On a Bayesian sample size determination problem with applications to auditing

Sujit K. Sahu and T. M. Fred Smith

Abstract

The problem motivating this article is the determination of sample size at the substantive testing stage of a financial audit. An error in an audit account is said to occur when there is a non-zero difference between its book value and true value. A typical financial auditing task involves several stages and usually a large number of potential items are available for testing at each stage. The sample size determination problem is to find an optimum fixed number of items which must be tested so that the quantitative risk of a wrong decision is bounded by a pre-specified quantity. Senior auditors often have strong subjective opinions regarding the state of the accounts being audited which naturally leads to the choice of Bayesian methods. Solutions are proposed under various model assumptions. A combination of analytical and simulation based techniques is proposed and some theoretical results are obtained. The methods developed, however, are quite general and can be applied to other sample size determination (SSD) problems. A number of numerical illustrations are given.

S^{3}RI Methodology Working Paper M04/04

On a Bayesian sample size determination problem with applications to auditing

S. K. Sahu and T. M. F. Smith†
University of Southampton, UK

Summary. The problem motivating this article is the determination of sample size at the substantive testing stage of a financial audit. An error in an audit account is said to occur when there is a non-zero difference between its book value and true value. A typical financial auditing task involves several stages and usually a large number of potential items are available for testing at each stage. The sample size determination problem is to find an optimum fixed number of items which must be tested so that the quantitative risk of a wrong decision is bounded by a pre-specified quantity. Senior auditors often have strong subjective opinions regarding the state of the accounts being audited which naturally leads to the choice of Bayesian methods. Solutions are proposed under various model assumptions. A combination of analytical and simulation based techniques is proposed and some theoretical results are obtained. The methods developed, however, are quite general and can be applied to other sample size determination (SSD) problems. A number of numerical illustrations are given.

Keywords: Auditing; Bayesian inference; book values; fitting prior; mixture distribution; rare errors; simulation based approach; sampling prior; taints.

1. Introduction

The problem motivating this article is sample size determination (SSD) at the substantive testing stage of a financial audit. Financial auditing involves several stages. At the first stage senior auditors review the system generating the accounts and compare the current results with those of previous years and with those of similar entities. This is often done together with the senior management of the entity being audited. The objective of system review is to identify any changes from previous years, any weaknesses in the system and any other areas of possible concern. In the light of this review a strategy for more detailed explorations and tests is developed. The next stage is to test the working of the accounting system and, in particular, the implementation of controls and checks. This phase is known as compliance testing and may exceptionally be done using a computer generated set of transactions, running them through the system and checking for compliance. The substantive testing of actual transactions follows. This procedure is usually carried out by junior auditors who identify the items selected into the sample and then check that the money values are correct. This is an essential part of the audit process, and of the training of auditors, but it is often described as being of little value since errors in money values are rarely found in samples selected from well-designed accounting systems.

The overall objective is to test whether the monetary totals recorded in the accounts are correct or not. It should be noted that the substantive testing of items tests only the

†*Address for correspondence:* Sujit K. Sahu, School of Mathematics, S^3RI, University of Southampton, Southampton, SO17 1BJ, United Kingdom
E-mail: S.K.Sahu@maths.soton.ac.uk

accuracy of the totals generated by the system as specified at the first stage. This is not a procedure designed to discover large faults in the design of the system that may have led to recent accounting scandals such as those at Enron and Parmalat. Discovering these system faults is the responsibility of senior auditors at the system review stage.

It is now being argued by some auditors that the sampling of actual transactions and the checking of money values is a waste of time because the information from the earlier stages, especially the comparisons of totals with earlier years and similar entities, a process called analytical review, and the compliance checks of the system, provides sufficient evidence to reach a conclusion about the accuracy of the accounts. However, accounting standards require that some form of substantive testing be carried out, but this may be interpreted broadly to embrace analytical review as well as sampling, and it is believed by many auditors that analytical review may in itself provide adequate information about the system. This conclusion has been challenged by Swan (2001) on the grounds that analytical review has very low power for detecting errors. A counter argument is that in order to be able to defend an audit publicly some actual items should always be tested, but the sample size need not be large since all that is required is confirmation that the accounting process, as evaluated at the earlier stages, is working well. In this view sampling has a quality assurance role within an audit rather than an estimation role. Sampling can be seen to be a protective procedure for auditors and the absence of sampling can be viewed as an invitation to abuse the audit process. The problems identified above are clearly statistical in nature. They relate to the quality assurance of a complex system and to the design and analysis of studies aimed at achieving high levels of assurance.

There is a considerable literature, see e.g. Laws and O'Hagan (2000) and the references therein, on the analysis of audit data. The information from the early stages of an audit is qualitative and often leads to strong opinions about the quality of the system. Combining this prior information with the hard data generated by sampling at the substantive stage may be done in an ad-hoc manner within the frequentist tradition (see e.g. Patterson, 1993 and Shrivastava and Shafer, 1994) or more formally using Bayes' theorem. An important reference is Cox and Snell (1979) who propose a Bayesian mixture model for the analysis of substantive data. See also Laws and O'Hagan (2002). The practical problem is that if money errors are rare then the number of errors found in small or medium sized samples will be very small, and possibly zero. Thus the effective sample size for frequentist inference about the total of money errors is small and the resulting inferences will be unreliable. Using the available prior information within a Bayesian methodology may lead to more reliable conclusions about the unknown error totals.

Despite the proposed use of Bayesian methods in the analysis of auditing data the SSD for substantive testing has mainly been based on frequentist ideas, although there are exceptions, see e.g. van Batenburg *et al.* (1994). The usual approach is to assume that the sample will yield no money errors and to determine the sample size to guarantee that if no errors are found then with, say, 95% confidence the proportion of items in error in the population will not exceed a given percentage. Assuming a Poisson distribution then a sample of size 100 gives 95% confidence that the population error rate is less than 3% if no errors are found. This is a large sample in the audit context and most auditors would refute this conclusion saying that they knew almost certainly that the error rate was less than 3% before the sample was selected. For this reason various ad-hoc methods have been proposed to employ the prior information to reduce the required sample size within the frequentist framework. Another problem is that if any errors are found in the sample then the upper confidence limit for the error rate increases dramatically. The usual response from auditors

is to argue that the errors found in the sample are unique, and can be ring-fenced and removed from the population, and that in the remaining part of the sample there were no errors and inferences about the rest of the population can be made accordingly. We find none of the frequentist arguments for determining sample size to be satisfactory in either theory or practice and so we explore some Bayesian methods for SSD, which would force auditors to be explicit about their assumptions prior to sampling.

Before setting up the model framework within which the statistical decisions will be made we need to refine further some aspects of the audit process. The final accounts about which a decision will be made comprise a set of sub-accounts such as, income (possibly by category) and expenditure on specific functions, for example payroll, or on products that are particular to the audited entity, and on administrative expenditure. Different sub-accounts have different accounting processes, and hence different types of error, and so the audit can be broken down into separate audits for each sub-account. If any sub-account is in serious error then the final audit conclusion will identify this and qualify this section of the accounts. Statistically the audit is stratified and inferences are made within strata as well as overall. Auditors use a concept called material error to define the value of monetary error that would lead them to qualify an account. We assume that the auditor has set the value of material error within each sub-account; typically this will be a percentage of the total money value of the sub-account, say 1% or 2%. Samples will be drawn from within strata and so we concentrate on SSD within each sub-account separately. In the rest of the paper the term account will refer to the sub-account being audited.

For SSD the only information available is prior information. Frequentist rules require the input of point estimates of parameters, such as variances, without any measure of uncertainty, which runs counter to one of the main motivations for the use of statistical methods. Introducing uncertainty into prior estimates is a quintessentially Bayesian procedure and we explore the use of Bayesian methods within various distributional frameworks which may be relevant for auditing. Ideally samples should be evaluated sequentially, but this is not possible within the audit context where the junior auditors carry out the sampling. We assume that the objective is to determine an optimal fixed size sample that satisfies a criterion based on the Bayes' risk. We further assume that the auditor can specify the implicit losses of their decision process and possibly also the cost of sampling measured in the same units as the losses. Given the loss functions and the sampling cost function it is possible to carry out a fully Bayesian determination of sample size, see Lindley (1997) and the references therein. In the absence of information about sampling costs we adopt a partial Bayes approach in the spirit of Adcock (1997), Joseph *et al.* (1995), and Wang and Gelfand (2002). We adopt the framework proposed in Wang and Gelfand where two different prior distributions are used for SSD problems. The prior for inference, the fitting prior, can differ from the prior used for averaging to calculate the Bayes risk, the sampling prior. We explore some of the consequences of using different fitting and sampling priors within our chosen model framework.

