ELSEVIER

International Journal of Approximate Reasoning 25 (2000) 43–70

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Monetary unit sampling: a belief-function implementation for audit and accounting applications

Peter R. Gillett *

Department of Accounting and Information Systems, Faculty of Management, Rutgers: The State University of New Jersey, School of Business – New Brunswick, Janice H. Levin Building, Rockafeller Road, Piscataway, NJ 08854-8054, USA

Received 1 May 1999; received in revised form 1 January 2000; accepted 1 February 2000

Abstract

Audit procedures may be planned and audit evidence evaluated using monetary unit sampling (MUS) techniques within the context of the Dempster–Shafer theory of belief functions. This article shows: (1) how to determine an appropriate sample size for MUS in order to obtain a desired degree of belief that the upper bound for misstatements lies within a given interval; and (2) what level of belief in a specified interval is obtained given a sample result. The results are consistent with the view that a specified level of belief in an interval is semantically a stronger claim than the same numerical level of probability. The paper describes two variants of MUS in both probability and belief-function forms, emphasizing the systematic similarities and the numerical differences between the two frameworks. The results, based on the Poisson distribution, extend results already available for mean-per-unit variables sampling, and may readily be developed to give similar results for the binomial distribution.

Keywords: Belief functions; Auditing; Monetary unit sampling

* Tel.: +1-732-445-4765.
E-mail address: gillett@business.rutgers.edu (P.R. Gillett).

PII: S 0 8 8 8 - 6 1 3 X (0 0) 0 0 0 4 6 - 3

1. Introduction

The assurance provided by the evidence gathered during the audit process is often represented in analytical models by probabilities, but it has been argued [1] that Dempster–Shafer belief functions provide a preferable alternative. The level of assurance provided by certain audit procedures (e.g., inquiry and observation) may be subjectively assessed by the auditor within this framework, just as it may within the probability framework. If belief functions are to provide a reasonable alternative for modeling the aggregation of audit evidence, however, it is important that they are able to represent the assurance provided by the application of statistical procedures such as audit sampling. The present study shows how this may be done in the case of monetary unit sampling (MUS); the method is first described as it is applied under probability theory, and then a technique for planning and evaluating monetary unit samples using belief functions is described and contrasted. It is not the purpose of this paper to make the case for the advantages of belief functions over probabilities, although the issue is discussed in outline below. Rather, the study presumes that modelers will wish to use whatever is the most appropriate form of uncertain reasoning for the problem in hand, and concentrates on showing how MUS may be carried out in a belief-function framework. Although the focus of the examples discussed in detail is clearly on auditing, the methods set forth in this study are available equally for other accounting applications of MUS.

The process of auditing is essentially concerned with the aggregation of evidence in support of the auditor's opinion. Indeed, auditing has been defined as "*a systematic process of objectively obtaining and evaluating evidence* regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users" [2, italics added]. Arens and Loebbecke [3] emphasize the central role of evidence-gathering: "We believe that the most fundamental concepts in auditing relate to determining the nature and amount of evidence the auditor should accumulate after considering the unique circumstances of each engagement." The purpose of this article is to show that audit evidence obtained using MUS techniques may be combined with other audit evidence within a Dempster–Shafer belief-function framework. In particular, the article shows: (1) how to determine an appropriate sample size for MUS in order to obtain a desired degree of belief that the upper bound for misstatements lies within a given interval; and (2) what level of belief in a specified interval is obtained given a sample result.

Although there are a number of different types of audit, this paper is concerned with the audit of financial statements in order to determine their compliance with generally accepted accounting principles (GAAP). The financial statements are the responsibility of management, whose representations regarding the company's financial position they contain. Implied by the financial statements are certain management assertions regarding the various account balances and classes of transactions that are included [4], and these assertions form the basis of a number of audit objectives investigated by the auditor. For example, for accounts receivable balances, management asserts via the financial statements that recorded accounts receivable exist (existence), that all existing accounts receivable are included (completeness), that accounts receivable are accurate (valuation), etc. For full details, any standard auditing text may be consulted; see, for example, [3].

For each of the audit objectives based on management assertions the auditor will plan to gather evidence (except where the amount involved be considered immaterial, or the risk is considered insignificant). However, audit evidence is expected to be persuasive rather than conclusive, and even after audit evidence has been gathered and evaluated, there remains the possibility that an auditor's conclusion that an account is not materially misstated may still be incorrect. This possibility is commonly referred to as audit risk, and its effective management is crucial to proper audit planning. In attempting to reduce audit risk to an acceptable level, the auditor may obtain different types of evidence from a variety of different sources. Note also that certain items of evidence may bear on more than one assertion or objective (as for example, the confirmation of accounts receivable bears on both existence and valuation assertions) and that the various audit objectives are not independent of each other. For example, the existence of sales transactions and the completeness of cash receipts clearly impact the existence of residual accounts receivable. Proper evaluation of a large body of evidence gathered during the audit process is, therefore, a rich and complex problem in the management of uncertainty.

