bayes factor has not been discussed in the auditing literature before, no research has been done into how auditors determine an appropriate evidence threshold. Further research could map the different evidence thresholds used by auditors and audit firms to gain an understanding about how much audit evidence is sufficient in their eyes.

Second, future research could investigate how to elicit a prior distribution from auditors' professional judgment. Despite the fact that the later chapters of this thesis emphasize the use of weakly informative prior distributions, Chapter 2 demonstrates that auditors can increase efficiency by specifying an informed prior distribution. However, auditors may not always have access to the information that is used to set up and justify these prior distributions, but they might still want to make a prior distribution based on their professional judgment or other audit information. This requires the need for prior elicitation (O'Hagan et al., 2006), a structured interview process designed for constructing a prior distribution based on the knowledge of one or multiple experts (i.e., auditors). Because prior elicitation is a structured process, it can help auditors to construct and justify a prior distribution that aligns with the situation in practice under many circumstances. However, prior elicitation comes with its own practical challenges (Stefan et al., 2022a). For instance, what method should be used to elicit the prior distribution, which model parameters are of interest and thus call for prior elicitation, and (how) should prior distributions from various auditors be combined? Further research into these matters in the context of an audit is warranted.

Last but not least, it is crucial to keep statistical models understandable in order to ensure that auditors can explain the results of their audits. The complexity of the statistical model is not necessarily constrained by this, but it does necessitate more practical guidance on the role of the prior distribution in more advanced Bayesian models for auditing. In Chapter 3, we specified weakly informative prior distributions mostly because it was more convenient. This made the prior distribution play a more supporting role. Nonetheless, the prior distribution has an impact on the results of the analysis, particularly for small sample sizes, and so it remains important to consider its interpretation. However, in more complex (e.g., multilevel) Bayesian models, this can be relatively challenging. For instance, it can be difficult to determine what the prior distribution on a (set of) transformed parameter(s) implies for the probability of misstatement. Further research into this topic can clarify the role of the prior distribution in more complex models, as well as how to construct them on the basis of pre-existing information and how to interpret them in a statistically sound way.

7.2.3 Building capacity for statistical auditing

One of the major factors limiting the expansion of research in the field of statistical auditing, and particularly in the area of Bayesian statistics, is the small number of researchers and professionals entering the field. As auditing firms can attest, the number of clients they serve is growing and will continue to grow in the future, necessitating the need for audits that are more effective and efficient. This means that the role of data and statistical inference in auditing will only continue to grow in the future. Unfortunately, there is currently not much capacity to develop and apply innovative methods in this field.

It is therefore critical that interest in statistical methods among auditing researchers and practitioners rises. In my opinion, our primary goals should be to increase the general knowledge and competency of statistical methodology among auditing students and practitioners; to broaden the guidance audit standard-setters give regarding statistical inference; and to increase the accessibility of complex statistical methodology to auditors. Of course, Bayesian statistics should take center stage in this.

Furthermore, it is important to create new strategies for attracting and retaining capable researchers and practitioners. These strategies may be put into place at different points along the career path, from the college years early on to spark interest in the field of statistical auditing to the professional years later on to retain the best researchers and practitioners. As history has shown, Bayesian statistics can thrive in this field if a large enough community collaborates to actively promote it.

7.3 Concluding comments

The benefits and possibilities of Bayesian statistics that initially piqued the interest of auditing researchers in the twentieth century are more relevant today than they have ever been. The ability of computers to perform advanced calculations continues to grow year after year, and as a result, the accessibility and complexity of Bayesian techniques available to auditors grows as well. A number of practical advantages currently lie at the auditor's fingertips, yet many of them do not know about—or find it difficult to act upon—these opportunities and continue to evaluate their samples in a familiar way. However, this thesis demonstrates that careful consideration of the statistical framework on which auditors base their opinions can unlock these practical benefits for auditors today.

Looking beyond the obvious, this thesis contains a more important message: Auditors must consider how to get the most out of their data. In practice, this means they should carefully consider the questions they want to address and decide on the best approach to take in order to provide the best possible answers. Seen from this perspective, I believe that applying Bayesian inference is merely a logical decision that will help the auditor match the situation in practice as closely as possible and provide a fitting answer to their questions. If this critical perspective on analyzing data can find a foothold in auditing theory and practice, it is not difficult to predict that auditors will employ more complex statistical planning and evaluation techniques in the near future.

Ultimately, the adoption of Bayesian methods rests with auditors in practice. I hope that this thesis will act as a roadmap for those who are willing to use these methods in the field.

References

Abdolmohammadi, M. J. (1985). Bayesian inference research in auditing: Some methodological suggestions. *Contemporary Accounting Research*, 2(1), 76–94. 17, 25

Abdolmohammadi, M. J. (1987). Bayesian inference in auditing: Evidence on the most appropriate assessment techniques. *Accounting and Business Research*, 17(68), 291–300. 17, 25

Ahmi, A. and Kent, S. (2013). The utilisation of generalized audit software (GAS) by external auditors. *Managerial Auditing Journal*, 28(2), 88–113. 172

Albert, M. (2003). Bayesian rationality and decision making: A critical review. *Analyse & Kritik*, 25(1), 101–117. 25

AlMarzouq, M., Zheng, L., Rong, G., and Grover, V. (2005). Open source: Concepts, benefits, and challenges. *Communications of the Association for Information Systems*, 16(1), 756–784. 9, 173

Altman, D. G. and Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *British Medical Journal*, 311(7003), 485. 100

American Institute of Certified Public Accountants (AICPA) (2019). *Audit Guide: Audit Sampling*. American Institute of Certified Public Accountants, New York. 3, 6, 16, 26, 49, 70, 72, 94, 95, 98, 106, 118, 141, 146, 152, 163, 176, 177, 191, 192

American Institute of Certified Public Accountants (AICPA) (2021). *Clarified Statements on Auditing Standards*. AICPA, New York, NY. 6, 94, 96, 97, 98, 101, 118, 189

Appelbaum, D., Kogan, A., and Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory*, 36(4), 1–27. 36, 42, 94, 132, 172

Appelbaum, D. A., Kogan, A., and Vasarhelyi, M. A. (2018). Analytical procedures in external auditing: A comprehensive literature survey and framework for external audit analytics. *Journal of Accounting Literature*, 40(1), 83–101. 8, 94

Armitage, P., McPherson, C., and Rowe, B. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society: Series A (General)*, 132(2), 235–244. 105

Arzhenovskiy, S. V., Bakhteev, A. V., Sinyavskaya, T. G., and Hahonova, N. N. (2019). Audit risk assessment model. *International Journal of Economics and Business Administration*, 7(1), 74–85. 37

Asadi, E., da Silva, M. C. G., and Costa, J. (2013). A systematic indoor air quality audit approach for public buildings. *Environmental Monitoring and Assessment*, 185(1), 865–875. 24

Beck, P. J., Solomon, I., and Tomassini, L. A. (1985). Subjective prior probability distributions and audit risk. *Journal of Accounting Research*, 23(1), 37–56. 17, 95

Bednarek, P. et al. (2016). Evaluating the usefulness of quantitative methods as analytical auditing procedures. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 434, 9–18. 94

Belastingdienst (2021). *Controleaanpak Belastingdienst—Uitgangspunten en Modellen*. Belastingdienst. 185

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, pages 551–572. 108

Bennett, G. B. and Hatfield, R. C. (2013). The effect of the social mismatch between staff auditors and client management on the collection of audit evidence. *The Accounting Review*, 88(1), 31–50. 94

Berger, J. O. and Mortera, J. (1999). Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association*, 94(446), 542–554. 142

Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American statistical Association*, 82(397), 112–122. 100, 132

Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle (2nd ed.)*. Institute of Mathematical Statistics. 105

Berkson, J. (1942). Tests of significance considered as evidence. *Journal of the American Statistical Association*, 37(219), 325–335. 100, 132

Betancourt, M. and Girolami, M. (2015). Hamiltonian monte carlo for hierarchical models. In *Current Trends in Bayesian Methodology with Applications*, pages 79–103. Chapman Hall/CRC Press. 57

Bickel, P. J. (1992). Inference and auditing: The Stringer bound. *International Statistical Review*, 60, 197–209. 72, 184, 188, 195, 198

Binck, D. (2012). Het effect van auditsoftware op de kwaliteit van de accountantscontrole. *Maandblad voor Accountancy en Bedrijfseconomie*, 88(1/2), 14–22. 172

Birnberg, J. G. (1964). Bayesian statistics: A review. *Journal of Accounting Research*, 2, 108–116. 3

Blocher, E. (1981). Assessment of prior distributions: The effect on required sample size in Bayesian audit sampling. *Accounting and Business Research*, 12(45), 11–20. 25