The plan of the remainder of this article is as follows. In Section 2 we develop the general methodology. Section 3 discusses the results for the normal error distribution. In Section 4 we consider the exponential distribution for the errors and in Section 5 we discuss a mixture distribution for the errors in an account. The article ends with a few summary remarks in Section 6.

2. Method

2.1. Error distribution

In financial audits the recorded value of a transaction is often called the *book value* which can be matched to a true value called the *audit value*. The error in a transaction is defined as the difference, $X_i = B_i - A_i$, between its book value, B_i, and audit value, A_i. In some accounts only overstatement errors can occur in which case $0 < A_i < B_i$. For modelling purposes, however, we can model either the difference itself or the absolute difference and we develop models for both of these scenarios.

Let θ denote the mean of the error random variable X. Thus the parameter θ is the mean error rate per item in the account. The primary objective of the auditors is to draw inference on the parameter θ, or equivalently $N\theta$, where N is the number of items in the account. In the subsequent sections we parameterise the distribution of the random variable X using θ and assume that $X|\theta$ follows the distribution $f(x|\theta)$. We shall work with three different possibilities for the density function f.

Usually in Bayesian modelling a choice has to be made between non-parametric and parametric modelling approaches. For the general SSD problem non-parametric models using Dirichlet processes can be considered, see e.g. Walker (2003). Here we work with three different tailor-made models each of which corresponds to a different practical auditing problem. Moreover, the parametric models can also be used for SSD problems which arise in other application areas, such as medicine and clinical trials.

For many SSD problems the Gaussian assumption is used because of its simplicity and tractability. Even when using a non-parametric model the central limit theorem may be used to approximate some key probabilities required for the SSD problem, see e.g. Walker (2003), Clarke and Yuan (2002) and Section 2.2 for more in this regard. In Section 3 we assume that each X_i is normally distributed. As is expected, this turns out to be an analytically tractable situation where our methods provide some exact solutions, though the final sample size needs to be calculated using computer intensive methods.

In Section 4 we model the absolute difference, i.e. we set $X_i = |B_i - A_i|$, because sometimes auditors are interested in the absolute errors only. In such situations there are often many small errors and a few large errors. For SSD we assume that the X_i's are exponentially distributed. Note that this is also a possible model for the positive overstatement errors where a skewed error distribution is appropriate for modelling purposes.

A drawback of both the above models is that they do not allow the particular value zero to have a positive probability, and in typical auditing problems a large number of items may give either no items in error or perhaps very few items in error. Thus in modelling we need to allow the case for $|X_i| = 0$ with non-zero probability. In Section 5 we entertain this possibility for SSD, and employ a mixture model using the approach introduced by Cox and Snell (1979). The mixture model is a mixture of an exponential distribution for the non-zero errors and a point mass at $|X_i| = 0$. In their paper Cox and Snell model the proportional overstatement errors or taints, $|X_i|/B_i$, using the above mixture and ignore the important fact that the taints are between 0 and 1. We, however, do not ignore this fact and model the actual error values $|X_i|$ so that the mixture distribution is supported on the non-negative half of the real line.

2.2. The hypothesis testing problem

Let $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)$ denote a random sample of size n from a population with density $f(x|\theta)$ and let $\pi(\theta)$ denote the prior distribution for the unknown parameter θ. Let $\pi(\theta|\mathbf{x}^{(n)})$ denote the posterior distribution of θ given the observed sample $\mathbf{x}^{(n)}$.

We follow the development in Berger (1985, Chapter 7) to set up the hypothesis testing problem which is to choose between the two hypotheses:

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1,$$

where Θ_0 is less than Θ_1 in the sense that, if $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ then $\theta_0 < \theta_1$. In this article we shall take $\Theta_0 = \{\theta : -\infty < \theta \leq \theta_0\}$ and $\Theta_1 = \{\theta : \theta_0 < \theta < \infty\}$.

In the auditing context θ_0 represents a value corresponding to a material error per item. If $\theta < \theta_0$ the error is not material and the account will be accepted. If, however, $\theta \geq \theta_0$ the error is material and the account will be rejected and the auditors will qualify that section of the accounts in their conclusions. Note that θ_0 *is a positive quantity* set in advance by the auditors.

Let a_i denote the action of accepting H_i for $i = 0, 1$ and $L(\theta, a_i)$ denote the loss for taking decision a_i when θ is the true value. The Bayes decision rule, denoted by δ_n^π, is to select a_0 if

$$\int_{\Theta_1} L(\theta, a_0)\pi(\theta|\mathbf{x}^{(n)})d\theta < \int_{\Theta_0} L(\theta, a_1)\pi(\theta|\mathbf{x}^{(n)})d\theta. \tag{1}$$

Under a parametric assumption it is often possible to find a suitable function $g\left(\mathbf{x}^{(n)}\right)$ such that (1) holds if and only if $g\left(\mathbf{x}^{(n)}\right) < k^\pi(n)$ where $k^\pi(n)$ is the value of $g\left(\mathbf{x}^{(n)}\right)$ for which equality holds in (1) instead of the inequality. In the parametric family $f(x|\theta)$, if $\bar{X}_n$ is sufficient for θ then Berger (1985) establishes that $g\left(\mathbf{x}^{(n)}\right) = \bar{x}_n$. This will be the case for our normal and exponential error distributions in Sections 3 and 4. However, this simplification is not possible for our mixture error distribution and in Section 5 we work with the appropriate g function.

The Bayes decision risk, denoted by $r\left(\pi, \delta_n^\pi\right)$, is given by

$$\begin{aligned} r(\pi, \delta_n^\pi) &= \int_{\Theta_1} L(\theta, a_0) P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) < k^\pi(n) \right\} \pi(\theta) d\theta \\ &\quad + \int_{\Theta_0} L(\theta, a_1) P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) \geq k^\pi(n) \right\} \pi(\theta) d\theta. \end{aligned} \tag{2}$$

For SSD we may also define a cost function, $c(n)$ say, for obtaining the samples. In general SSD problems the cost function $c(n)$ is often chosen to be an increasing linear function of n while the risk decreases with n. The SSD problem is to minimise

$$r(\pi, \delta_n^\pi) + c(n)$$

over the values of the sample size, n. The smallest n which minimises the above is the required sample size.

Note that the parametric assumption enters the sample size calculation through the probability $P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) \geq k^\pi(n) \right\}$. The non-parametric approach of Walker (2003) approximates this probability using the central limit theorem. Therefore, the large optimum sample sizes for the Gaussian model will be similar to the ones obtained from an equivalent non-parametric approach.

2.3. Fitting and sampling priors

All Bayesian model fitting exercises need a prior distribution for the unknown parameters in the model. This is the prior distribution which would have been used for model fitting if the data were available. Following Wang and Gelfand (2002) we call this the fitting prior and denote it by $\pi^{(f)}(\theta)$. Often, $\pi^{(f)}(\theta)$ is assumed to be vague (or non-informative) so that the modeller encourages the data to drive the inference, thus it is a general purpose working prior distribution.

The fitting prior is to be used to obtain the posterior distribution $\pi(\theta|\mathbf{x}^{(n)})$ and to emphasise this dependence we write the posterior distribution as $\pi^{(f)}(\theta|\mathbf{x}^{(n)})$. Thus the decision rule is denoted by $\delta_n^{\pi^{(f)}}$ and it selects a_0 if (1) holds for the posterior distribution $\pi^{(f)}(\theta|\mathbf{x}^{(n)})$. The quantity $k^{\pi}(n)$ will also depend on the fitting prior used to calculate the posterior distribution and we emphasise this dependence by writing $k^{\pi^{(f)}}(n)$.

In the frequentist approach to the SSD problems it is usually of interest to investigate the sensitivity of the SSD procedure when the 'true' parameter θ assumes some particular values. This is not considered to be satisfactory from a Bayesian perspective where the unknown parameter θ is assumed to be random. To perform sensitivity analysis in a coherent Bayesian framework it is natural to assume that the parameter θ follows an informative prior distribution concentrated around some specific values of θ which are of particular interest to the practitioner. This is the prior that a pure Bayesian would employ after full consideration of all the available prior information. Wang and Gelfand (2002) formalised this concept by calling this informative prior distribution the sampling prior. Here this prior is denoted by $\pi^{(s)}(\theta)$ and it replaces the familiar assumption of fixing θ in the classical SSD problem.