The most widely used and discussed model for the aggregation of audit evidence is commonly called the audit risk model [5,6]. It proposes that audit risk is the combined effect of *inherent risk* (the susceptibility of an account balance or class of transactions to material error, assuming that there were no related internal accounting controls), *control risk* (the risk that material error that could occur in an account balance or class of transactions will not be prevented or detected on a timely basis by the system of internal accounting control), and *detection risk* (the risk that an auditor's procedures will lead to the conclusion that material error in an account balance or class of transactions does not exist when in fact such error does exist). This model is often summarized in the formula

$$\mathrm{AR} = \mathrm{IR} \times \mathrm{CR} \times \mathrm{DR}.$$

Since the beginning, however, this model has been subject to widespread criticism by academics. Critics have argued that it is inappropriate because the risks are not properly independent [7], that the probability model is not properly specified [8,9], that inherent risk should be treated as a Bayesian prior [10], and that the outcome space is not properly considered [11,12].

In considering alternative approaches to the audit risk model, some authors, mindful of the rich interdependencies between assertions, and between audit procedures and assertions, have suggested that a network approach is appropriate [13–16]. These authors have additionally proposed that Dempster–Shafer belief functions be used to represent uncertainty in the audit process rather than probability theory [17]. There has been some support for this approach from other authors in recent years [18,19], partly because of semantic considerations regarding the representation of ignorance. These issues arise even in the case of statistical evidence. To illustrate the semantic limitations of the probability theory model, consider an auditor who, having sent out a certain number of positive confirmation requests for accounts receivable, finds that most of them are returned and agree with the recorded amounts, but that the remainder are not returned. Assuming that the auditor concludes that the evidence warrants a 70% probability that accounts receivable are not materially misstated, probability theory imposes the constraint that (absent any other evidence) there is a 30% probability that accounts receivable *are* materially misstated. Yet the auditor has no evidence that accounts receivable are misstated (such as it might have arisen if any of the returned confirmations had disagreed with the recorded amounts). There is merely an insufficiency of evidence in support of the recorded amounts. Belief functions, on the other hand, allow the auditor to assign a belief that accounts receivable are not misstated based on the evidence, no belief that accounts receivable are misstated, and a residual amount that represents ignorance (that may require the gathering of further evidence). Thus the auditor using belief functions can distinguish this case from another where all the confirmations were returned, most agreeing, but the remainder disagreeing, with the recorded amounts.

The majority of Srivastava and Shafer's published work in this area has concentrated on introducing belief functions to the auditing literature, showing how the aggregation of audit evidence may be represented using belief functions, and discussing some of the complexities of a network approach. Whether audit evidence is represented by probabilities or belief functions, the auditor will gather a variety of different kinds of evidence: assessments of inherent risk, tests of controls, analytical procedures, and tests of details of transactions and balances, some of which may be performed on a sample basis. For some audit procedures, an estimate of associated risk may be made subjectively within either framework (for example, inquiry and observation procedures). The relative ease or difficulty the auditor experiences in formulating estimates of the assurance provided by such procedures within either the probability or the belief-function framework is an interesting empirical question that is presently the subject of behavioral studies. For tests carried out on a sample basis, however, there are well-known statistical techniques for relating sample sizes and results to probabilities associated with audit conclusions. How can this be done for belief functions? Based on principles described in Shafer's seminal work [20], this question has been answered for mean-per-unit variables sampling [21]. Concern about the skewed nature of accounting populations, coupled with ever-increasing demands to improve audit efficiency, however, have lead to much greater use of an alternative sampling approach, MUS, in cases where substantive sampling is to be applied. The purpose of this paper is to show how sample sizes and results may be related to beliefs in specified intervals using MUS, so that a network of both statistical and non-statistical audit evidence may be represented within the belief-function framework.

The remainder of the paper is divided into four further sections. Section 2 of the paper describes the main principles of MUS, concentrating on two popular approaches including the method proposed by Kaplan [22] for controlling alpha risk. Section 3 gives an outline of consonant belief functions and their relation to statistical evidence, and Section 4 provides a belief-function approach to MUS. Section 5 summarizes and concludes with future research issues.

2. Monetary unit sampling

The sampling techniques classically used in auditing are divided into two categories: attribute sampling (including discovery sampling, sequential sampling etc.) and variables sampling (including mean-per-unit sampling, stratified sampling, difference and ratio methods etc.). These techniques are described in a variety of statistical texts such as [23], and their application to auditing is more fully discussed in such texts as [24] or [25]. The use of attribute sampling for tests of controls has been relatively uncontroversial; although the hypergeometric distribution is theoretically correct for most audit applications, the binomial distribution or the Poisson distribution have been widely used as conservative approximations yielding more tractable calculations and simpler tables prior to the widespread use of computers in the audit process.