Blokdijk, J. (2004). Tests of control in the audit risk model: effective? efficient? *International Journal of Auditing*, 8(2), 185–194. 42

Boersma, M., Maliutin, A., Sourabh, S., Hoogduin, L., and Kandhai, D. (2020). Reducing the complexity of financial networks using network embeddings. *Scientific Reports*, 10(1), 1–15. 132, 172

Bollen, K. A., Ray, S., Zavisca, J., and Harden, J. J. (2012). A comparison of Bayes factor approximation methods including two new methods. *Sociological Methods & Research*, 41(2), 294–324. 103

Bradford, M., Henderson, D., Baxter, R. J., and Navarro, P. (2020). Using generalized audit software to detect material misstatements, control deficiencies and fraud: How financial and it auditors perceive net audit benefits. *Managerial Auditing Journal*, 35(4), 521–547. 172

Brazel, J. F. and Agoglia, C. P. (2007). An examination of auditor planning judgements in a complex accounting information system environment. *Contemporary Accounting Research*, 24(4), 1059–1083. 172

Brivot, M., Roussy, M., and Mayer, M. (2018). Conventions of audit quality: The perspective of public and private company audit partners. *Auditing: A Journal of Practice & Theory*, 37(2), 51–71. 16

Broeze, E. (2006). *Validation of Risk Assessments in Auditing.* PhD thesis, Amsterdam. 19

Brown, H. and Prescott, R. (2015). *Applied Mixed Models in Medicine.* John Wiley & Sons. 7

Brown-Liburd, H., Issa, H., and Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. *Accounting Horizons*, 29(2), 451–468. 16, 42, 171

Brown-Liburd, H. and Vasarhelyi, M. A. (2015). Big data and audit evidence. *Journal of Emerging Technologies in Accounting*, 12(1), 1–16. 172

Brydges, C. R. and Gaeta, L. (2019). An introduction to calculating Bayes factors in JASP for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research*, 62(12), 4523–4533. 174

Buitenhuis, C. (2022). Homogeen, of niet? Over homogeniteit van de populatie bij toepassing van de geldsteekproef. *Weekblad Fiscaal Recht*, 7413, 40–46. 176, 185

Cao, M., Chychyla, R., and Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–429. 42

Carpenter, B. (2016). Hierarchical partial pooling for repeated binary trials. `https://mc-stan.org/users/documentation/case-studies/pool-binary-trials.html`. 71

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32. 7, 61, 78

Casella, G. and Berger, R. L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of the American Statistical Association*, 82(397), 106–111. 142

CaseWare Analytics (2022). IDEA (Version 12)[Computer software]. 8, 43, 176

Chan, H. and Smieliauskas, W. (1990). Further tests of the modified moment bound in audit sampling of accounting populations. *Auditing: A Journal of Practice & Theory*, 9(3), 167–182. 17

Chesley, G. (1975). Elicitation of subjective probabilities: A review. *The Accounting Review*, 50(2), 325–337. 36

Chesley, G. (1978). Subjective probability elicitation techniques: A performance comparison. *Journal of Accounting Research*, 16(2), 225–241. 25

Christensen, B. E., Elder, R. J., and Glover, S. M. (2015). Behind the numbers: Insights into large audit firm sampling policies. *Accounting Horizons*, 29(1), 61–81. 4, 16, 132, 172

Cipriani, M., Costantini, R., and Guarino, A. (2012). A Bayesian approach to experimental analysis: Trading in a laboratory financial market. *Review of Economic Design*, 16(2), 175–191. 103

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4), 404–413. 188

Cochran, W. G. (1977). *Sampling Techniques*. John Wiley & Sons, New York. 59

Cockcroft, S. and Russell, M. (2018). Big data opportunities for accounting and finance practice and research. *Australian Accounting Review*, 28(3), 323–333. 42

Commissie Eindtermen Accountantsopleiding (CEA) (2020). *Eindtermen Accountantsopleidingen*. CEA. 172

Coram, P., Ng, J., and Woodliff, D. (2003). A survey of time budget pressure and reduced audit quality among australian auditors. *Australian Accounting Review*, 13(29), 38–44. 16

Corless, J. C. (1972). Assessing prior distributions for applying Bayesian statistics in auditing. *The Accounting Review*, 47(3), 556–566. 17, 96, 101

Crosby, M. A. (1980). Implications of prior probability elicitation on auditor sample size decisions. *Journal of Accounting Research*, 18(2), 585–593. 17

Crosby, M. A. (1981). Bayesian statistics in auditing: A comparison of probability elicitation techniques. *Accounting Review*, 56(2), 355–365. 19, 25

Cushing, B. E. and Loebbecke, J. K. (1983). Analytical approaches to audit risk: A survey and analysis. *Auditing: A Journal of Practice & Theory*, 3(1), 23–41. 37

Daroca, F. P. and Holder, W. W. (1985). The use of analytical procedures in review and audit engagements. 94

de Swart, J. J., Wille, J., and Majoor, B. (2013). Het ‘push left’-principe als motor van data analytics in de accountantscontrole. *Maandblad voor Accountancy en Bedrijfseconomie*, 87(oktober), 425–433. 10, 22, 28, 43, 171, 172, 176, 180

de Swart, J. J. B., Wille, F. J., and Zuur, M. (2016). Data analytics volgens het push-left principe. *Audit Magazine*, 1, 40–42. 172

Deakin, E. B. and Granof, M. H. (1974). Regression analysis as a means of determining audit sample size. *The Accounting Review*, 49(4), 764–771. 32

Deloitte (2021). Data and analytics: Riding the digitalization wave. `https://www2.deloitte.com/content/dam/Deloitte/my/Documents/risk/my-risk-data-analytics-brochure.pdf`. 42

Derks, K. (2022). *jfa: Bayesian and Classical Audit Sampling*. R package version 0.6.4. 12, 37, 152, 166, 176

Derks, K., de Swart, J., van Batenburg, P. C., Wagenmakers, E.-J., and Wetzels, R. (2021a). Priors in a Bayesian audit: How integration of existing information into the prior distribution can improve audit transparency and efficiency. *International Journal of Auditing*, 25(3), 621–636. 44, 45, 63, 101, 107, 133, 134, 172, 175, 180

Derks, K., de Swart, J., Wagenmakers, E.-J., and Wetzels, R. (2022a). The Bayesian approach to audit evidence: Quantifying evidence using the Bayes factor. *PsyArXiv*. 44

Derks, K., de Swart, J., Wagenmakers, E.-J., Wille, J., and Wetzels, R. (2021b). JASP for audit: Bayesian tools for the auditing practice. *Journal of Open Source Software*, 6(68), 2733. 37, 76, 121, 152, 166, 173, 174, 176, 189

Derks, K., de Swart, J., and Wetzels, R. (2022b). Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor. *Maandblad voor Accountancy en Bedrijfseconomie*, 96(1/2), 37–46. 65, 71, 172

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective*. CRC Press, Boca Raton. 47

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274–290. 16, 29

Dilligent (2022). Audit Command Language (ACL)[Computer software]. 9, 43, 176

Dowling, C. and Leech, S. (2007). Audit support systems and decision aids: Current practice and opportunities for future research. *International Journal of Accounting Information Systems*, 8(2), 92–116. 94

Drake, P. D. and Nigrini, M. J. (2000). Computer assisted analytical procedures using Benford's law. *Journal of Accounting Education*, 18(2), 127–146. 109

Durtschi, C., Hillison, W., and Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, 5(1), 17–34. 108

Dutka, J. (1981). The incomplete beta function—a historical profile. *Archive for History of Exact Sciences*, 24, 11–29. 153

Dyckman, T. R. (2016). Significance testing: We can do better. *Abacus*, 52(2), 319–342. 132

Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, 58(5), 493–500. 16

Edmonds, M., Miller, T., and Savage, A. (2019). Accounts receivable: An audit simulation. *Journal of Accounting Education*, 47, 75–92. 3, 118

Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193–242. 106, 132

Efron, B. and Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, 236(5), 119–127. 58, 68

Elliott, R. K. and Rogers, J. R. (1972). Relating statistical sampling to audit objectives. *Journal of Accountancy*, 134, 46–55. 97, 98, 118

Etz, A. (2018). Introduction to the concept of likelihood and its applications. *Advances in Methods and Practices in Psychological Science*, 1(1), 60–69. 19, 46, 135

European Commission (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. 113

Faulkenberry, T. J. (2019). A tutorial on generalizing the default Bayesian t-test via posterior sampling and encompassing priors. *Communications for Statistical Applications and Methods*, 26(2), 217–238. 137

Faulkenberry, T. J., Ly, A., and Wagenmakers, E.-J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition*, 6(2), 231–259. 174

Felix, W. L. (1976). Evidence on alternative means of assessing prior probability distributions for audit decision making. *The Accounting Review*, 51(4), 800–807. 7, 17