What are the differences between the fitting and sampling priors and why should they not be the same? The sampling prior is the prior distribution used to generate the parameter values which are then conditioned upon to generate the data from $f(x|\theta)$ in substantive experiments. That is, data $\mathbf{X}^{(n)}$ are generated from the joint hierarchical model $\pi^{(s)}(\theta)f(x|\theta)$. Once data are available we would like to pretend that the informative prior distribution which generated the data is unknown to us; and we would like to make inference with the assumption of a relatively non-informative prior distribution. The sampling and fitting prior distributions should not be the same because they serve two different purposes in the SSD problems. The sampling prior distribution addresses the 'what if' type sensitivity scenarios, whereas the fitting prior distribution is used to form the posterior distribution for making inference. In our numerical illustrations we will investigate the situation where the sampling prior is the same as the fitting prior, the conventional Bayesian approach, and also explore the effect of different sampling and fitting priors.

The distinction between the sampling and fitting prior distributions will naturally affect the calculation of the Bayes risk, $r(\pi, \delta_n^{\pi})$ given in (2). As mentioned above the decision rule δ_n^{π} will need to be written as $\delta_n^{\pi^{(f)}}$. The prior distribution $\pi(\theta)$, used as the averaging measure in the integrals of (2), will be the sampling prior distribution $\pi^{(s)}(\theta)$. Thus the Bayes risk (2) will have the following form:

$$\begin{aligned} r\left(\pi^{(s)}, \delta_n^{\pi^{(f)}}\right) &= \int_{\Theta_1} L(\theta, a_0) P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) < k^{\pi^{(f)}}(n) \right\} \pi^{(s)}(\theta) d\theta \\ &\quad + \int_{\Theta_0} L(\theta, a_1) P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) \geq k^{\pi^{(f)}}(n) \right\} \pi^{(s)}(\theta) d\theta. \end{aligned} \tag{3}$$

2.4. Specific losses and bounding the risk

The methodology outlined in the above sections is quite general in nature. As a result it can be specialised to particular problems such as the auditing problem motivating this article. Here we shall make several assumptions which are particular to the auditing problems. Auditors have some difficulty in specifying a particular cost function, $c(n)$, although there is a general consensus that the cost function is an increasing function of n. In view of this we reformulate the SSD problem as that of bounding the risk $r\left(\pi^{(s)}, \delta_n^{\pi^{(f)}}\right)$ by a pre-specified quantity. We then find the minimum sample size which bounds the risk function and this takes care of the increasing nature of the cost function. The actual upper-bound on the risk function is related to the loss functions which we describe below.

From practical considerations in the audit problem we assume the simplest constant loss functions $L(\theta, a_0) = L_0$ for $\theta > \theta_0$ and $L(\theta, a_1) = L_1$ for $\theta \leq \theta_0$. The loss function is assumed to be zero if a correct decision is made. Auditors are reluctant to specify absolute values of L_0 and L_1 for a general problem. However, they feel more comfortable in specifying the ratio L_0/L_1 or equivalently

$$\eta = \frac{L_0}{L_0 + L_1}.$$

Henceforth, we shall work with this ratio wherever possible. Now we have the following simpler form of the risk function (3)

$$\begin{aligned} r\left(\pi^{(s)}, \delta_n^{\pi^{(f)}}\right) &= L_0 \int_{\Theta_1} P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) < k^{\pi^{(f)}}(n) \right\} \pi^{(s)}(\theta) d\theta \\ &\quad + L_1 \int_{\Theta_0} P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) \geq k^{\pi^{(f)}}(n) \right\} \pi^{(s)}(\theta) d\theta \\ &= L_0 \left[\int_{\Theta_1} P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) < k^{\pi^{(f)}}(n) \right\} \pi^{(s)}(\theta) d\theta \right. \\ &\quad \left. + \tfrac{1-\eta}{\eta} \int_{\Theta_0} P_\theta \left\{ g\left(\mathbf{X}^{(n)}\right) \geq k^{\pi^{(f)}}(n) \right\} \pi^{(s)}(\theta) d\theta \right]. \end{aligned}$$

The above risk function is a multiple of the loss L_0 and it depends on the ratio of the losses η. In the absence of the absolute values of the losses we re-formulate the SSD problem as one of finding the minimum n such that

$$\frac{r\left(\pi^{(s)}, \delta_n^{\pi^{(f)}}\right)}{L_0} \leq M(\eta)$$

for given values of η and $M(\eta)$. Note that this is a canonical version of the SSD problem which bounds the risk by $L_0 M(\eta)$. Also under the assumption that $L_0 = L_1$, i.e. the losses are equal for the two possible wrong decisions, we see that the quantity to be bounded for the SSD is the sum of two error probabilities, which is an appealing quantity to bound for practical problems. In our numerical illustrations we shall experiment with three values of $M(\eta)$, viz. 0.10, 0.05 and 0.01. The last one implies a very strict condition on the two error probabilities and then we shall see that many sample sizes will be very large. We set the optimum sample size to be ∞ if it is greater than 5000.

3. Normal error distribution

Suppose that $X|\theta \sim N(\theta, \sigma^2)$ where σ^2 is known and assume $\pi^{(f)}(\theta) = N(\mu_f, \tau_f^2)$ and $\pi^{(s)}(\theta) = N(\mu_s, \tau_s^2)$. All hyper-parameters are assumed to be known. The posterior distri-

bution of θ is normal with mean

$$E(\theta|\bar{x}_n) = \lambda_f^2\left(\frac{n\bar{x}_n}{\sigma^2} + \frac{\mu_f}{\tau_f^2}\right), \text{ and } \text{var}(\theta|\bar{x}_n) = \lambda_f^2$$

where $\lambda_f^2 = 1/\left(\frac{n}{\sigma^2} + \frac{1}{\tau_f^2}\right)$. We now derive $k^{\pi^{(f)}}(n)$. The Bayes rule chooses action a_0 if

$$\begin{array}{lcl} L_0\int_{\theta_0}^{\infty}\pi^{(f)}(\theta|\bar{x}_n)d\theta & < & L_1\int_{-\infty}^{\theta_0}\pi^{(f)}(\theta|\bar{x}_n)d\theta \\ i.e. \quad L_0(1-p) & < & L_1 p, \text{ say,} \\ \Longrightarrow \quad p & > & \frac{L_0}{L_0+L_1} \equiv \eta, \end{array}$$

where

$$p = \int_{-\infty}^{\theta_0}\pi^{(f)}(\theta|\bar{x})d\theta = \Phi\left(\frac{\theta_0 - \lambda_f^2(\frac{n\bar{x}_n}{\sigma^2} + \frac{\mu_f}{\tau_f^2})}{\lambda_f}\right).$$

Let $q = \Phi^{-1}(\eta)$ where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution. Now it is clear that $p > \eta$ if

$$\bar{x}_n < k^{\pi^{(f)}}(n) = \frac{\sigma^2}{n}\left\{\frac{\theta_0 - q\lambda_f}{\lambda_f^2} - \frac{\mu_f}{\tau_f^2}\right\}.$$

We now have

$$Pr(\bar{X}_n < k^{\pi^{(f)}}(n)|\theta) = \Phi\left(\frac{k^{\pi^{(f)}}(n) - \theta}{\sigma/\sqrt{n}}\right).$$

Let $\phi(\cdot)$ be the density function of the standard normal random variable. The following calculations reduce the risk function to an analytic form. The risk is given by

$$\begin{array}{rcl} r(\pi^{(s)}, \delta_n^{\pi^{(f)}}) & = & L_0\int_{\theta_0}^{\infty}\Phi\left(\frac{k^{\pi^{(f)}}(n)-\theta}{\sigma/\sqrt{n}}\right)\frac{1}{\tau_s\sqrt{2\pi}}e^{-\frac{1}{2\tau_s^2}(\theta-\mu_s)^2}d\theta \\ & & +L_1\int_{-\infty}^{\theta_0}\left\{1-\Phi\left(\frac{k^{\pi^{(f)}}(n)-\theta}{\sigma/\sqrt{n}}\right)\right\}\frac{1}{\tau_s\sqrt{2\pi}}e^{-\frac{1}{2\tau_s^2}(\theta-\mu_s)^2}d\theta, \\ & = & L_0\int_{\frac{\theta_0-\mu_s}{\tau_s}}^{\infty}\Phi\left(\frac{k^{\pi^{(f)}}(n)-\mu_s-\tau_s u}{\sigma/\sqrt{n}}\right)\phi(u)du \\ & & +L_1\int_{-\infty}^{\frac{\theta_0-\mu_s}{\tau_s}}\left\{1-\Phi\left(\frac{k^{\pi^{(f)}}(n)-\mu_s-\tau_s u}{\sigma/\sqrt{n}}\right)\right\}\phi(u)du \\ & = & L_0 Pr(U^* < -a, V^* < b) + L_1 Pr\left(U^* < a, V^* < -b\right) \end{array}$$

where

$$a = \frac{\theta_0 - \mu_s}{\tau_s},\ b = \frac{d}{\sqrt{1+c^2}},\ c = -\frac{\sqrt{n}\tau_s}{\sigma},\ \text{and } d = \frac{\sqrt{n}\left(k^{\pi^{(f)}}(n) - \mu_s\right)}{\sigma}.$$

and U^* and V^* jointly follow the bivariate normal distribution with zero means, unit variances and correlation $\rho^* = \frac{c}{\sqrt{1+c^2}}$. We have used the following two identities:

$$\begin{array}{rcll} \int_a^{\infty}\phi(z)\Phi(cz+d)dz & = & Pr\left(U^* < -a, V^* < b\right) & (4) \\ \int_{-\infty}^{a}\phi(z)\left(1-\Phi(cz+d)\right)dz & = & Pr\left(U^* < a, V^* < -b\right). & \end{array}$$

These two results are proved similarly, the proof of the first identity (4) is given below.