Variables sampling, however, relies on the use of the central limit theorem, and this has given rise to some concern since accounting populations are often quite skewed. As a result, appropriate sample sizes for the application of the central limit theorem are likely to be larger than desirable in the audit context. This difficulty is compounded by the effect on sample sizes of variability within the accounting population. MUS has been developed in response to these concerns. Although the idea of using an individual dollar as the sampling unit was suggested by Deming [26], the idea was first introduced into the auditing literature by van Heerden [27], and into the US auditing profession by Stringer [28], among others. Its use was popularized, however, by the work of Albert Teitlebaum in conjunction with Rod Anderson and Donald Leslie [29–31].

Over the intervening years the technique, originally known as dollar unit sampling (DUS), has also been called MUS in deference to its more general application. Since individual dollars (or other monetary units) are used as the sampling units, higher value items containing more dollars are more likely to be selected, and the technique is sometimes called sampling with probability proportional to size (PPS) as a result. One popular method of selection is based on a systematic sample [1] of the total monetary units, and this has given rise to the alternative nomenclature: cumulative monetary amount sampling (CMA). The advantages and disadvantages of the MUS approach are discussed, for example, in the AICPA's audit sampling guide [32].

In the MUS approach, the underlying idea is to treat the population as consisting of individual dollars, each of which may or may not be misstated. It is, therefore, essentially an attribute sampling application. Its use in practice is nevertheless subject to a wide range of variants. Suppose that an invoice for \$1,000 is selected for audit by virtue of its 850th dollar being randomly selected, and that the audited value of the invoice turns out to be only \$800; i.e., there is a \$200 overstatement. In van Heerden's original formulation, for example, the 200 overstated dollars might be allocated to be the first \$200 dollars of the invoice, or the final \$200; depending on this choice, the 850th dollar may be classified as overstated or not. Teitlebaum showed that this leads to obviously undesirable variability in audit applications [30], and proposed instead that each of the 1000 dollars in the invoice be considered as 20% overstated. The present article presumes the use of this method; the associated evaluation technique is usually referred to as the Stringer bound. Since this technique mixes attribute sampling with values it has been described as a combined attributes and variables method (CAV).

Individual dollars may be selected for audit by simple *random* sampling, by a *systematic* sample based on cumulative monetary amounts, or by a method known as *cell-selection* in which the population is divided into cells of equal value, and a dollar is selected from each [30,31]. Although cell-selection has some advantages, the more basic systematic sampling technique is widely taught and used, and its use will be presumed for the remainder of this paper.

In practice, it is common for auditors using MUS to control only the beta risk (the risk of incorrect acceptance) and to ignore alpha risk (the risk of incorrect rejection). However, Kaplan has shown how the technique may be extended to take account of alpha risk [22], and both methods are considered below. Teitlebaum's and Kaplan's presentations rely on use of the Poisson distribution for attribute sampling, and this is still the basis of many modern auditing texts, see for example, [25]. For consistency with their presentations,

[1] In a systematic sample, the population recorded book value B is divided by the calculated sample size n to give a sampling interval S. From a random start s chosen in the range 1–S, every Sth monetary unit is selected from a cumulative total of the values of the population items; i.e., the selected monetary units are s, $s + S$, $s + 3S$ etc.

the Poisson distribution is used in this paper, although the binomial distribution (which gives slightly smaller sample sizes) is used in some texts, e.g., [3], as well as in certain audit software.

The Stringer bound is widely used, and is presumed in this paper, However, it has no rigorous mathematical justification, [2] and has been found in many simulations to be significantly conservative. A number of alternative evaluation techniques have been proposed, but are not addressed in the present study; for example, the modified moment bound [34,35], the multinomial-Dirichlet bound [36], the beta-normal bound [37], or the robust Bayesian bound [38].

2.1. Sample size determination

In a MUS application, let n be the sample size, p the probability of an individual monetary unit being misstated, $\lambda = np$ the mean number of misstatements, k the critical value for the number of misstatements for the sampling application, B the recorded book value for the population total, T the tolerable misstatement based on planning materiality, E the expected amount of misstatement in the population, and β the acceptable level for the risk of incorrect acceptance. Then since the probability p is given by $p = T/B$, so that $\lambda = np = nT/B$, and up to k errors will be accepted, the sample size for the application is determined [29] by solving

$$\sum_{j=0}^{k} \mathrm{e}^{-\lambda} \frac{\lambda_j}{j!} = \beta \tag{1}$$

for λ, and then letting $n = B\lambda/T$.