Fisher, R. A. (1934). *Statistical Methods for Research Workers (5th ed.)*. Oliver and Boyd, London. 99

Flint, D. (1988). *Philosophy and Principles of Auditing: An Introduction*. Macmillan Education. 2

Foster, E. D. and Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, 105(2), 203–206. 8

Fragoso, T. M., Bertoli, W., and Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86(1), 1–28. 76, 207

Gao, Y., Kennedy, L., Simpson, D., and Gelman, A. (2021). Improving multilevel regression and poststratification with structured priors. *Bayesian Analysis*, 16(3), 719–744. 62, 63

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–534. 64, 65

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27(15), 2865–2873. 53

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2013). *Bayesian Data Analysis (3rd ed.)*. Chapman and Hall/CRC, London. 8, 47, 58, 61, 63, 64, 135

Gelman, A. and Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge university press. 48, 53, 57, 58, 63, 77

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4), 1360–1383. 62

Gelman, A. and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2), 127–135. 59

Gepp, A., Linnenluecke, M. K., O'Neill, T. J., and Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, 40(1), 102–115. 16, 42, 132

Ghosh, J., Li, Y., and Mitra, R. (2018). On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, 13(2), 359–383. 62

Gillett, P. R. and Srivastava, R. P. (2000). Attribute sampling: A belief-function approach to statistical audit evidence. *Auditing: A Journal of Practice & Theory*, 19(1), 145–155. 94

Glover, S. M., Prawitt, D. F., and Drake, M. S. (2015). Between a rock and a hard place: A path forward for using substantive analytical procedures in auditing large P&L accounts: Commentary and analysis. *Auditing: A Journal of Practice & Theory*, 34(3), 161–179. 4

Godfrey, J. and Neter, J. (1984). Bayesian bounds for monetary unit sampling in accounting and auditing. *Journal of Accounting Research*, 22(2), 497–525. 17

Good, I. J. (1967). A Bayesian significance test for multinomial distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 29(3), 399–418. 110

Goodman, S. N. and Royall, R. (1988). Evidence and scientific research. *American Journal of Public Health*, 78(12), 1568–1574. 95

Gronau, Q. F., Ly, A., and Wagenmakers, E.-J. (2019). Informed Bayesian t-tests. *The American Statistician*, 74(2), 137–143. 174

Gunel, E. and Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, 61(3), 545–557. 114

Hall, T. W., Hoogduin, L. A., Pierce, B. J., and Tsay, J. J. (2021). Designing efficient stratified mean-per-unit sampling applications in accounting and auditing. *Journal of Accounting, Auditing & Finance*. 59

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, pages 3315–3323. 113

Harrison, K. E., Srivastava, R. P., and Plumlee, R. D. (2002). Auditors' evaluations of uncertain audit evidence: Belief functions versus probabilities. In *Belief Functions in Business Decisions*, pages 161–183. Springer. 27

Hartley, H. and Fitch, E. (1951). A chart for the incomplete beta-function and the cumulative binomial distribution. *Biometrika*, 38(3/4), 423–426. 153

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York. 58

Heck, D. W., Böhm, U., Böing-Messing, F., Bürkner, P.-C., Derks, K., Dienes, Z., Fu, Q., Gu, X., Karimova, D., Kiers, H. A. L., Klugkist, I., Kuiper, R. M., Lee, M. D., Leenders, R., Leplaa, H. J., Linde, M., Ly, A., Meijerink-Bosman, M., Moerbeek, M., J, M., B, P., Schönbrodt, F. D., Tendeiro, J. N., van den Bergh, D., van Lissa, C., van Ravenzwaaij, D., Vanpaemel, W., Wagenmakers, E.-J., Williams, D. R., Zondervan-Zwijnenburg, M., and Hoijtink, H. (2022). A review of applications of the Bayes factor in psychological research. *Psychological Methods*. 103

Higgins, H. N. and Nandram, B. (2009). Monetary unit sampling: Improving estimation of the total audit error. *Advances in Accounting*, 25(2), 174–182. 7

Hinne, M., Gronau, Q. F., van den Bergh, D., and Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, 3(2), 200–215. 206

Hirst, D. E. and Koonce, L. (1996). Audit analytical procedures: A field investigation. *Contemporary Accounting Research*, 13(2), 457–486. 94

Hitzig, N. B. (1995). Audit sampling: A survey of current practice. *The CPA Journal*, 65(7), 54. 16

Hoekstra, R., Morey, R. D., Rouder, J. N., and Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5), 1157–1164. 9, 21, 172

Hoijtink, H., Mulder, J., van Lissa, C., and Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, 24(5), 539–556. 37

Huang, F., No, W. G., Vasarhelyi, M. A., and Yan, Z. (2022). Audit data analytics, machine learning, and full population testing. *The Journal of Finance and Data Science*, 8, 138–144. 42

Hubbard, R. and Lindsay, R. M. (2008). Why p values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology*, 18(1), 69–88. 95

International Auditing and Assurance Standards Board (IAASB) (2018). *Handbook of international quality control, auditing review, other assurance, and related services pronouncements (Vol. I)*. International Federation of Accountants, New York, NY. 6, 16, 17, 26, 27, 28, 42, 93, 94, 96, 97, 98, 101, 117, 118, 132, 172, 189

International Forum of Independent Audit Regulators (IFIAR) (2020). Survey of inspection findings 2019. `https://www.ifiar.org/?wpdmdl=10453`. 173

International Forum of Independent Audit Regulators (IFIAR) (2021). Survey of inspection findings 2020. `https://www.ifiar.org/?wpdmdl=12436`. 173

International Forum of Independent Audit Regulators (IFIAR) (2022). Survey of inspection findings 2021. `https://www.ifiar.org/?wpdmdl=13957`. 173

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., and Wagenmakers, E.-J. (2017). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, 49(2), 638–652. 114, 127

JASP Team (2022). JASP (Version 0.16.3)[Computer software]. 9, 12, 37, 76, 103, 116, 152, 166, 174

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, 31, 203–222. 132

Jeffreys, H. (1939). *Theory of probability (1st ed.)*. Oxford University Press, Oxford. 2, 101, 132

Jeffreys, H. (1961). *Theory of probability (3rd ed.)*. Oxford University Press, Oxford. 103, 104, 138

Jiambalvo, J. and Waller, W. (1984). Decomposition and assessments of audit risk. *Auditing: A Journal of Practice & Theory*, 3(2), 80–88. 17, 37

Johnstone, D. J. (1986). Tests of significance in theory and practice. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 35(5), 491–498. 95

Johnstone, D. J. (1990). Sample size and the strength of evidence: A Bayesian interpretation of binomial tests of the information content of qualified audit reports. *Abacus*, 26(1), 17–35. 95, 99, 132

Johnstone, D. J. (1994). A statistical paradox in auditing. *Abacus*, 30(1), 44–49. 7, 118, 132

Johnstone, D. J. (2018). Accounting theory as a Bayesian discipline. *Foundations and Trends in Accounting*, 13(1-2), 1–266. 8, 95, 96

Johnstone, D. J. (2021). Accounting research and the significance test crisis. *Critical Perspectives on Accounting*, page 102296. 95, 132

Kachelmeier, S. J., Majors, T., and Williamson, M. G. (2014). Does intent modify risk-based auditing? *The Accounting Review*, 89(6), 2181–2201. 16

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American statistical association*, 90(430), 773–795. 8, 76, 95, 102, 132

Kearns, M., Neel, S., Roth, A., and Wu, Z. S. (2018). Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *International Conference on Machine Learning*, pages 2564–2572. PMLR. 113

Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to Bayesian inference with JASP. *BMC Medical Research Methodology*, 20(1), 1–12. 174

Kerman, J. (2011a). A closed-form approximation for the median of the beta distribution. *arXiv preprint 1111.0433*. 61, 143, 144, 154, 160

Kerman, J. (2011b). Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics*, 5, 1450–1470. 135

Keysers, C., Gazzola, V., and Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, 23(7), 788–799. 100

Kim, J. H., Ahmed, K., and Ji, P. I. (2018). Significance testing in accounting research: A critical evaluation based on evidence. *Abacus*, 54(4), 524–546. 95

Kinney, W. R. (1975). Decision theory aspects of internal control system design/-compliance and substantive tests. *Journal of Accounting Research*, 13, 14–29. 7, 95

Kinney, W. R. (1983). A note on compounding probabilities in auditing. *Auditing: A Journal of Practice & Theory*, 2(2), 13–22. 17, 37

Kinney, W. R. (1984). Discussant's response to leslie (1984): "An analysis of the audit framework focusing on inherent risk and the role of statistical sampling in compliance testing". In *Proceedings of the 1984 Touche Ross University of Kansas Symposium on auditing problems*, pages 126–132. Lawrence, KA: School of Business, University of Kansas. 3