We have

$$\int_a^{\infty} \phi(z)\Phi(cz+d)dz = \int_a^{\infty} \phi(z) \int_{-\infty}^{cz+d} \phi(y)dydz = \int_{-\infty}^{-a} \phi(z) \int_{-\infty}^{-cz+d} \phi(y)dydz.$$

Now we work with the right hand side as follows:

$$\begin{aligned} Pr(U^* < -a, V^* < \tfrac{d}{\sqrt{1+c^2}}) &= \int_{-\infty}^{-a} \int_{-\infty}^{\frac{d}{\sqrt{1+c^2}}} \frac{1}{2\pi\sqrt{1-\rho^{*2}}} e^{-\frac{1}{2(1-\rho^{*2})}\left(u^2-2\rho^* uv+v^2\right)} dudv \\ &= \int_{-\infty}^{-a} \int_{-\infty}^{\frac{1}{\sqrt{1-\rho^{*2}}}\left(\frac{d}{\sqrt{1+c^2}}-\rho^* z\right)} \phi(y)\phi(z)dydz, \\ &= \int_{-\infty}^{-a} \phi(z) \int_{-\infty}^{-cz+d} \phi(y)dydz. \end{aligned}$$

by using the transformation $z = u, \quad y = \frac{1}{\sqrt{1-\rho^{*2}}}(v - \rho^* u)$, and then by substituting the value of ρ^*. This completes the proof. Thus we have an analytic expression for the risk function which can be evaluated for different values of the sample size n and the optimum can be found.

3.1. Results

We investigate the risk function for obtaining analytical solutions. Using a transformation we re-write the risk function as:

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}}) = L_0 Pr\left(U > a, V < b\right) + L_1 Pr\left(U < a, V > b\right), \tag{5}$$

where U and V jointly follow the bivariate normal distribution with zero means, unit variances and correlation

$$\rho = \left(1 + \frac{\sigma^2}{n\tau_s^2}\right)^{-1/2},$$

and

$$a = \frac{\theta_0 - \mu_s}{\tau_s}, \quad b = \rho\frac{k^{\pi^{(f)}}(n) - \mu_s}{\tau_s}.$$

Note that ρ is always non-negative. The joint bivariate distribution comes from the joint probability distribution of $\bar{X}_n$ and θ as implied by the hierarchical modelling of the likelihood and the prior. The quantity a depends on the sampling prior alone while b depends on the sampling prior, the fitting prior and the sample size n. The correlation between θ and $\bar{X}_n$ is ρ which also depends on n.

In order to fix ideas, we provide a particular contour plot of the joint distribution of U and V in Figure 1. The two regions: (1) $U > a, V < b$ and (2) $U < a, V > b$ have been shaded as well. These two regions intersect at the point (a, b). The location of the point (a, b) and the shape of the contours of the bivariate normal distribution will change depending on the values of the sample size, n and the prior parameters. Note, however, that the correlation will always be non-negative. The probabilities of these two regions under the bivariate normal distribution must be controlled to bound the risk function. How will it be possible to make the two probabilities very small? Unfortunately, there is no simple answer to this as the probabilities will be dependent on the actual prior parameters used

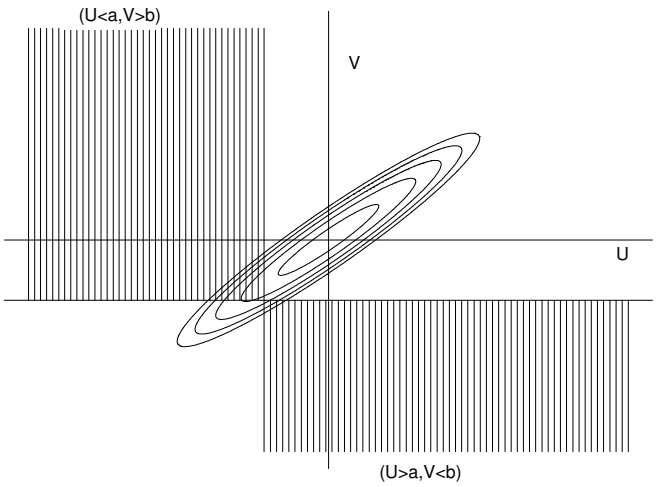

Fig. 1. A particular contour plot.

and the sample size n. However, we provide the following remarks and illustrate the results numerically.

The two probabilities will be small (even for small n) if a and b are of the same sign, and both $|a|$ and $|b|$ are large. This happens when the point (a, b) is far away from the origin in either direction along the major axis of the elliptical contours. When a and b are of opposite sign and at least one of $|a|$ and $|b|$ is large then one of the probabilities will be zero and the other will be large for small values of n. Both the probabilities will be large for small n if the point (a, b) falls inside the high probability region of the contours. To reduce the high probabilities in the last two cases a large value of n will be required. The large value of n will make the value of ρ close to 1 and as a result the contours will shrink to the major axis and both the probabilities will approach zero.

Suppose that τ_f^2 is large corresponding to a non-informative fitting prior. Straightforward calculation yields that

$$b = \rho \left(a - \frac{q\,\sigma}{\tau_s \sqrt{n}} \right). \tag{6}$$

With a further assumption that $L_0 = L_1$, (equivalently $\eta = 1/2$) we have $q = 0$; now b will be a positive multiple of a. Thus a large value of $|a|$ will yield a large value of $|b|$ of the same sign even for small values of n. As a result, even a very small sample size will be sufficient to make the two probabilities small. The quantity a will be large if the mean of the sampling prior μ_s is quite far away from θ_0 in units of τ_s, the standard deviation of the sampling prior. Thus a smaller sample size can be expected if the prior mean is quite far away from the boundary value θ_0 in either direction in units of τ_s when $L_0 = L_1$ and the variance of the fitting prior is large.

If we assume that both the sampling prior and the fitting prior are non-informative (in the sense that both τ_s^2 and τ_f^2 are large) then b will be approximately equal to a and as a

result the sampling prior alone may dictate the sample size. That is, a smaller sample size can be expected if the prior mean is quite far away from the boundary value θ_0 in either direction in units of τ_s. Note that this conclusion does not require the equality assumption of the losses made in the preceding paragraph, since from (6) we have $b \to a$ as $\tau_s^2 \to \infty$ even when $q \neq 0$.

The two probabilities will be moderately large for small values of n if the point (a, b) is near the origin. The origin is the worst position of the point (a, b) for making the probabilities of the two regions small since each of the two regions will intersect heavily with high probability areas of the bivariate normal distribution. Thus the $a = 0$ case for which the mean of the sampling prior is equal to θ_0 will require a larger sample size than the $a \neq 0$ cases. The actual sample size, however, will depend on the magnitude of the quantity b and the tightness of the upper bound on the risk function.

We now turn to numerical illustrations. We have used Monte Carlo integration with 5000 replications to estimate the error probabilities. As a result the optimal sample sizes reported in the tables below are subject to sampling fluctuations. In most cases the sampling variabilities are small and do not affect the main conclusions we have reported. It is possible to remove the sampling fluctuations by additional programming to evaluate probabilities of rectangular regions under bivariate normal distributions.

We make several convenient assumptions; in real applications the auditors must specify the values of all the unknown constants. We suppose that $L_0 = L_1$, i.e. $\eta = 1/2$. We further let $\sigma^2 = \omega\theta_0^2$ and subsequently illustrate with $\omega = 1$. For simplicity and easy interpretation, we parameterise the mean and variance of the fitting prior distribution by:

$$\mu_f = k_1^{(f)}\theta_0, \quad \text{and} \quad \tau_f = k_2^{(f)}\theta_0,$$

assuming $\theta_0 > 0$. Thus the parameter $k_1^{(f)}$ is the ratio of the prior mean and the materiality, θ_0. The parameter $k_2^{(f)}$ does not have any such direct interpretation. We similarly define $k_1^{(s)}$ and $k_2^{(s)}$ for the sampling prior distribution.