When no errors are to be accepted, Eq. (1) may be solved explicitly to give $\lambda = -\ln\beta$ and hence

$$n = \frac{-T\ln\beta}{B}. \tag{2}$$

When errors are expected, the solution of (1) rapidly becomes tedious, and although microcomputers may now readily be used for this purpose, tables were used previously to provide values for λ. Table 1, which is similar to the table given in [31] may be used for this purpose. Values of λ may be found in the columns headed UEL, the table is accessed in the row for the appropriate expected number of errors, in the column for confidence $= 1 - \beta$.

Consider, for example, a population with a recorded book value of \$5,000,000 that the auditor wishes to audit using MUS with a tolerable misstatement of \$500,000 and an acceptable risk of incorrect acceptance of 5%;

[2] However, an asymptotic result has recently been obtained, see [33].

Table 1
UEL is $P_k(\beta)$ such that $\sum_{j=0}^{k} e^{-P_k}((P_k)^j/j!) = \beta$, where k = number of errors; PGW $= P_k(\beta) - P_{k-1}(\beta) - 1$

Number of errors	Confidence levels (1−β)													
	75%		80%		85%		90%		95%		98%		99%	
	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW
0	1.386		1.609		1.897		2.303		2.996		3.689		4.605	
1	2.693	0.306	2.994	0.385	3.372	0.475	3.890	0.587	4.744	0.748	5.572	0.883	6.638	1.033
2	3.920	0.228	4.279	0.285	4.723	0.351	5.322	0.433	6.296	0.552	7.225	0.653	8.406	0.768
3	5.109	0.189	5.515	0.236	6.014	0.290	6.681	0.358	7.754	0.458	8.767	0.543	10.045	0.639
4	6.274	0.165	6.721	0.206	7.267	0.253	7.994	0.313	9.154	0.400	10.242	0.474	11.605	0.560
5	7.423	0.148	7.906	0.185	8.495	0.228	9.275	0.281	10.513	0.360	11.668	0.427	13.108	0.504
6	8.558	0.136	9.075	0.169	9.703	0.208	10.532	0.257	11.842	0.329	13.059	0.391	14.571	0.462
7	9.684	0.126	10.233	0.157	10.897	0.193	11.771	0.239	13.148	0.306	14.423	0.363	16.000	0.429
8	10.802	0.118	11.380	0.147	12.078	0.181	12.995	0.224	14.435	0.287	15.763	0.341	17.403	0.403
9	11.914	0.111	12.519	0.139	13.249	0.171	14.206	0.211	15.705	0.271	17.085	0.322	18.783	0.380
10	13.020	0.106	13.651	0.132	14.411	0.162	15.407	0.201	16.962	0.257	18.390	0.306	20.145	0.362
11	14.121	0.101	14.777	0.126	15.566	0.155	16.598	0.191	18.208	0.245	19.682	0.292	21.490	0.345
12	15.217	0.097	15.897	0.121	16.715	0.149	17.782	0.183	19.443	0.235	20.962	0.280	22.821	0.331
13	16.310	0.093	17.013	0.116	17.857	0.143	18.958	0.176	20.669	0.226	22.230	0.269	24.139	0.318
14	17.400	0.090	18.125	0.112	18.995	0.138	20.128	0.170	21.886	0.218	23.490	0.259	25.446	0.307
15	18.486	0.087	19.233	0.108	20.128	0.133	21.292	0.164	23.097	0.211	24.740	0.251	26.743	0.297
16	19.570	0.084	20.338	0.105	21.257	0.129	22.452	0.159	24.301	0.204	25.983	0.243	28.030	0.288
17	20.652	0.081	21.439	0.102	22.382	0.125	23.606	0.155	25.499	0.198	27.219	0.236	29.310	0.279
18	21.731	0.079	22.538	0.099	23.504	0.122	24.756	0.150	26.692	0.193	28.448	0.229	30.581	0.271