Kinney, W. R. and Bailey, A. D. (1976). Regression analysis as a means of determining audit sample size: A comment. *The Accounting Review*, 51(2), 396–401. 32

Kirkham, E. and Gaa, C. (1939). Is there a theory basis for audit procedure? *The Accounting Review*, 14(2), 139–146. 2

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. Sage. 41

Kloosterman, H. (2018). Korte samenvatting Leslie, Teitlebaum & Anderson. In *Essaybundel Statistical Auditing*, pages 34–42. Limperg Instituut. 176, 179, 183, 184

Klugkist, I., Kato, B., and Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, 59(1), 57–69. 137

Knoblett, J. A. (1970). The applicability of Bayesian statistics in auditing. *Decision Sciences*, 1(3–4), 423–440. 17, 133

Kogan, A., Mayhew, B. W., and Vasarhelyi, M. A. (2019). Audit data analytics research—An application of design science methodology. *Accounting Horizons*, 33(3), 69–73. 172

Koninklijke Nederlandse Beroepsorganisatie van Accountants (NBA) (2017). Nba-handreiking 1141, data-analyse bij de controle. `https://www.nba.nl/globalassets/wet--en-regelgeving/nba-handreikingen/1141/nba-handreiking-1141-data-analyse---18-juni-2019.pdf`. 42

Kraft Jr, W. H. (1968). Statistical sampling for auditors: A new look. *Journal of Accountancy*, 126(2), 49. 2

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS (2nd ed.)*. Academic Press, New York. 46

Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour*, 5(10), 1282–1291. 75

Kruschke, J. K. and Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, 25(1), 155–177. 21, 46

Laws, D. J. and O'Hagan, A. (2000). Bayesian inference for rare errors in populations with unequal unit sizes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(4), 577–590. 8, 136

Laws, D. J. and O'Hagan, A. (2002). A hierarchical Bayes model for multilocation auditing. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(4), 431–450. 8, 17, 58

Lee, L., Kerler, W., and Ivancevich, D. (2018). Beyond Excel: Software tools and the accounting curriculum. *AIS Educator Journal*, 13(1), 44–61. 16

Lee, M. D. and Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge university press. 50

Lee, T.-H. and Ali, A. M. (2008). The evolution of auditing: An analysis of the historical development. *Journal of Modern Accounting and Auditing*, 4(12), 1–8. 2

Lehmann, E. L. and Romano, J. P. (2006). *Testing Statistical Hypotheses*. Springer. 99, 136

Leslie, D. A. (1984). An analysis of the audit framework focusing on inherent risk and the role of statistical sampling in compliance testing. In *Proceedings of the 1984 Touche Ross University of Kansas Symposium on auditing problems*, pages 126–132. Lawrence, KA: School of Business, University of Kansas. 3, 7, 17, 37, 96

Leslie, D. A., Teitlebaum, A. D., and Anderson, R. J. (1979). *Dollar-Unit Sampling: A Practical Guide for Auditors*. Copp Clark Pitman, London. 181

Leung, P., Coram, P., and Cooper, B. (2007). *Modern Auditing & Assurance Services*. John Wiley & Sons. 2

Li, C., Raman, K., Sun, L., and Yang, R. (2020). The SOX 404 control audit and the effectiveness of additional audit effort in lowering the risk of financial misstatements. *Review of Quantitative Finance and Accounting*, 54(3), 981–1009. 94

Li, X. (2022). Behavioral challenges to professional skepticism in auditors' data analytics journey. *Maandblad voor Accountancy en Bedrijfseconomie*, 96(1/2), 27–36. 172

Lin, G. D. and Hu, C.-Y. (2008). On characterizations of the logistic distribution. *Journal of Statistical Planning and Inference*, 138(4), 1147–1156. 50

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, 43(1/2). 148

Lindley, D. V. (1993). The analysis of experimental data: The appreciation of tea and wine. *Teaching Statistics*, 15(1), 22–25. 105

Lindley, D. V. (2004). That wretched prior. *Significance*, 1(2), 85–87. 17

Liu, C. C. and Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, 52(6), 362–375. 37

Lombardi, D., Bloch, R., and Vasarhelyi, M. (2014). The future of audit. *Journal of Information Systems and Technology Management*, 11(1), 21–32. 16

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., et al. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, 88(2), 1–17. 8, 103, 152, 166, 174

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs - a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337. 7

Ly, A. (2018). *Bayes Factors for Research Workers*. PhD thesis, Amsterdam. 151

Ly, A., Raj, A., Etz, A., Marsman, M., Gronau, Q. F., and Wagenmakers, E.-J. (2018). Bayesian reanalyses from summary statistics: A guide for academic consumers. *Advances in Methods and Practices in Psychological Science*, 1(3), 367–374. 174

Ly, A., Verhagen, J., and Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, 72, 19–32. 103, 145

Lynch, S. M. and Bartlett, B. (2019). Bayesian statistics in sociology: Past, present, and future. *Annual Review of Sociology*, 45, 47–68. 103

Ma'arif, M. Y., Mohd Satar, N. S., Samah, A., and Armi, M. (2020). Detecting ERP fraud using the first digits formula of Benford's law. *Science International*, 32(3), 439–444. 108

Maingot, M. and Quon, T. K. (2009). Sampling practices of internal auditors at corporations on the standard & poor's toronto stock exchange composite index. *Accounting Perspectives*, 8(3), 215–234. 16

Martel-Escobar, M., Vázquez-Polo, F. J., and Hernández-Bastida, A. (2005). Analysing the independence hypothesis in models for rare errors: An application to auditing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(4), 795–804. 37, 151

Martel-Escobar, M., Vázquez-Polo, F.-J., and Hernández-Bastida, A. (2018). Bayesian inference in auditing with partial prior information using maximum entropy priors. *Entropy*, 20(12), 919–928. 7, 8, 17, 25, 118

Mcbride, G. B. and Ellis, J. C. (2001). Confidence of compliance: A Bayesian approach for percentile standards. *Water Research*, 35(5), 1117–1124. 141

Meeden, G. (2003). A Bayesian solution for a statistical auditing problem. *Journal of the American Statistical Association*, 98(463), 735–740. 7, 17

Morey, R. D. and Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, 16(4), 406–419. 137

Morgan, L. and Finnegan, P. (2007). Benefits and drawbacks of open source software: An exploratory study of secondary software firms. In *IFIP International Conference on Open Source Systems*, pages 307–312. Springer. 9, 173

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90–100. 46, 47

Neter, J. and Godfrey, J. (1985). Robust Bayesian bounds for monetary-unit sampling in auditing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(2), 157–168. 136

Neter, J. and Loebbecke, J. K. (1975). *Auditing research monograph No. 2: Behavior of major statistical estimators in sampling accounting populations*. American Institute of Certified Public Accountants, New York. 16, 17

Neter, J. and Loebbecke, J. K. (1977). On the behavior of statistical estimators when sampling accounting populations. *Journal of the American Statistical Association*, 72(359), 501–507. 16

Nichols, D. R. and Baker, R. (1977). Testing the consistency of auditors' prior distributions and sampling results. *Abacus*, 13(2), 91–105. 132

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5(2), 241. 9

Nigrini, M. J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. John Wiley & Sons. 109

Nigrini, M. J. and Mittermaier, L. J. (1997). The use of Benford's law as an aid in analytical procedures. *Auditing*, 16(2), 52–67. 109

Normanton, E. L. (1966). *The Accountability and Audit of Governments: A Comparative Study*. Manchester University Press. 2

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). Uncertain judgements: Eliciting experts' probabilities. 208

Park, D. K., Gelman, A., and Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, 12(4), 375–385. 59

Pearson, K. (1934). *Tables of the Incomplete Beta Function*. Biometrika, University College, London. 153, 188

Perry, T. P. M. (2011). SAS 106—Audit evidence: Is it effective? *Available at SSRN 1907438*. 94

Power, M. K. (1992). From common sense to expertise: Reflections on the prehistory of audit sampling. *Accounting, Organizations and Society*, 17(1), 37–62. 16, 132, 171

Pratt, J. W. (1965). Bayesian interpretation of standard inference statements. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2), 169–203. 141

Public Company Accounting Oversight Board (PCAOB) (2017). Briefing paper: Use of data and technology in the conduct of audits. `https://pcaob-assets.azureedge.net/pcaob-dev/docs/default-source/news/events/documents/05242017-sag-meeting/dta-briefing-paper-may-2017.pdf?sfvrsn=b03c2128_0`. 42

Public Company Accounting Oversight Boards (PCAOB) (2020). *Auditing Standards*. PCAOB, Washington, DC. 6, 94, 97, 98, 189