The above choices lead to the following values of ρ, a, and b:

$$\rho = \left(1 + \frac{1}{n\left(k_2^{(s)}\right)^2}\right)^{-1/2}, \; a = \frac{1 - k_1^{(s)}}{k_2^{(s)}}, \; b = \rho\left(a + \frac{1 - k_1^{(f)}}{nk_2^{(s)}\left(k_2^{(f)}\right)^2}\right).$$

Note that these three parameters are now free of the value of the per item material error θ_0. We now have the following results:

- Suppose that the sampling and the fitting priors are same; that is, $k_1 = k_1^{(f)} = k_1^{(s)}$ and $k_2 = k_2^{(f)} = k_2^{(s)}$. Then we have

$$\rho = \left(1 + \frac{1}{nk_2^2}\right)^{-1/2}, \; a = \frac{1 - k_1}{k_2}, \; b = \rho\left(a + \frac{1 - k_1}{nk_2^3}\right).$$

  We report the optimal sample sizes in Table 1. As expected the largest sample size is required for the $k_1 = 1$ case where the prior mean is equal to θ_0, the boundary value between the two hypotheses. Note that in this case $a = b = 0$; and sample size decreases as k_2 increases since ρ increases. Thus in this case a larger sample size is required for a tighter prior distribution. The sample sizes are symmetric on either

side of the $k_1 = 1$ column. For very small or large values of k_1 the sample sizes are very small.

- Suppose that $k_1^{(f)} = k_1^{(s)} = 1$, the most difficult case for decision making, but $k_2^{(s)}$ and $k_2^{(f)}$ can be different. In this case we have $a = b = 0$. Hence the optimum sample size will not depend on the value of $k_2^{(f)}$ and it will be decreasing in $k_2^{(s)}$ as we have seen previously. Optimum sample sizes reported in Table 2 confirm this. Here the variance of the fitting prior does not affect the sample sizes (the rows are almost the same) but the sample size decreases as the variance of the sampling prior becomes larger.

- Now suppose that there is a mismatch between the means of the fitting and sampling prior distributions. To illustrate we assume that $k_1^{(s)} = 0.5$ and $k_1^{(f)} = 1$. In this case $k_2^{(f)}$ will not affect the optimum sample size since b is free of $k_2^{(f)}$ and $k_2^{(f)}$ enters into the sample size calculation only through b. Now the optimum sample size will depend on the value of $k_2^{(s)}$. Note that ρ increases as $k_2^{(s)}$ increases but a and b decrease as $k_2^{(s)}$ increases. That is why the optimum sample sizes will first increase and then decrease. Numerical results in Table 3 confirm this.

Observe that the sample sizes are larger if a tighter bound, $M(\eta)$, is required for the risk function. A sample size of ∞ will require a complete audit.

The above results support the auditors' intuition that if the prior mean error is very much less than the material mean error, θ_0, then they only need small sample sizes. The Bayesian approach forces them to make their assumptions more explicit.

4. Exponential error distributions

We now suppose that the errors follow an exponential distribution with mean θ. We thus assume that

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}},\ x > 0,\ \theta > 0.$$

The conjugate prior distribution for θ is the inverse gamma distribution $IG(\alpha, \lambda)$ with the density:

$$\pi(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{\theta^{\alpha+1}} e^{-\lambda/\theta},\ \theta > 0. \tag{7}$$

The sampling prior will be taken as $\pi^{(s)}(\theta) = IG(\alpha_s, \lambda_s)$ and the fitting prior will be assumed to be $\pi^{(f)}(\theta) = IG(\alpha_f, \lambda_f)$. The posterior distribution has the probability density function

$$\pi^{(f)}(\theta|\mathbf{x}^{(n)}) = \frac{(\lambda_f + n\bar{x}_n)^{n+\alpha_f}}{\Gamma(n+\alpha_f)} \frac{1}{\theta^{n+\alpha_f+1}} e^{-(\lambda_f + n\bar{x}_n)/\theta}, \theta > 0.$$

Note that $\pi^{(f)}(\theta|\mathbf{x}^{(n)})$ only depends on $\bar{x}_n$ and henceforth we shall write $\pi^{(f)}(\theta|\bar{x}_n)$. The Bayes rule chooses action a_0 if

$$\begin{array}{lccl} & L_0 \int_{\theta_0}^{\infty} \pi^{(f)}(\theta|\bar{x}_n)d\theta & < & L_1 \int_0^{\theta_0} \pi^{(f)}(\theta|\bar{x}_n)d\theta \\ i.e. & L_0(1-p) & < & L_1 p, \\ \Longrightarrow & p & > & \frac{L_0}{L_0+L_1} \equiv \eta, \end{array}$$

where

$$\begin{aligned} p &= \int_0^{\theta_0} \pi^{(f)}(\theta|\bar{x}_n)d\theta \\ &= \int_0^{\theta_0} \frac{(\lambda_f+n\bar{x}_n)^{n+\alpha_f}}{\Gamma(n+\alpha_f)} \frac{1}{\theta^{n+\alpha_f+1}} e^{-(\lambda_f+n\bar{x}_n)/\theta} d\theta \\ &= \int_{\frac{\lambda_f+n\bar{x}_n}{\theta_0}}^{\infty} \frac{y^{n+\alpha_f-1}e^{-y}}{\Gamma(n+\alpha_f)} dy \\ &= 1 - \int_0^{\frac{\lambda_f+n\bar{x}_n}{\theta_0}} \frac{y^{n+\alpha_f-1}e^{-y}}{\Gamma(n+\alpha_f)} dy. \end{aligned}$$

Thus $p > \eta$ implies

$$\int_0^{\frac{\lambda_f+n\bar{x}_n}{\theta_0}} \frac{y^{n+\alpha_f-1}e^{-y}}{\Gamma(n+\alpha_f)} dy < 1-\eta$$

which in turn implies

$$n\bar{x}_n < k^{\pi^{(f)}}(n) = \theta_0 q_f - \lambda_f,$$

where q_f is the $(1-\eta)$th quantile of the gamma distribution with shape parameter $n+\alpha_f$ and rate parameter 1 denoted by $G(n+\alpha_f, 1)$. Let $G_m(\cdot)$ be the cdf of $G(m,1)$. Then $q_f = G^{-1}_{n+\alpha_f}(1-\eta)$. Now we have

$$Pr(n\bar{X}_n < k^{\pi^{(f)}}(n)|\theta) = G_n\left(\frac{k^{\pi^{(f)}}(n)}{\theta}\right).$$

Note that this probability is zero when $k^{\pi^{(f)}}(n)$ is negative. Now the risk function is given by

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}}) = L_0 \int_{\theta_0}^{\infty} G_n\left(\frac{k^{\pi^{(f)}}(n)}{\theta}\right) \pi^{(s)}(\theta)d\theta + L_1 \int_0^{\theta_0} \left\{1 - G_n\left(\frac{k^{\pi^{(f)}}(n)}{\theta}\right)\right\} \pi^{(s)}(\theta)d\theta.$$

An equivalent expression for this is

$$\begin{aligned} r(\pi^{(s)}, \delta_n^{\pi}) = \; & L_0 \int_0^{\lambda_s/\theta_0} G_n\left(\frac{k^{\pi^{(f)}}(n)y}{\lambda_s}\right) \frac{y^{\alpha_s-1}}{\Gamma(\alpha_s)} e^{-y} dy \\ & + L_1 \int_{\lambda_s/\theta_0}^{\infty} \left\{1 - G_n\left(\frac{k^{\pi^{(f)}}(n)y}{\lambda_s}\right)\right\} \frac{y^{\alpha_s-1}}{\Gamma(\alpha_s)} e^{-y} dy. \end{aligned}$$

We solve the above two integrals using Monte Carlo integration by sampling from $G(\alpha_s, 1)$. In our numerical investigation below we have used 5000 samples from the $G(\alpha_s, 1)$ distribution.