19	22.808	0.077	23.634	0.096	24.622	0.118	25.903	0.146	27.879	0.187	29.671	0.223	31.845	0.264
20	23.883	0.075	24.728	0.094	25.737	0.115	27.045	0.143	29.062	0.183	30.888	0.218	33.103	0.258
21	24.956	0.073	25.819	0.091	26.850	0.113	28.184	0.139	30.240	0.178	32.101	0.212	34.355	0.252
22	26.028	0.072	26.909	0.089	27.960	0.110	29.320	0.136	31.415	0.174	33.308	0.208	35.601	0.246
23	27.098	0.070	27.996	0.087	29.068	0.108	30.453	0.133	32.585	0.171	34.511	0.203	36.841	0.241
24	28.167	0.069	29.082	0.086	30.173	0.105	31.584	0.130	33.752	0.167	35.710	0.199	38.077	0.236
25	29.234	0.067	30.166	0.084	31.276	0.103	32.711	0.128	34.916	0.164	36.905	0.195	39.308	0.231
26	30.300	0.066	31.248	0.082	32.378	0.101	33.836	0.125	36.077	0.161	38.096	0.191	40.534	0.227
27	31.365	0.065	32.329	0.081	33.477	0.099	34.959	0.123	37.234	0.158	39.284	0.188	41.757	0.222
28	32.428	0.064	33.408	0.079	34.575	0.098	36.080	0.121	38.389	0.155	40.468	0.184	42.975	0.218
29	33.491	0.062	34.486	0.078	35.671	0.096	37.199	0.119	39.541	0.152	41.649	0.181	44.190	0.215
30	34.552	0.061	35.563	0.077	36.765	0.094	38.315	0.117	40.691	0.150	42.827	0.178	45.401	0.211
35	39.845	0.059	40.928	0.073	42.215	0.090	43.872	0.111	46.404	0.143	48.677	0.170	51.408	0.201
40	45.118	0.055	46.269	0.068	47.635	0.084	49.390	0.104	52.069	0.133	54.469	0.158	57.347	0.188
45	50.375	0.051	51.589	0.064	53.029	0.079	54.878	0.098	57.695	0.125	60.214	0.149	63.231	0.177
50	55.618	0.049	56.893	0.061	58.403	0.075	60.339	0.092	63.287	0.118	65.919	0.141	69.067	0.167
55	60.850	0.046	62.181	0.058	63.758	0.071	65.779	0.088	68.851	0.113	71.590	0.134	74.863	0.159
60	66.071	0.044	67.457	0.055	69.098	0.068	71.199	0.084	74.390	0.108	77.232	0.128	80.625	0.152
65	71.283	0.042	72.722	0.053	74.424	0.065	76.602	0.081	79.907	0.103	82.848	0.123	86.355	0.146
70	76.487	0.041	77.977	0.051	79.738	0.063	81.990	0.078	85.405	0.100	88.441	0.119	92.059	0.141
75	81.685	0.039	83.223	0.049	85.041	0.061	87.364	0.075	90.885	0.096	94.013	0.114	97.738	0.136
80	86.875	0.038	88.461	0.048	90.334	0.059	92.727	0.072	96.350	0.093	99.567	0.111	103.395	0.131
85	92.060	0.037	93.692	0.046	95.618	0.057	98.078	0.070	101.801	0.090	105.104	0.107	109.032	0.127
90	97.240	0.036	98.916	0.045	100.894	0.055	103.419	0.068	107.239	0.088	110.625	0.104	114.650	0.124
95	102.415	0.035	104.134	0.044	106.162	0.054	108.751	0.066	112.664	0.085	116.133	0.101	120.252	0.120
100	107.585	0.034	109.346	0.042	111.424	0.052	114.075	0.065	118.079	0.083	121.627	0.099	125.839	0.117

suppose also that three of the selected dollars are expected to be in error. From Table 1, $\lambda = 7.754$ for three errors and a confidence of 95%. The required sample size is therefore for three errors and a confidence of 95%. The required sample size is therefore

$$\frac{5{,}000{,}000 \times 7.754}{500{,}000} = 77.54 \approx 78.$$

Tolerable misstatement is determined by the auditor in the light of planning materiality, and the beta risk may be determined by reference to the audit risk model – but how is the expected number of errors to be determined? When the auditor has reason to expect a particular number of errors, the calculation proceeds in a straightforward manner, as outlined above.

More commonly, the auditor does not have in mind the required expected number of errors, but rather some amount E of expected misstatement (perhaps based on the experience of prior years). When MUS is based on binomial attribute sampling tables, this poses no difficulty. The expected error rate is calculated as

$$\mathrm{EER} = \frac{E \times 100\%}{B},$$

the tolerable error rate is calculated as

$$\mathrm{TER} = \frac{T \times 100\%}{B}$$

and the tables are used in the normal way to give the appropriate sample size [32]. In the Poisson distribution approach, matters become a little more complicated. The (conservative) assumption made in MUS planning is that all misstatements are 100% misstatements. Thus $k/n = E/B$ or, equivalently

$$E = B \times \frac{k}{n} = Bk \times \frac{T}{B\lambda} = \frac{kT}{\lambda}. \tag{3}$$

For $k = 0, 1, 2, \ldots$, Eq. (1) may be solved iteratively for λ, and the resulting value used to compute E from Eq. (3) until the required level of expected error is reached.

In the example considered earlier, suppose that the expected aggregate error was \$193,449. For $k = 0$, Table 1 gives $\lambda = 2.996$, but yields a value for E of 0. For $k = 1$, Table 1 gives $\lambda = 4.744$, giving a value for E of \$105,396. Similarly, for $k = 2$, Table 1 gives $\lambda = 6.296$, giving a value for E of \$158,831. Finally, for $k = 3$, Table 1 gives $\lambda = 7.754$, giving a value for E of \$193,449. This somewhat complicated procedure is necessary because a given number of errors correspond to a different amount of error at different levels of beta risk.