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. MIT Press, Cambridge. 153

Ramage, J. G., Krieger, A. M., and Spero, L. L. (1979). An empirical study of error characteristics in audit populations. *Journal of Accounting Research*, 17, 72–102. 16

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3), 23–49. 9, 173

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 50, 67, 72, 76, 78, 116, 174, 176, 189

Richard, M. and Vecer, J. (2021). Efficiency testing of prediction markets: Martingale approach, likelihood ratio and Bayes factor analysis. *Risks*, 9(2), 31. 103

Roberts, D. M. (1975). Discussion of the real risks in audit sampling. *Journal of Accounting Research*, 13, 92–94. 117

Roberts, D. M. (1978). *Statistical Auditing*. American Institute of Certified Public Accountants, New York. 54, 59, 117

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21(2), 301–308. 29, 44, 96, 104, 105, 206

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. 103, 112, 125

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237. 103, 112, 116, 133

Sahu, S. and Smith, T. (2006). A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(2), 235–253. 17

Salijeni, G., Samsonova-Taddei, A., and Turley, S. (2019). Big Data and changes in audit technology: Contemplating a research agenda. *Accounting and Business Research*, 49(1), 95–119. 4, 7, 171

Sarafoglou, A., Haaf, J. M., Ly, A., Gronau, Q. F., Wagenmakers, E.-J., and Marsman, M. (2021). Evaluating multinomial order restrictions with bridge sampling. *Psychological Methods*. 110, 123

Schönbrodt, F. D. and Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. 207

Schouten, R. (2007). De toegevoegde waarde van audit software. *Maandblad voor Accountancy en Bedrijfseconomie*, 81(11), 521–530. 172

Scott, W. R. (1973). A Bayesian approach to asset valuation and audit size. *Journal of Accounting Research*, 11(2), 304–330. 95

Sellke, T. (1983). On the fitting of Pearson curves to sums of independent random variables. Technical report, Department of Statistics, Stanford University. 207

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71. 71

Sorensen, J. E. (1969). Bayesian analysis in auditing. *The Accounting Review*, 44(3), 555–561. 99

Srivenkataramana, T. (2018). Application of statistical sampling to audit and control. *DHARANA–Bhavan's International Journal of Business*, 12(1), 14–19. 16

Steele, A. (1992). *Audit Risk and Audit Evidence: The Bayesian Approach to Statistical Auditing*. Academic Press, London. 7, 19, 28, 43, 107, 139

Stefan, A. M., Evans, N. J., and Wagenmakers, E.-J. (2022a). Practical challenges and methodological flexibility in prior elicitation. *Psychological Methods*, 27(2), 177–197. 208

Stefan, A. M., Katsimpokis, D., Gronau, Q. F., and Wagenmakers, E.-J. (2022b). Expert agreement in prior elicitation and its effects on Bayesian inference. *Psychonomic Bulletin & Review*, pages 1–19. 207

Stewart, T. (2012). *Technical Notes on the AICPA Audit Guide Audit Sampling*. American Institute of Certified Public Accountants, New York. 3, 19, 42, 43, 46, 50, 70, 72, 95, 98, 136, 141, 172

Stewart, T., Strijbosch, L., Moors, H., and van Batenburg, P. C. (2007). A simple approximation to the convolution of gamma distributions. *CentER Discussion Paper, no. 2006-27, ISSN 0924-7815*. 19, 207

Stewart, T. R. (2013). *A Bayesian Audit Assurance Model with Application to the Component Materiality Problem in Group Audits*. PhD thesis, Amsterdam. 3, 7, 16, 17, 19, 21, 36, 43, 44, 95, 101, 107, 136, 139, 162, 172, 207

Strang, K. D. (2022). Petroleum industry contingency planning using auditing theories and inferential statistics. In *Global Risk and Contingency Management Research in Times of Crisis*, pages 15–28. IGI Global. 205

Stringer, K. W. and Stewart, T. R. (1986). *Statistical Techniques for Analytical Review in Auditing*. John Wiley & Sons. 7, 32

Swinamer, K., Lesperance, M., and Hartmut, W. (2004). Optimal bounds used in dollar-unit sampling: A comparison of reliability and efficiency. *Communications in Statistics—Simulation and Computation*, 33(1), 109–143. 17

Symon, I. W. (1974). The place of statistics in the education of chartered accountants. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 23(1), 67–70. 172

Thatcher, A. R. (1964). Relationships between Bayesian and confidence limits for predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 176–192. 141

Titera, W. R. (2013). Updating audit standard——Enabling audit data analysis. *Journal of Information Systems*, 27(1), 325–331. 16

Touw, P. and Hoogduin, L. (2012). *Statistiek voor Audit en Controlling*. Academic Service, Den Haag. 3, 16, 19, 29, 175, 188

Trompeter, G. and Wright, A. (2010). The world has changed—Have analytical procedure practices? *Contemporary Accounting Research*, 27(2), 669–700. 94

Tsui, K.-W., Matsumura, E. M., and Tsui, K.-L. (1985). Multinomial-Dirichlet bounds for dollar-unit sampling in auditing. *Accounting Review*, 60(1), 76–96. 17

Tuyl, F., Gerlach, R., and Mengersen, K. (2008). A comparison of Bayes-Laplace, Jeffreys, and other priors: the case of zero events. *The American Statistician*, 62(1), 40–44. 25

van Batenburg, P. C. (2018a). Data-analyse en statistical auditing: Maakt data-analyse steekproeven overbodig? (deel 1). In *Essaybundel Statistical Auditing*, pages 147–148. Limperg Instituut. 171, 172

van Batenburg, P. C. (2018b). Integrale controle: Als het kan, dan moet het? (deel 5). In *Essaybundel Statistical Auditing*, pages 156–157. Limperg Instituut. 171, 172

van Batenburg, P. C. (2018c). Rekenen met plaatjes: Een proefschrift over een nieuwe kijk op steekproeven. In *Essaybundel Statistical Auditing*, pages 225–229. Limperg Instituut. 180

van Batenburg, P. C. (2018d). Steekproefselectie, de basis. In *Essaybundel Statistical Auditing*, pages 103–105. Limperg Instituut. 181

van Batenburg, P. C. (2018e). Uitbreiden van steekproeven om beter te schatten. In *Essaybundel Statistical Auditing*, pages 119–121. Limperg Instituut. 95

van Batenburg, P. C. and Kriens, J. (1989). Bayesian discovery sampling: A simple model of Bayesian inference in auditing. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 38(4), 227–233. 7, 30

van Batenburg, P. C., O'Hagan, A., and Veenstra, R. H. (1994). Bayesian discovery sampling in financial auditing: A hierarchical prior model for substantive test sample sizes. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1), 99–110. 31, 43

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., et al. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), 1–26. 9, 203

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217. 2

van den Acker, C. (2000). Belief-function representation of statistical audit evidence. *International Journal of Intelligent Systems*, 15(4), 277–290. 94

van den Bergh, D., Clyde, M. A., Gupta, A. R., de Jong, T., Gronau, Q. F., Marsman, M., Ly, A., and Wagenmakers, E.-J. (2021). A tutorial on Bayesian multi-model linear regression with BAS and JASP. *Behavior Research Methods*, 53, 1–21. 174, 207

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, Š., Gupta, A. R. K. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., and Wagenmakers, E.-J. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique*, 120(1), 73. 174

Van Der Nest, D., Smidt, L., and Lubbe, D. (2015). The application of statistical and/or non-statistical sampling techniques by internal audit functions in the south african banking industry. *Risk Governance and Control: Financial Markets & Institutions*, 5(1), 71–80. 16, 172

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Komarlu Narenda Gupta, A. R., Sarafoglou, A., Stefan, A., Voekel, J. G., and Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3), 813–826. 75, 103, 138

van Ravenzwaaij, D. and Etz, A. (2021). Simulation studies as a tool to understand Bayes factors. *Advances in Methods and Practices in Psychological Science*, 4(1). 145

Varma, D. T. and Khan, D. (2012). Fraud detection in supply chain using Benford distribution. *International Journal of Research in management*, 5(2). 109

Verkuilen, J. and Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, 37(1), 82–113. 71

Von Krogh, G. and Spaeth, S. (2007). The open source software phenomenon: Characteristics that promote research. *The Journal of Strategic Information Systems*, 16(3), 236–253. 173

Voorhoeve, K. (2018). Statistical auditing voor dummies. In *Essaybundel Statistical Auditing*, pages 267–275. Limperg Instituut. 180

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. 95, 96, 100, 104, 105, 132

Wagenmakers, E.-J., Gronau, Q. F., and Vandekerckhove, J. (2019). Five Bayesian intuitions for the stopping rule principle. *PsyArXiv*. 104, 106, 146