4.1. Numerical results

In our illustration we choose the parameters in the prior distribution, λ and α, by specifying the mean and the standard deviation. We choose the mean to be a multiple, k_1 say, of θ_0 the boundary value of θ. Note that the mean and variance of the prior distribution of θ in (7) are given by

$$\mu = E(\theta) = \frac{\lambda}{\alpha - 1}, \text{ if } \alpha > 1, \quad \text{and} \quad \tau^2 = \text{var}(\theta) = \frac{\lambda}{(\alpha-1)^2(\alpha-2)}, \text{ if } \alpha > 2.$$

We suppose that

$$\mu = k_1\theta_0, \quad \tau = k_2\theta_0,$$

and choose k_1 and k_2 first to specify the prior parameters. These two constraints imply that

$$\alpha = \frac{1}{2}\left(3 + \sqrt{1 + 4k_1/(\theta_0 k_2^2)}\right), \quad \lambda = (\alpha - 1)k_1\theta_0. \tag{8}$$

Thus we can obtain the values of the parameters α and λ once we have specified θ_0, k_1 and k_2.

Now we consider the sampling and fitting prior distributions. The four parameters $k_1^{(s)}$, $k_2^{(s)}$, $k_1^{(f)}$ and $k_2^{(f)}$ will determine the prior parameters α_s, λ_s, α_f and λ_f.

- Suppose that the sampling and the fitting priors are same; that is, $\lambda_s = \lambda_f$ and $\alpha_s = \alpha_f$. Note that these parameters are obtained from (8) by first assuming a particular value for each of $k_1^{(s)} = k_1^{(f)} = k_1$ and $k_2^{(s)} = k_2^{(f)} = k_2$. The optimal sample sizes are reported in Table 4. As expected, we see that the largest sample sizes are needed when k_1 is close to 1, but small sample sizes are needed otherwise. Thus, as intuition suggests, the tighter the concentration of the prior distribution around θ_0, the more samples are required. Note that some sample sizes are ∞ for the $M(\eta) = 0.01$ case. A complete audit will be required in each of those situations.

- In Table 5 we assume that $k_1^{(s)} = k_1^{(f)} = 1$, but we specify different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the sampling and fitting prior. The optimum sample sizes are not affected by the fitting prior distribution; the small variation between the columns is due to the sampling fluctuations in the simulation. As seen before, higher sample sizes are needed for tighter sampling prior distributions (see the variations between the rows of the table).

- Now we suppose that there is a mismatch between the means of the fitting and sampling prior distributions. To illustrate we assume that $k_1^{(s)} = 0.5$ and $k_1^{(f)} = 1$. We report the optimum sample sizes in Table 6 for different values of $k_2^{(s)}$ and $k_2^{(f)}$. The optimum sample size increases when $k_2^{(s)}$ increases and those are not affected by the variance of the fitting prior.

5. A mixture model

Following Cox and Snell (1979) we assume that X_i is non-zero with probability ψ. Let there be m items which resulted in positive errors. Denote these m positive values of X by $Z_1, Z_2, \ldots, Z_m$. Further, we assume that the random sample $Z_1, Z_2, \ldots, Z_m$ follows the exponential distribution with mean μ. Now the parameter of interest is given by $\theta = \psi\mu$. As mentioned earlier Cox and Snell assume that it is the taints Z_i/B_i that have an exponential distribution.

As in Cox and Snell (1979) we assume that a-priori $\psi \sim G(a, a/\psi_0)$ and $\mu \sim IG(b, (b-1)\mu_0)$ independently for suitable values of a, b, ψ_0 and μ_0. These prior distributions are adopted because these are conjugate, and as is well known a simpler analysis ensues under conjugate prior distributions. The joint prior density of ψ and μ is given by:

$$\pi(\psi, \mu) = \left(\frac{a}{\psi_0}\right)^a \frac{1}{\Gamma(a)}\psi^{a-1}e^{-a\psi/\psi_0} \; \frac{\{(b-1)\mu_0\}^b}{\Gamma(b)}\frac{1}{\mu^{b+1}}e^{-(b-1)\mu_0/\mu}, \; \psi > 0, \mu > 0.$$

After some calculation, we see that the induced prior distribution of the parameter of interest θ, $\pi(\theta)$, is given by

$$\pi(\theta) = c\{\pi(\theta)\}\ F_{2a,2b} \text{ where } c\{\pi(\theta)\} = \left\{\frac{(b-1)\psi_0\mu_0}{b}\right\},$$

where F_{ν_1,ν_2} is the standard F random variable with (ν_1, ν_2) degrees of freedom.

Note that the prior mean of $\theta = \psi\mu$ is given by the product $\psi_0\mu_0$; the other hyper-parameters a and b cancel out in the mean. However, the variance of θ depends on all the hyper-parameters and we shall return to their choices later.

The likelihood is obtained by arguing that $m|n,\psi$ follows the Poisson distribution with parameter $n\psi$ and given $m > 0$, $Z_1, \ldots, Z_m$ are i.i.d. exponential random variables with mean μ. The resulting likelihood is given by:

$$L(\psi, \mu; n, m, \mathbf{z}) \propto e^{-n\psi}(n\psi)^m \frac{1}{\mu^m} e^{-\frac{1}{\mu}\sum_{i=1}^m z_i}.$$

The joint posterior distribution of ψ and μ is proportional to $L(\psi, \mu; n, m, \mathbf{z}) \times \pi(\psi, \mu)$, and is given by

$$\pi(\psi, \mu | n, m, \mathbf{z}) \propto e^{-n\psi}(n\psi)^m \frac{1}{\mu^m} e^{-\frac{1}{\mu}\sum_{i=1}^m z_i} \psi^{a-1} e^{-a\psi/\psi_0} \frac{1}{\mu^{b+1}} e^{-(b-1)\mu_0/\mu},$$

for $\psi > 0$ and $\mu > 0$. If $m = 0$ then we simply drop the terms involving m from the above expression to obtain the posterior distribution.

After some integration, we see that the posterior distribution of the quantity $\theta = \psi\mu$ is given by

$$\pi(\theta|\mathbf{x}^{(n)}) = c\left\{\pi(\theta|\mathbf{x}^{(n)})\right\}\ F_{2(m+a),2(m+b)},$$

where

$$c\left\{\pi(\theta|\mathbf{x}^{(n)}\right\} = \left\{\frac{m\bar{z}_m + (b-1)\mu_0}{n + a/\psi_0}\right\}\left(\frac{m+a}{m+b}\right).$$

Note that if $m = 0$ then the posterior distribution is given by

$$\theta = c\left\{\pi(\theta|\mathbf{x}^{(n)})\right\}\ F_{2a,2b}, \text{ where } c\left\{\pi(\theta|\mathbf{x}^{(n)})\right\} = \frac{a}{b}\left\{\frac{(b-1)\mu_0}{n + a/\psi_0}\right\}.$$

Further, when $n = 0$ it is easy to see that the prior and posterior distributions of θ coincide, as expected.

The Bayes rule chooses action a_0 if

$$\int_0^{\theta_0} \pi(\theta|\mathbf{x}^{(n)})d\theta > \frac{L_0}{L_0 + L_1} \equiv \eta,$$

as before. This holds if,

$$\frac{\theta_0}{c\{\pi(\theta|\mathbf{x}^{(n)})\}} \geq q(m, a, b, \eta), \tag{9}$$

where $q(m, a, b, \eta)$ satisfies

$$Pr\left\{F_{2(m+a),2(m+b)} < q(m, a, b, \eta)\right\} = \eta.$$

For the inequality (9), two cases arise depending on the value of m. If $m > 0$, then the Bayes rule chooses action a_0 if

$$\sum_{i=1}^{m} z_i < \theta_0 \frac{m+b}{m+a} \frac{n+a/\psi_0}{q(m,a,b,\eta)} - (b-1)\mu_0. \tag{10}$$

On the other hand, if $m = 0$ then the Bayes rule chooses action a_0 if

$$\theta_0 \frac{b}{a} \frac{n+a/\psi_0}{q(0,a,b,\eta)} > (b-1)\mu_0. \tag{11}$$

Consequently, depending on the value of m the probability $P_\theta\left\{g\left(\mathbf{X}^{(n)}\right) < k(n)\right\}$ will have two different forms. When $m = 0$, the probability is 1 if (11) is satisfied and 0 otherwise. If, however, m is non-zero then the probability is given by

$$P\left(Y < \frac{\theta_0}{\mu} \frac{m+b}{m+a} \frac{n+a/\psi_0}{q(m,a,b,\eta)} - (b-1)\frac{\mu_0}{\mu}\right)$$

where Y follows the gamma distribution $G(m,1)$. Note that this probability will be zero when the right hand side of (10) is negative.