It does not always happen that the amount of expected error corresponds to an exact number of errors. If the expected error falls between k and $k+1$ errors, giving rise to λ_k and λ_{k+1}, respectively, producing expected errors of E_k and E_{k+1}, a conservative procedure would be to use a sample size based on λ_{k+1}. Alternatively, linear interpolation may be used. Thus,

$$\lambda = \lambda_k + \frac{E - E_k}{E_{k+1} - E_k} \times (\lambda_{k+1} - \lambda_k). \tag{4}$$

In the above example, if the auditor expected aggregate error of \$172,678, then Eq. (4) would give

$$\lambda = 6.296 + \frac{172,678 - 158,831}{193,449 - 158,831} \times (7.754 - 6.296) = 6.879$$

and

$$n = \frac{5{,}000{,}000 \times 6.879}{500{,}000} = 68.79 \approx 69.$$

2.1.1. Controlling alpha risk

The methods described above for determining sample sizes for MUS do not explicitly control the risk of incorrect rejection. A method for doing this has been devised [22] as follows. Suppose that the auditor determines some low rate of error, μ, such that it would be undesirable to reject a population with this error rate more often than alpha percent of the time. Then the sample size is determined by finding the smallest value of k such that

$$1 - \sum_{j=0}^{k} \mathrm{e}^{-n\mu} \frac{(n\mu)^j}{j!} \leqslant \alpha \tag{5}$$

and

$$\sum_{j=0}^{k} \mathrm{e}^{-nT/B} \frac{(nT/B)^j}{j!} \beta. \tag{6}$$

Tables 2 and 3 provided by Kaplan [22] are used to reduce the tedious calculation for alpha and beta risks, respectively.

Suppose, for example, that the auditor wishes to audit a balance of \$5,000,000 with a tolerable misstatement of \$500,000 at an acceptable risk of incorrect acceptance of 5%, and an alpha risk of 10% of rejecting the population when the misstatement rate is only 0.5%.

If the auditor only accepts the population when there are no errors, then from Tables 2 and 3 we see that

$$\lambda = 0.005n \leqslant 0.105 \quad \text{or} \quad n \leqslant 21,$$

Table 2
$\lambda(\alpha, k)$ such that $\sum_{j=0}^{k} e^{-\lambda}(\lambda^j/j!) = 1-\alpha$, where k = number of errors

Number of errors	Alpha risk α				
	0.01	0.05	0.10	0.25	0.50
0	0.010	0.051	0.105	0.288	0.693
1	0.149	0.355	0.532	0.961	1.68
2	0.436	0.818	1.10	1.73	2.67
3	0.823	1.37	1.74	2.54	3.67
4	1.28	1.97	2.43	3.37	4.67
5	1.79	2.61	3.15	4.22	5.67

Table 3
$\lambda(\beta, k)$ such that $\sum_{j=0}^{k} e^{-\lambda}(\lambda^j/j!) = \beta$, where k = number of errors

Number of errors	Beta risk β				
	0.01	0.05	0.10	0.25	0.50
0	4.61	3.00	2.30	1.39	0.693
1	6.64	4.74	3.89	2.69	1.68
2	8.41	6.30	5.32	3.92	2.67
3	10.0	7.75	6.68	5.11	3.67
4	11.6	9.15	7.99	6.27	4.67
5	13.1	10.5	9.27	7.42	5.67

$$\lambda = 0.1n \geqslant 3.00 \quad \text{or} \quad n \geqslant 30.$$

This is clearly inconsistent, and there is no solution for n. However, if the auditor accepts the population when there is one error or less, then from Tables 2 and 3 we see that

$$\lambda = 0.005n \leqslant 0.532 \quad \text{or} \quad n \leqslant 106.4,$$

$$\lambda = 0.1n \geqslant 4.74 \quad \text{or} \quad n \geqslant 47.4.$$

A sample of 48 will therefore suffice, with the population being rejected if there is more than one error. Of course, if we wish to accept the population with up to three errors, we see that

$$\lambda = 0.005n \leqslant 1.75 \quad \text{or} \quad n \leqslant 350,$$

$$\lambda = 0.1n \geqslant 7.75 \quad \text{or} \quad n \geqslant 77.5,$$

so that the sample size of 78 determined earlier will certainly be sufficient to control both the alpha and beta risks. An alternative possibility suggested by Kaplan is to adopt a decision rule based on beta risk only, compute the required sample size and see what this implies about alpha risk. In this case, with a sample size of 78, Eq. (5) may be used to compute an alpha risk of 0.0007.