Wagenmakers, E.-J. and Kucharský, Š. (2020). *The JASP Data Library*. 177

Wagenmakers, E.-J., Lee, M., Lodewyckx, T., and Iverson, G. J. (2008). Bayesian versus frequentist inference. In *Bayesian evaluation of informative hypotheses*, pages 181–207. Springer. 4, 16, 17, 29, 44, 94, 175

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60(3), 158–189. 123

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Smira, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., and Morey, R. D. (2018a). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76. 174

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., and Morey, R. D. (2018b). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. 7

Wasserstein, R. L. and Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. 104

Wei, A. and Vellwock, A. E. (2020). Is COVID–19 data reliable? A statistical analysis with Benford's law. 108

Westfall, P. H., Johnson, W. O., and Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika*, 84(2), 419–427. 112

Wetzels, R., Grasman, R. P., and Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for ANOVA designs. *The American Statistician*, 66(2), 104–111. 103, 112, 116, 133

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., and Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6(3), 291–298. 103, 138, 147, 148, 149, 150

Wetzels, R., Raaijmakers, J. G., Jakab, E., and Wagenmakers, E.-J. (2009). How to quantify support for and against the null hypothesis: A flexible winbugs implementation of a default Bayesian t test. *Psychonomic Bulletin & Review*, 16(4), 752–760. 103, 112

Wetzels, R. and Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19(6), 1057–1064. 103, 116, 133

Yang, R. and Berger, J. O. (1996). *A Catalog of Noninformative Priors*. Institute of Statistics and Decision Sciences, Duke University Durham. 61

Yin, X. (2020). Audit evidence classification and collection techniques in China and the US. *The Frontiers of Society, Science and Technology*, 2(14), 182–190. 94

Yoon, K., Hoogduin, L., and Zhang, L. (2015). Big data as complementary audit evidence. *Accounting Horizons*, 29(2), 431–438. 42

Yoon, K. and Pearce, T. (2021). Can substantive analytical procedures with data and data analytics replace sampling as tests of details? *Journal of Emerging Technologies in Accounting*, 18(2), 185–199. 16, 42, 172

Summary

In this thesis, I make the case that auditors can benefit greatly by adopting Bayesian statistics into their methodological toolbox. Auditors often use statistical sampling to efficiently form an opinion about the misstatement in a population. At this stage of the audit, they typically have access to pre-existing information from earlier audit activities. However, the currently dominant frequentist techniques for statistical audit sampling cannot make optimal use of this pre-existing information. This is unfortunate because it means that much relevant information is discarded during the statistical analysis. Bayesian statistics can incorporate the pre-existing information into the statistical analysis in a responsible way, leading to a potential decrease in uncertainty and an increase in efficiency for the auditor.

The thesis has three objectives: First, it aims to advance Bayesian inference in the auditing profession by updating its theoretical underpinnings and outlining its key practical benefits; second, it seeks to develop innovative and simple-to-use Bayesian statistical methods for parameter estimation and hypothesis testing specifically tailored to audit sampling; and third, it seeks to make these techniques openly available to all audit practitioners via user-friendly open-source statistical software.

The first part of the thesis discusses Bayesian parameter estimation in audit sampling, focusing on the use of pre-existing information. Chapter 2 introduces the basic ideas behind Bayesian parameter estimation and covers how information can be incorporated into the prior distribution. The key idea that this chapter introduces is that the prior distribution enables auditors to statistically build upon pre-existing information about the misstatement in their sampling procedures. In this chapter, I go over how this has a number of practical benefits for the auditor, including the potential for a smaller sample size and an increase in transparency. In addition, the chapter presents five methods for creating a prior distribution in the context of audit sampling. These prior distributions are created using familiar audit information, enabling the auditor to justify them in a logical manner. Chapter 3 builds on the fundamental ideas of Bayesian parameter estimation from Chapter 2 and discusses how additional data can be incorporated into the statistical model. The key idea introduced in this chapter is that the statistical model enables auditors to statistically base their sampling procedures on multiple sources

of data. In this chapter, I go over how this has two benefits for auditors: First, it helps them develop a more nuanced understanding of the population because they can explain the impact of the integrated data; and second, it makes it easier for them to identify potential misstatements in the population because they can more sharply differentiate between items. The chapter uses two examples to demonstrate how a Bayesian modeling approach can help the auditor develop a statistical model that is aligned with the situation in practice.

The second part of the thesis discusses Bayesian hypothesis testing, focusing on the quantification of statistical audit evidence. Chapter 4 discusses the concept of audit evidence and introduces Bayesian hypothesis testing as a method for quantifying such evidence from statistical samples. The key idea this chapter introduces is that statistical audit evidence can be quantified via the Bayes factor, a measure of relative evidence comparing two hypotheses being tested. It is demonstrated in the chapter that frequentist hypothesis testing has a number of practical drawbacks that Bayesian hypothesis testing using the Bayes factor can address. The use of the Bayes factor is then illustrated in a number of scenarios from a modern auditing context. Furthermore, Chapter 5 develops a default Bayesian hypothesis test for audit sampling. The Bayes factor is sensitive to the specification of the prior distribution, but the impact of prior on the Bayes factor has not been previously investigated in the context of audit sampling. Unfortunately, prior distributions that are tempting to use in the context of audit sampling can produce Bayes factors that quantify evidence in the opposite direction to what the data indicate. In order to address this issue, this chapter proposes an impartial prior distribution. The resulting Bayes factor is consistent, that is, it will always quantify evidence in favor of the hypothesis that is best supported by the data. The key ideas in this chapter are that the impartial Bayesian hypothesis test is appropriate in a variety of circumstances, that it is straightforward to use, and that it is easy to justify.

The third part of this thesis discusses a software implementation of the Bayesian statistical framework advocated in the preceding chapters. Improving the accessibility of Bayesian statistical techniques for auditors was one of the goals of this project. Because of this, JASP for Audit was created as an add-on module for the existing software JASP, an open-source statistical program with a graphical user interface. Chapter 6 gives an overview of JASP for Audit, which provides both frequentist and Bayesian statistical techniques for audit sampling with the goal of assisting auditors in the statistical aspects of an audit. It accomplishes this by providing, among other analyses, a guided workflow that adheres to the familiar four-step audit sampling process, performs the necessary statistical calculations, and automatically produces an audit report containing the statistical results and the interpretation of these results. This chapter discusses the benefits of JASP for Audit for the auditing practice and offers a thorough walkthrough of three real-world examples. Furthermore, it provides recommendations for the use of JASP for Audit in practice.

In sum, this thesis aims to develop an innovative approach to statistical auditing. By providing a full Bayesian framework for parameter estimation and hypothesis testing and making these techniques freely available, this thesis ensures that auditors can reap the benefits of Bayesian inference at all times.

Nederlandse Samenvatting

In dit proefschrift betoog ik dat auditors veel baat kunnen hebben bij het opnemen van Bayesiaanse statistiek in hun methodologische gereedschapskist. Auditors maken vaak gebruik van statistische steekproeven om op efficiënte wijze een oordeel te vormen over de fouten in een populatie. In dit stadium van de controle hebben zij meestal toegang tot reeds bestaande informatie uit eerdere controlewerkzaamheden. De momenteel dominante frequentistische technieken voor statistische steekproeven kunnen echter niet optimaal gebruik maken van deze reeds bestaande informatie. Dit is jammer, want het betekent dat veel relevante informatie tijdens de statistische analyse buiten beschouwing wordt gelaten. Bayesiaanse statistiek kan de reeds bestaande informatie op een verantwoorde manier in de statistische analyse opnemen, wat leidt tot een potentiële afname van de onzekerheid en een toename van de efficiëntie voor de auditor.

Het proefschrift heeft drie doelen: Ten eerste beoogt het Bayesiaanse statistiek in het auditberoep te bevorderen door het actualiseren van de theoretische onderbouwing en het schetsen van de belangrijkste praktische voordelen; ten tweede beoogt het innovatieve en eenvoudig te gebruiken Bayesiaanse statistische methoden te ontwikkelen voor het schatten van parameters en het toetsen van hypothesen, specifiek ontworpen voor steekproeven in een audit; en ten derde beoogt het deze technieken openlijk beschikbaar te maken voor alle auditors via gebruiksvriendelijke open-source statistische software.