We now introduce the fitting and the sampling priors for calculating the risk function (3). Assume that the forms of the fitting and sampling prior distributions are the same. Let $a_f, b_f, \psi_0^{(f)}, \mu_0^{(f)}$ be the parameters under the fitting prior and $a_s, b_s, \psi_0^{(s)}, \mu_0^{(s)}$ be the parameters under the sampling prior. Now the probabilities $P_\theta\left\{g\left(\mathbf{X}^{(n)}\right) < k(n)\right\}$ and $P_\theta\left\{g\left(\mathbf{X}^{(n)}\right) \geq k(n)\right\}$ are to be calculated using the parameter values $a_f, b_f, \psi_0^{(f)}, \mu_0^{(f)}$ for the fitting prior.

The risk function (3) is now calculated using Monte Carlo sampling from the sampling prior distribution as follows. We first simulate ψ and μ from their sampling prior distributions which have hyper-parameters $a_s, b_s, \psi_0^{(s)}, \mu_0^{(s)}$. The product $\theta = \psi\mu$ is taken as a draw from the sampling prior distribution. Conditional on the draws from the prior distribution we simulate m for a given sample size n using the fact that $m|n,\psi$ follows the Poisson distribution with parameter $n\psi$.

The probability of choosing actions a_0 and a_1 are evaluated under the fitting prior distributions which have hyper-parameters $a_f, b_f, \psi_0^{(f)}, \mu_0^{(f)}$. That is, we set

$$P_\theta\left\{g\left(\mathbf{X}^{(n)}\right) < k^{\pi^{(f)}}(n)\right\} = \begin{cases} I\left(\theta_0 \frac{b_f}{a_f} \frac{n+a_f/\psi_0^{(f)}}{q(0,a_f,b_f,\eta)} > (b_f-1)\mu_0^{(f)}\right), & \text{if } m = 0 \\ G_m\left(\frac{\theta_0}{\mu} \frac{m+b_f}{m+a_f} \frac{n+a_f/\psi_0^{(f)}}{q(m,a_f,b_f,\eta)} - (b_f-1)\frac{\mu_0^{(f)}}{\mu}\right), & \text{otherwise,} \end{cases}$$

where $I(\cdot)$ denotes the indicator function. Subsequently the average risk over 2000 simulation replications produces accurate estimates of the risk $r\left(\pi^{(s)}, \delta_n^{\pi^{(f)}}\right)$.

5.1. Numerical results

The prior mean and variance of θ are given by:

$$\text{mean} = \psi_0\mu_0, \quad \text{variance} = \frac{a+b-1}{a(b-2)}(\psi_0\mu_0)^2.$$

We assume $\psi_0 = 0.01$ and obtain values of μ_0 using the relation $\psi_0\mu_0 = k_1\theta_0$ for different values of k_1. We now set the prior standard deviation k_2 times θ_0 as in the previous example, i.e.

$$\left(\frac{a+b-1}{a(b-2)}\right)^{1/2} \psi_0\mu_0 = k_2\theta_0.$$

This only provides one constraint for two undetermined parameters a and b, so many different strategies can be adopted. In order to ensure positivity of both a and b we require that

$$b > 2 + \frac{k_1^2}{k_2^2}.$$

We let

$$b = 2 + \frac{k_1^2}{k_2^2} + b_0 \text{ and } a = \frac{b-1}{k_2^2(b-2)/k_1^2 - 1},$$

where b_0 is a non-negative parameter. A small value of b_0 makes the prior distribution very spiky and as a result the sample sizes become very large. That is why we illustrate with a moderate value of $b_0 = 10$, although other values can be adopted.

In our illustration, we assume that $\psi_0^{(s)} = \psi_0^{(f)} = 0.01$ to reduce the number of parameters to be given as input for the method. The remaining parameters in the prior distributions are obtained by specifying particular values for k_1 and k_2. Note that we shall have four parameters $k_1^{(f)}, k_2^{(f)}, k_1^{(s)}$ and $k_2^{(s)}$ for the fitting and sampling priors.

- Suppose that the sampling and the fitting priors are same. In this case we have $a_s = a_f$ and $b_s = b_f$. Note that these parameters are obtained by first assuming a particular value for each of $k_1^{(s)} = k_1^{(f)} = k_1$ and $k_2^{(s)} = k_2^{(f)} = k_2$. The optimal sample sizes are reported in Table 7. Here the sample sizes are not symmetric around the $k_1 = 1$ column due to skewness of the mixture distribution. The sample sizes decrease when the prior variance increases as in the previous cases. Also note that there are some optimal sample sizes which are ∞. These are due to the corresponding very small prior variances assumed. The implied prior distribution for each of those cases resembles a spike (centered very close to θ_0) and huge number of samples are required to discriminate between the two hypotheses. In practical auditing terms these infinite sample sizes will require a complete audit.

- In Table 8 we assume that $k_1^{(s)} = k_1^{(f)} = 1$, but we specify different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the sampling and fitting prior. As in the normal and exponential distribution examples the optimum sample sizes are not affected by the fitting prior distribution; the small variation between the columns is due to sampling fluctuations in the simulation. Also as seen previously higher sample sizes are needed for tighter sampling prior distributions (see the variations between the rows of the table).

- Now we suppose that there is a mismatch between the means of the fitting and sampling prior distributions. To illustrate we assume that $k_1^{(s)} = 0.5$ and $k_1^{(f)} = 1$. We report the optimum sample sizes in Table 9 for different values of $k_2^{(s)}$ and $k_2^{(f)}$. The optimum sample size decreases when $k_2^{(s)}$ increases and those are not affected a great deal by the variance of the fitting prior.

In the above discussion we do not consider the $M(\eta) = 0.01$ case since most sample sizes were either very small or very large.

6. Discussion

In this article we have developed a Bayesian method to calculate optimal sample sizes using analytic and numerical simulation methods. Problems in auditing motivated the development of this method, although the approach is general and can be used for similar problems in statistical decision making.

We feel that a clear distinction should be made between the sampling and fitting prior distributions. The sampling prior distribution relates to the data generating mechanism while the fitting prior drives the inference through the posterior distribution. Intuition suggests that a non-informative fitting prior distribution should not influence the sample size and we have demonstrated this here. The sampling prior distribution captures the auditors' usually strong prior belief while the fitting prior distribution is a statistician's device to implement the method.

A key result in the auditing context is that if the prior mean is far away from the boundary value, θ_0 (or the per item material error) then the required sample size is very small which confirms the auditors' views about the value of sampling. In this case a minimum sample size should be set to satisfy auditing standards and to guarantee some level of quality assurance due to sampling. If the prior mean is very close to the material error then as expected, a large sample size is required. This sample size gets even larger for the tighter prior distributions. Also when the upper bound on the two error probabilities, $M(\eta)$, is small the sample sizes become very large. All these results confirm auditors' intuition regarding sample sizes.

The optimal sample sizes have been found under three different parametric assumptions on the error distribution. The assumption of a mixture distribution is appropriate in most auditing problems, although the key conclusions remained the same across the three models. The sample sizes are model dependent if the prior mean of θ is close to the material value θ_0. However, if the prior mean is very much smaller than θ_0, which is often the case in auditing, all models give very small sample sizes.

Acknowledgments

The authors would like to thank Prof Vic Barnett, John Haworth, Robin Swan and colleagues of the National Audit Office for many helpful discussions and suggestions.

References

Adcock, C. J. (1997) Sample size determination: a review. *The Statistician*, **46**, 261–283.

Cox, D. R. and Snell, E. J. (1979) On sampling and the estimation of rare errors. *Biometrika*, **66**, 125–132.

Clarke, B. S. and Yuan, A. (2002) A closed form expression for Bayesian sample sizes. *Technical Report*, Department of Statistics, University of British Columbia.

Johnstone, D. J. (1995) Statistically incoherent hypothesis testing in auditing. *Auditing-A journal of practice and theory*, **14**, 156–175.

Joseph, L., Wolfson, D. B. and du Berger, R. (1995) Sample size determination for binomial proportions via highest posterior density intervals. *The Statistician*, **44**, 143–154.

Laws, D. J. and O'Hagan, A. (2000) Bayesian inference for rare errors in populations with unequal unit sizes. *Applied Statistics*, **49**, 577–590.

Laws, D. J. and O'Hagan, A. (2002) A hierarchical Bayesian model for rare errors. *The Statistician*, **51**, 431–450.

Lindley, D. V. (1997) The choice of sample size. *The Statistician*, **46**, 129–138.

Patterson, E. R. (1993) Strategic Sample-size choice in auditing. *Journal of Accounting Research*, **31**, 272–293.

Shrivastava, R. P. and Shafer, G. R. (1994) Integrating statistical and non-statistical audit evidence using belief functions - a case of variable sampling *International journal of intelligent systems*, **9**, 519–539.

Swan, R. N. (2001) The auditor's application of analytical procedures, the extent of their use and the effectiveness of such procedures. Ph.D. Thesis, University of Southampton.