2.2. Sample evaluation

The MUS sample size determined as set out in the previous section gives the number of individual monetary units to be audited. These units will be identified by the auditor as falling within the monetary values of certain items in the audit application; e.g., sales invoices, or accounts receivable. The whole of the items containing the selected monetary units are audited, and any audit differences noted. If no differences are discovered, then the upper error limit (UEL) (for both overstatements and understatements) may be determined by

$$\mathrm{UEL} = \frac{B \times (-\ln\beta)}{n} \times 100\%. \tag{7}$$

For a number of values of confidence $= 1-\beta$, the value of $-\ln\beta$ is given in the first row of Table 1, in the UEL columns. Also, the value $B/n = T/\lambda$ is the sample interval used in the selection of individual monetary units.

When audit differences are found, overstatements and understatements must be separately evaluated. Suppose that k overstatements are found, and that the values of these differences are $d_1, \ldots, d_k$. Suppose further that the ratios of these differences to their respective recorded book values (known as the "tainting"), *sorted in descending* order, [3] are $t_1, \ldots, t_k$. Then the gross UEL for overstatements is given by

$$\begin{aligned}\mathrm{UEL_O} = {} & \frac{B \times P_0}{n} \times 100\% + \frac{B \times (P_1 - P_0)}{n} \times t_1 + \cdots \\ & + \frac{B \times (P_k - P_{k-1})}{n} \times t_k,\end{aligned} \tag{8}$$

where P_i is given by

$$\sum_{j=0}^{i} \mathrm{e}^{-P_i} \frac{(P_i)^j}{j!} = \beta. \tag{9}$$

For a number of values of confidence $= 1-\beta$, the value of P_i is given in the UEL columns of Table 1, and the value of $P_i - P_{i-1}$ can then be calculated. [4]

Suppose, for example, that an auditor is auditing a population with a recorded book value of \$5,000,000 with a tolerable misstatement of \$500,000 and an acceptable risk of incorrect acceptance of 5%, and that based on a sample of 78 the audited values for recorded amounts of \$1,000, \$2,000 and \$3,000 are \$500, \$1,400, and \$2,400, respectively. Then the taintings in descending order

[3] A recent simulation study by Lucassen et al. [39] has shown that other bounds are also conservative; the authors of this study favor a bound in which misstatements are ranked in descending order of the magnitude of the misstatements.

[4] The P_i notation commonly used in practice is of course, simply an alternative to λ.

are 0.5, 0.3 and 0.2. From Table 1, the values of $P_i - P_{i-1}$ for 1, 2 and 3 errors, respectively are 1.748, 1.552 and 1.458. Thus Eq. (8) gives

$$\begin{aligned}\mathrm{UEL_O} &= \frac{5{,}000{,}000 \times 2.996}{78} \times 100\% + \frac{5{,}000{,}000 \times 1.748}{78} \times 0.5 \\ &\quad + \frac{5{,}000{,}000 \times 1.552}{78} \times 0.3 + \frac{5{,}000{,}000 \times 1.458}{78} \times 0.2 \\ &= \$296{,}615.\end{aligned}$$

It is apparent that each additional error increases the UEL; since the values of $P_i - P_{i-1}$ are greater than 1, each error contributes a greater amount to the increased UEL than its own value. In other words, errors diminish the precision of the sample. This effect is often referred to as precision gap widening (PGW), and the PGW columns of Table 1 indicate by how much the precision gap between the upper error level and the most likely error is increased by each successive error.

Of course, the decision to evaluate the misstatements in descending order of tainting is one reason why the Stringer bound is generally conservative. When both k overstatements and l understatements are found, the UEL for overstatements is reduced to take account of the understatements found, and the most commonly used method is to subtract from UEL the value of the most likely error for understatements

$$\mathrm{MLE_U} = \frac{1}{n} \times \sum_{1}^{l} s_l \times B, \tag{10}$$

where $s_1, \ldots, s_l$ are the understatement taintings for the l understatements.

Suppose in the example above that recorded amounts of \$1,000 and \$800 had audited values of \$1,100 and \$1,000, respectively, giving rise to taintings of 0.10 and 0.25. Based on Eq. (10), $\mathrm{MLE_U} = 1/78 \times (0.10 + 0.25) \times 5{,}000{,}000 = \$22{,}436$. Thus the net upper bound $\mathrm{UB_O}$ for overstatements given by $\mathrm{UEL_O} - \mathrm{MLE_U}$ is \$296,615 − \$22,436 = \$274,179.

A net lower bound for understatements can then be calculated by reversing the roles of the overstatement and understatement taintings in the above calculations. The resulting lower bound $\mathrm{UB_U}$ for understatements is therefore given by $\mathrm{UEL_U} - \mathrm{MLE_O}$ = \$230,013 − \$64,103 = \$165,910.