Het eerste deel van het proefschrift bespreekt het op een Bayesiaanse manier schatten van parameters bij auditsteekproeven, met de nadruk op het gebruik van reeds bestaande informatie. Hoofdstuk 2 introduceert de basisideeën achter Bayesiaanse parameterschatting en behandelt hoe informatie in de prior-verdeling kan worden opgenomen. Het belangrijkste idee dat in dit hoofdstuk wordt geïntroduceerd is dat de prior-verdeling auditors in staat stelt om statistisch voort te bouwen op reeds bestaande kennis over de fouten in de populatie in hun steekproefprocedures. In dit hoofdstuk bespreek ik hoe het opnemen van reeds bestaande informatie in de prior-verdeling een aantal praktische voordelen heeft voor de auditor, waaronder de mogelijkheid tot een kleinere steekproefomvang en meer transparantie. Daarnaast presenteert het hoofdstuk vijf methoden voor het opstellen van een prior-verdeling in de context van auditsteekproeven. Deze prior-verdelingen worden gemaakt met behulp van bekende controleinformatie, waardoor de auditor ze op een logische manier kan verantwoorden. Hoofdstuk 3 bouwt voort op de fundamentele ideeën van Bayesiaanse parameterschatting uit Hoofdstuk 2 en bespreekt hoe aanvullende data kunnen worden opgenomen in het statistische model. Het belangrijkste idee dat in dit hoofdstuk wordt geïntroduceerd is dat het statistische model auditors in staat stelt hun steekproefprocedures statistisch te baseren op meerdere bronnen van data. In dit hoofdstuk ga ik in op hoe dit twee voordelen heeft voor auditors: Ten eerste helpt het hen een genuanceerder begrip van de populatie te ontwikkelen omdat ze de impact van de geïntegreerde data kunnen verklaren; en ten tweede wordt het voor hen gemakkelijkerer om potentiële fouten in de populatie te identificeren omdat ze scherper onderscheid kunnen maken tussen posten. Het hoofdstuk laat aan de hand van twee voorbeelden zien hoe een Bayesiaanse model-gebaseerde benadering de auditor kan helpen een statistisch model te ontwikkelen dat is afgestemd op de situatie in de praktijk.

Het tweede deel van het proefschrift bespreekt het op een Bayesiaanse manier toetsen van hypothesen, waarbij de nadruk ligt op de kwantificering van statistisch controlebewijs. Hoofdstuk 4 bespreekt het concept van controlebewijs en introduceert Bayesiaans hypothesetoetsen als een methode voor het kwantificeren van dergelijk bewijs uit statistische steekproeven. Het kernidee dat in dit hoofdstuk wordt geïntroduceerd is dat statistisch controlebewijs kan worden gekwantificeerd via de Bayes factor, een maatstaf voor de relatieve bewijskracht van twee getoetste hypothesen. In het hoofdstuk wordt gedemonstreerd dat frequentistische hypothesetoetsing een aantal praktische nadelen heeft die Bayesiaans hypothesetoetsen met gebruikmaking van de Bayes factor kan verhelpen. Het gebruik van de Bayes factor wordt vervolgens geïllustreerd aan de hand van een aantal scenarios uit een moderne audit context. Hoofdstuk 5 ontwikkelt een standaard Bayesiaanse hypothesetoets voor steekproeven in een audit context. De Bayes factor is gevoelig voor de specificatie van de prior-verdeling, maar de invloed van prior-verdeling op de Bayes factor in de context van audit steekproeven is nog niet eerder onderzocht. Helaas kunnen prior-verdelingen die aantrekkelijk zijn om te gebruiken in de context van een audit Bayes factoren opleveren die bewijs kwantificeren in de tegenovergestelde richting van wat de data aangeven. Om dit probleem aan te pakken, wordt in dit hoofdstuk een onpartijdige prior-verdeling voorgestelt. De resulterende Bayes factor is consistent, dat wil zeggen, deze kwantificeert altijd bewijs in het voordeel van de hypothese die het best door de data wordt ondersteund. Het belangrijkste idee in dit hoofdstuk is dat de onpartijdige Bayesiaanse hypothesetoets geschikt is in een verscheidenheid van omstandigheden en bovendien eenvoudig te gebruiken en te rechtvaardigen is.

Het derde deel van dit proefschrift bespreekt een software implementatie van het Bayesiaanse statistische raamwerk dat in de eerdere hoofdstukken wordt bepleit. Het verbeteren van de toegankelijkheid van Bayesiaanse statistische technieken voor auditors was één van de doelen van dit project. Daarom is JASP for Audit ontwikkeld als een add-on module voor de bestaande software JASP, een open-source statistisch programma met een grafische gebruikersinterface. Hoofdstuk 6 geeft een overzicht van JASP for Audit, dat zowel frequentistische als Bayesiaanse statistische technieken biedt voor audit steekproeven met als doel de auditor te helpen bij de statistische aspecten van een audit. Dit wordt bereikt door, naast andere analyses, een begeleidende workflow te bieden die de bekende vier stappen uit het steekproefproces doorloopt, de noodzakelijke statistische berekeningen uitvoert, en automatisch een controlerapport produceert met de statistische resultaten en de interpretatie van deze resultaten. Dit hoofdstuk bespreekt de voordelen van JASP for Audit voor de audit praktijk en biedt een grondige walkthrough van drie voorbeelden. Verder geeft het aanbevelingen voor het gebruik van JASP for Audit in de praktijk.

Kortom, dit proefschrift beoogt een innovatieve benadering van statistische auditsteekproeven te ontwikkelen. Door een volledig Bayesiaans raamwerk voor parameterschatting en hypothesetoetsing te bieden en deze technieken vrij beschikbaar te stellen, zorgt dit proefschrift ervoor dat auditors te allen tijde de vruchten kunnen plukken van Bayesiaanse statistiek.

Acknowledgements — Dankwoord

This thesis could not have been completed without the help and support of a large number of people. First and foremost, I would like to thank my supervisor, Ruud Wetzels, and my co-supervisors Jacques de Swart and Eric–Jan Wagenmakers. Additionally, I want to thank Stephen Asare, Joost van Buuren, Michael Lee, and Barbara Majoor who made up my defense committee.

Ruud, thank you for all of your time and effort in guiding me, for all of the fruitful discussions and writing sessions we have had together, and for all of the shots of home-brewed liquor. Whenever we had spent time together, I always went home with fresh ideas and insights (which can only partially be attributed to the drinks). Without your advice on how to organize my thoughts and writing, this thesis would have become a complete mess.

Thank you, Jacques, for always being willing to put in such a tremendous amount of effort into this project, and for wanting to show me the practical side of the auditing profession. Your enthusiasm for JASP is contagious and has frequently given me the drive to keep working on this project. Without your real-world knowledge and your willingness to play devil's advocate, I am sure this thesis would have become unrelatable to many readers.

Eric–Jan, thank you for always being the Bayesian voice of reason in my study time and during this project. Your extensive knowledge of Bayesian statistics, critical perspective and sharp writing suggestions have played a large role in making this thesis what it is. I also want to thank you for initially giving me the opportunity to work with JASP and be a part of a bigger team. I cannot emphasize enough how valuable the friendships and experiences I gained in your lab are to me, both as a researcher and as a person.

I want to express my gratitude to my colleagues at the Center of Accounting, Auditing & Control for being at my side for the last four years, in spite of the challenges posed by social constraints. I appreciate Remko's constant optimism and desire to involve me in a variety of fascinating side-content. Herman, thank you for your keen interest in and openness to learning about our research into Bayesian statistics. Thank you, René, for occasionally lending me a sympathetic ear. Without Barbara's and Cokky's outstanding leadership, our center obviously would not exist. Joost, Reinalt, Niels, Tim, Bob, Ivo, Jaap, Hans, Lineke, Wilbert, Roland, Inez and Marlies, thank you for the interesting hallway conversations and center meetings. Though our interactions over the last four years were often brief, they were always appreciated.

I also had the opportunity to join the statistical auditing steering committee at the Limperg Institute, which was instrumental in the development of JASP for Audit. These gatherings would not have been possible without Jacques' excellent chairmanship. Ferry, I want to thank you for delving so deeply into JASP for Audit, for all of your feedback on JASP itself, and for making all those notes after the meetings. Paul, for your commitment to helping with our first article, for your feedback on the early versions of JASP for Audit, and for your unwavering determination to make Bayesian inference accessible to audit practitioners. Thank you, Crist–Jan, for your enthusiasm for Bayesian statistics, your commitment to promoting open-source software at work, including JASP for Audit, and your relatable love of programming in R. Niels, Kees, Wouter, Marc, Angelique, Chantal, Peter, Rabin, Kees, Willem, Harry, and Hein, you were all a great source of inspiration and constructive feedback throughout this project.

I want to express my gratitude to the JASP team and the Psychological Methods department at the University of Amsterdam for accepting me into their group and giving me a second place to work from. I had the good fortune to spend a good deal of my time here, which has been very rewarding for me socially and professionally. Alexander, I want to thank you for taking me under your wing and giving me the opportunity to express myself from the beginning, even though I did not have much programming experience. Thank you Bruno, Joris, Frans, and Tim, for teaching me the ropes, and for always helping me out when the Windows version of JASP would not build. Don, thank you for being by my side throughout many JASP projects and for your extensive knowledge of programming, which has helped me out more than a few times. Šimon, thank you for always wanting to review my pull requests despite the fact that they had very brief descriptions. I can only say that I have developed a strong affection for the entire JASP team and am proud of everything we have accomplished together.