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. *The Statistician*, **43**, 99–110.

Walker, S. G. (2003) How many samples?: a Bayesian non-parametric approach. *The Statistician*, **52**, 475–482.

Wang, F. and Gelfand, A. E. (2002) A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, **17**, 193–208.

Table 1. Optimum sample size for different values of k_1 and k_2 for the normal example when the fitting and sampling prior are same. Here $\theta_0 = 0.01$, $\eta = 1/2$ and $\sigma^2 = \theta_0^2$.

$M(\eta) = 0.10$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	2	12	28	37	31	13	2
1.0	6	8	10	11	11	9	6
1.5	5	5	5	6	6	6	5
2.0	4	4	4	4	4	4	4
2.5	3	3	3	3	3	3	3
$M(\eta) = 0.05$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	12	55	106	157	114	57	11
1.0	23	30	38	39	35	31	25
1.5	16	16	18	19	19	19	14
2.0	10	11	11	11	12	11	10
2.5	7	7	8	8	8	8	8
$M(\eta) = 0.01$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	296	1149	2186	3005	2300	1118	346
1.0	448	598	709	746	749	618	417
1.5	290	296	374	357	317	330	287
2.0	183	175	206	209	194	202	160
2.5	112	144	134	114	143	129	124

Table 2. Optimum sample size for different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the normal example when $k_1^{(s)} = k_1^{(f)} = 1$. Here $\theta_0 = 0.01$, $\eta = 1/2$ and $\sigma^2 = \theta_0^2$.

$M(\eta) = 0.10$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	40	37	39	40	39
1.0	11	11	11	11	11
1.5	5	6	5	6	6
2.0	4	4	4	4	4
2.5	3	3	3	3	3
$M(\eta) = 0.05$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	140	148	152	151	152
1.0	41	43	41	41	41
1.5	19	19	20	18	18
2.0	11	12	11	11	12
2.5	8	8	8	8	8
$M(\eta) = 0.01$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	2804	2698	2966	2578	2877
1.0	746	661	763	698	765
1.5	341	326	348	307	352
2.0	168	189	208	183	186
2.5	134	121	138	125	145

Table 3. Optimum sample size for different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the normal example when $k_1^{(s)} = 0.5$ and $k_1^{(f)} = 1.0$. Here $\theta_0 = 0.01$, $\eta = 1/2$ and $\sigma^2 = \theta_0^2$.

$M(\eta) = 0.10$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	12	12	12	12	11
1.0	16	16	16	16	16
1.5	13	12	12	13	13
2.0	9	9	9	9	8
2.5	7	7	7	7	7
$M(\eta) = 0.05$					
			$k_2^{(f)}$		
0.5	26	27	27	28	28
1.0	59	60	57	57	60
1.5	47	46	44	47	48
2.0	32	33	31	29	30
2.5	24	24	22	24	23
$M(\eta) = 0.01$					
			$k_2^{(f)}$		
0.5	269	220	235	248	282
1.0	1070	1129	1021	1146	1027
1.5	855	834	837	770	808
2.0	587	636	596	615	572
2.5	463	378	441	369	379

Table 4. Optimum sample size for different values of k_1 and k_2 for the exponential example when the fitting and sampling prior are same. Here $\theta_0 = 0.01$ and $\eta = 1/2$.

$M(\eta) = 0.10$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	2	2	2	193	79	2	2
1.0	2	2	18	97	86	2	2
1.5	2	2	17	67	76	30	2
2.0	2	2	14	48	64	37	2
2.5	2	2	14	43	52	38	12
$M(\eta) = 0.05$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	2	2	80	739	377	2	2
1.0	2	2	89	371	355	78	2
1.5	2	2	73	259	300	131	2
2.0	2	4	67	182	253	154	36
2.5	2	5	56	162	215	151	68
$M(\eta) = 0.01$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	2	2	1822	∞	∞	287	2
1.0	2	67	1734	∞	∞	1750	118
1.5	2	118	1593	∞	∞	2562	598
2.0	2	177	1481	3086	4391	2846	1038
2.5	2	159	1229	3082	3646	2938	1277

Table 5. Optimum sample size for different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the exponential example when $k_1^{(s)} = k_1^{(f)} = 1$. Here $\theta_0 = 0.01$ and $\eta = 1/2$. All sample sizes are ∞ for the $M(\eta) = 0.01$ case.

$M(\eta) = 0.10$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	189	184	180	184	183
1.0	100	94	94	98	92
1.5	67	64	64	63	65
2.0	49	50	49	53	51
2.5	43	42	40	42	43
$M(\eta) = 0.05$					
			$k_2^{(f)}$		
0.5	735	709	732	717	731
1.0	364	364	382	370	369
1.5	256	266	246	249	253
2.0	204	185	202	196	201
2.5	164	155	171	157	166
$M(\eta) = 0.01$					
			$k_2^{(f)}$		
0.5	∞	∞	∞	∞	∞
1.0	4219	4280	4584	3937	4999
1.5	3250	3376	3253	2771	2870
2.0	2116	2672	2432	2366	2521
2.5	2186	2000	2197	1814	1736

Table 6. Optimum sample size for different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the exponential example when $k_1^{(s)} = 0.5$ and $k_1^{(f)} = 1$. Here $\theta_0 = 0.01$ and $\eta = 1/2$.

$M(\eta) = 0.10$					
	$k_2^{(f)}$				
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	2	2	2	2	2
1.0	2	2	2	2	2
1.5	2	2	2	2	2
2.0	2	2	2	2	2
2.5	2	2	2	2	2
$M(\eta) = 0.05$					
	$k_2^{(f)}$				
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	2	2	2	2	2
1.0	5	4	5	5	4
1.5	6	7	7	6	6
2.0	7	8	8	7	8
2.5	8	8	8	7	8
$M(\eta) = 0.01$					
	$k_2^{(f)}$				
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	23	25	25	27	25
1.0	88	81	91	93	86
1.5	147	147	148	152	155
2.0	153	182	156	175	168
2.5	159	185	179	196	169

Table 7. Optimum Sample size for different values of k_1 and k_2 for the mixture example when the fitting and sampling prior are same. Here $\theta_0 = 0.01$, $\eta = 1/2$, $\psi_0^{(s)} = \psi_0^{(f)} = 0.01$ and $\mu_0^{(s)} = \mu_0^{(f)}$.

$M = 0.10$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	2	65	1500	∞	∞	3341	2
1.0	2	37	275	1011	2285	3153	2809
1.5	2	6	80	307	746	1405	1871
2.0	2	2	27	103	318	701	1017
2.5	2	2	12	43	131	317	589
$M = 0.05$							
	k_1						
k_2	0.25	0.5	0.75	1.0	1.25	1.5	1.75
0.5	22	962	∞	∞	∞	∞	2
1.0	3	251	1572	4526	∞	∞	∞
1.5	2	74	451	1457	3403	∞	∞
2.0	2	18	167	653	1674	2614	4578
2.5	2	9	69	287	817	1490	2462

Table 8. Optimum Sample size for different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the mixture example. Here $\theta_0 = 0.01$; $k_1^{(s)} = k_1^{(f)} = 1$, $\psi_0^{(s)} = \psi_0^{(f)} = 0.01$, $\eta = 1/2$.

$M(\eta) = 0.10$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	∞	∞	∞	∞	∞
1.0	1086	1032	1058	1126	1039
1.5	337	334	304	301	325
2.0	128	142	99	129	114
2.5	75	66	64	50	48
$M(\eta) = 0.05$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	∞	∞	∞	∞	∞
1.0	4516	4215	4768	4644	4261
1.5	1555	1497	1424	1560	1537
2.0	633	630	545	639	538
2.5	315	284	274	241	239

Table 9. Optimum Sample size for different values of $k_2^{(s)}$ and $k_2^{(f)}$ for the mixture example. Here $\theta_0 = 0.01$; $k_1^{(s)} = 0.5, k_1^{(f)} = 1$, $\psi_0^{(s)} = \psi_0^{(f)} = 0.01$, $\eta = 1/2$.

$M(\eta) = 0.10$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	499	984	1003	991	971
1.0	37	58	63	65	63
1.5	4	6	5	5	8
2.0	2	3	2	2	3
2.5	2	2	2	2	2
$M(\eta) = 0.05$					
			$k_2^{(f)}$		
$k_2^{(s)}$	0.5	1.0	1.5	2.0	2.5
0.5	1843	2584	2482	2651	2613
1.0	451	761	871	879	1010
1.5	74	68	66	66	76
2.0	24	18	22	16	23
2.5	9	9	8	10	8