My family has always been a huge source of support for me, but especially over the last four years. Thank you, mom and dad, for always being there for me and letting me follow the path I want to. Laura, thank you for being such a motivated learner. You have always been a shining example of sincerity, eagerness, and hard work to me. During this time, I unfortunately got to spend less time with friends like Koen, Dion, Rickert, Rens, Frank, Zenab, Gaby, Ria, Leonie, Sofieke, Jessica, Femke, Casper, and Anfernee than I had hoped, but they were always a welcome distraction. Thank you for the fun times we had in the bar, the shared laughs and holidays, and for being a good friend to me. Thank you Jason, Tessel, Iris, and Maike, for the times we spend drinking beers at CREA on a Friday afternoon. Most importantly, I want to thank my dearest Sophie for always being there for me, no matter if it was in a good or a bad moment. You have truly been a source of love, encouragement, and inspiration for me during the last four years.

I want to express my gratitude to everyone who has given input on the articles that make up this thesis in any way. Without your suggestions, expertise, and advice, I would not have been able to complete this project. Naturally, any remaining mistakes in this thesis are entirely my own.

Publications

1. Wagenmakers, E.–J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.–J., van Doorn, J., Smira, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., **Derks, K.**, Rouder, J. N., & Morey, R. D. (2017). Bayesian inference for psychology. part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. doi: `https://doi.org/10.3758/s13423-017-1323-7`

2. **Derks, K**., Burger, J., van Doorn, J., Kossakowski, J. J., Matzke, D., Atticciati, L., Beitner, J., Benzesin, V., de Bruijn, A. L., Cohen, T. R. H., Cordesius, E. P. A., van Dekken, M., Delvendahl, N., Dobbelaar, S., Groenendijk, E. R., Hermans, M. E., Hiekkaranta, A. P., Hoekstra, R. H. A., Hoffmann, A. M., Hogenboom, S. A. M., Kahveci, S., Karaban, I. J., Kevenaar, S. T., te Koppele, J. L., wil Kramer, A., Kroon, E., Kucharský, Š., Lieuw–On, R., Lunansky, G., Matzen, T. P., Meijer, A., Nieper, A., de Nooij, L., Poelstra, L., van der Putten, W. J., Sarafoglou, A., Schaaf, J. V., van de Schraaf, S. A. J., van Schuppen, S., Schutte, M. H. M., Seibold, M., Slagter, S. K., Snoek, A. C., Stracke, S., Tamimy, Z., Timmers, B., Tran, H., Uduwa–Vidanalage, E. S., Vergeer, L., Vossoughi, L., Yücel, D. E., & Wagenmakers, E.–J. (2018). Network models to organize a dispersed literature: The case of misunderstanding analysis of covariance. *Journal of European Psychology Students*, *9*(1), 48–57. doi: `http://doi.org/10.5334/jeps.458`

3. van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., **Derks, K.**, Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.–J. (2020). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*. doi: `https://doi.org/10.3758/s13423-020-01798-5`

4. Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., van

den Bergh, D., Marsman, M., **Derks, K.**, Wagenmakers, E.–J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., Cheung, F., Cimpian, A., Dohle, S., Donnellan, M. B., Hahn, A., Hall, M. P., Jiménez–Leal, W., Johnson, D. J., Lucas, R. E., Monin, B., Montealegre, A., Mullen, E., Pang, J., Ray, J., Reinero, D. A., Reynolds, J., Sowden,W., Storage, D., Su, R., Tworek, C. M., Bavel, J. J. V.,Walco, D.,Wills, J., Xu, X., Yam, K. C., Yang, X., Cunningham,W. A., Schweinsberg, M., Urwitz, M., Collaboration, T. C. H. T., & Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451–479. doi: `https://doi.org/10.1037/bul0000220`

5. van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.–J., **Derks, K.**, Dablander, F., Gronau, Q. F., Kucharský, S., Gupta, A. R. K. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.–J. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année psychologique*, *120*(1), 73–96. doi: `https://doi.org/10.3917/anpsy1.201.0073`

6. Ly, A., Stefan, A., van Doorn, J., Dablander, F., van den Bergh, D., Sarafoglou, A., Kucharský, S., **Derks, K.**, Gronau, Q. F., Raj, A., Boehm, U., van Kesteren, E.–J., Hinne, M., Matzke, D., Marsman, M., & Wagenmakers, E.–J. (2020). The Bayesian methodology of Sir Harold Jeffreys as a practical alternative to the p value hypothesis test. *Computational Brain & Behavior*, *3*(2), 153–161. doi: `https://doi.org/10.1007/s42113-019-00070-x`

7. van Doorn, J., van den Bergh, D., Dablander, F., van Dongen, N. N. N., **Derks, K.**, Evans, N. J., Gronau, Q. F., Haaf, J. M., Kunisato, Y., Ly, A., Marsman, M., Sarafoglou, A., Stefan, A., & Wagenmakers, E.–J. (2021). Strong public claims may not reflect researchers' private convictions. *Significance*, *18*(1), 44–45. doi: `https://doi.org/10.1111/1740-9713.01493`

8. **Derks, K.**, de Swart, J., van Batenburg, P., Wagenmakers, E.–J., & Wetzels, R. (2021). Priors in a Bayesian audit: How integration of existing information into the prior distribution can improve audit transparency and efficiency. *International Journal of Auditing*, *25*(3), 621–636. doi: `https://doi.org/10.1111/ijau.12240`

9. **Derks, K.**, de Swart, J., Wagenmakers, E.–J., Wille, J., & Wetzels, R. (2021). JASP for Audit: Bayesian tools for the auditing practice. *Journal of Open Source Software*, *6*(68), 2733. doi: `https://doi.org/10.21105/joss.02733`

10. **Derks, K.**, de Swart, J., & Wetzels, R. (2022). Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor. *Maandblad voor Accountancy en Bedrijfseconomie*, *96*(1/2), 37–46. doi: `https://doi.org/10.5117/mab.96.78836`

11. Heck, D. W., Boehm, U., Böing–Messing, F., Bürkner, P.–C., **Derks, K.**, Dienes, Z., Fu, Q., Gu, X., Karimova, D., Kiers, H., Klugkist, I.,

Kuiper, R. M., Lee, M. D., Leenders, R., Leplaa, H. J., Linde, M., Ly, A., Meijerink–Bosman, M., Moerbeek, M., Mulder, J., Palfi, B., Schönbrodt, F., Tendeiro, J., van den Bergh, D., van Lissa, C. J., van Ravenzwaaij, D., Vanpaemel, W., Wagenmakers, E.-J., Williams, D. R., Zondervan–Zwijnenburg, M., & Hoijtink, H. (2022). A review of applications of the Bayes factor in psychological research. *Psychological Methods.* doi: https://doi.org/10.1037/met0000454

12. **Derks, K.**, de Swart, J., Wagenmakers, E.–J., & Wetzels, R. (2022a). The Bayesian approach to audit evidence: Quantifying statistical evidence using the Bayes factor. Manuscript under review for publication. doi: https://doi.org/10.31234/osf.io/kzqp5

13. **Derks, K.**, de Swart, J., Wagenmakers, E.–J., & Wetzels, R. (2022b). An impartial Bayesian hypothesis test for audit sampling. Manuscript under review for publication. doi: https://doi.org/10.31234/osf.io/8nf3e

14. **Derks, K.**, de Swart, J., Wagenmakers, E.–J., & Wetzels, R. (2022c). How to incorporate audit information into the statistical model: Bayesian generalized linear modeling for audit sampling. Manuscript submitted for publication. doi: https://doi.org/10.31234/osf.io/byj2a

15. **Derks, K.**, de Swart, J., & Wetzels, R. (2022). Open-source software als brug tussen de auditor en de statisticus. Manuscript under review for publication.

16. JASP Team. (2022). JASP (Version 0.16.4)[Computer software]. https://jasp-stats.org/

17. **Derks, K.** (2022). jfa: Bayesian and Classical Audit Sampling. R package version 0.6.4. https://cran.r-project.org/package=jfa

BAYESIAN BENEFITS FOR AUDITING:

A Proposal to Innovate Audit Methodology

In this thesis, I make the case that auditors can benefit greatly by adopting Bayesian statistics into their methodological toolbox. Specifically, I investigate and discuss the practical benefits of Bayesian inference in the context of modern auditing practices, I develop innovative Bayesian statistical methods that are specifically tailored to audit sampling and I make these techniques available to all audit practitioners via the open-source software program JASP (**https://jasp-stats.org**).

NYENRODE
BUSINESS UNIVERSITEIT

ISBN 978-90-8980-982-7