3. Consonant belief functions and statistical evidence

A belief function [20] on a frame Θ is a function Bel: $2^\Theta \rightarrow [0, 1]$ (where 2^Θ is the set of all subsets of Θ) satisfying the conditions:

1. $\mathrm{Bel}(\emptyset) = 0$;
2. $\mathrm{Bel}(\Theta) = 1$;

3. for every positive integer n and every collection $A_1 \dots A_n$ of subsets of Θ,

$$\mathrm{Bel}(A_1 \cup \cdots \cup A_n) \geqslant \sum_i \mathrm{Bel}(A_i) - \sum_{i<j} \mathrm{Bel}(A_i \cap A_j) + \cdots + (-1)^{n+1} \mathrm{Bel}(A_1 \cap \cdots \cap A_n).$$

We may then define the plausibility function as a function $\mathrm{Pl}\colon 2^{\Theta} \to [0,1]$ given by

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\sim A) \quad \text{for} A \subseteq \Theta.$$

An alternative formulation of belief functions begins with a basic probability assignment, which is a function $m : 2^{\Theta} \to [0,1]$ such that:
1. $m(\emptyset) = 0$;
2. $\sum_{A \subseteq \Theta} m(A) = 1$.

Then we may define a belief function by $\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$. Furthermore, the basic probability assignment is unique, and can be recovered from

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathrm{Bel}(B) \quad \text{for all } A \subseteq \Theta. \tag{11}$$

Belief functions allow for a simple representation of ignorance, by assigning mass to non-singleton subsets (or, of course, the frame itself). The focal elements of a belief function are subsets $A \subseteq \Theta$ such that $m(A) > 0$ (i.e., to which non-zero mass is assigned).

In modeling audit evidence, a suitable frame might contain two values such as: "accounts receivable are not materially overstated" and "accounts receivable are materially overstated." Probability theory requires that if 70% probability is assigned to "accounts receivable are not materially overstated," 30% probability is assigned to "accounts receivable are materially overstated." Suppose, however, that in the belief-function framework a 70% belief is assigned to "accounts receivable are not materially overstated." The remaining 30% belief may be assigned to "accounts receivable are materially overstated" (representing conflicting evidence), or to the set containing *both* options (representing ignorance as to which option is correct). Alternatively, part of it may be assigned to conflicting evidence (say, 10%) and the remainder (20%) will then represent ignorance. This additional flexibility (non-specificity) gives rise to the distinctive characteristics of belief functions, their claimed advantages in representing certain situations, and a variety of computational complexities that are not addressed here.

Among belief functions, *consonant* belief functions are of particular relevance to this paper. A belief function is said to be consonant if its focal elements are nested; i.e., if its focal elements can be arranged in order so that each is contained in the following one. In this situation, the different subsets of outcomes to which positive belief is assigned do not contradict each other;

rather, some are simply more precisely focused than others. This is a form of consistency in a belief system.

Given a frame Θ, and probability density functions $f_\theta : X \to [0,1]$ for each $\theta \in \Theta$, how can we build a belief function in support of the elements of Θ? Shafer [20] describes two conventions for deriving such a belief function:

1. the plausibility should be proportional to the probability; i.e., for $\theta \in \Theta, \mathrm{Pl}_x(\{\theta\}) = cf_\theta(x)$;
2. $\mathrm{Bel}_x(\cdot)$ given by $\mathrm{Bel}_x(A) = 1 - \mathrm{Pl}_x(\sim A)$ is a consonant belief function.

Under these intuitively appealing conditions, the plausibility function is uniquely defined, and the constant c is given by $1/\max_{\theta\in\Theta} f_\theta(x)$. Suppose that $f(\cdot)$ is a real valued function with:

1. $0 \leqslant f(\theta) \leqslant 1$ for all $\theta \in \Theta$;
2. $f(\theta) = 1$ for at least one $\theta \in \Theta$.

Then, [21], the relevant plausibility function is given by $\mathrm{Pl}(A) = \max_{\theta\in A} f(\theta)$. The technique may be extended to the continuous case, where it becomes $\mathrm{Pl}(A) = \sup_{\theta\in A} f(\theta)$. How should we choose the function f meeting these conditions? Srivastava and Shafer propose that we use the renormalized likelihood function (i.e., the likelihood function normalized by its maximum value). The value of this method is that the resulting plausibilities are proportional to the likelihood function; in addition, we will have increasing belief in wider intervals around the true value.

Consider the case of the Poisson distribution. The likelihood function is given by

$$L(\lambda; x) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}. \tag{12}$$

Differentiating this with respect to λ and setting the result equal to zero shows that the maximum value is reached when $\lambda = x$. The value of $L(\lambda; x)$ at this point is $\mathrm{e}^{-x}x^x/x!$, and the renormalized likelihood function becomes

$$f(A) = \frac{\sup_{\lambda\in A}(\mathrm{e}^{-\lambda}\lambda^x/x!)}{(\mathrm{e}^{-x}x^x/x!)} = \sup_{\lambda\in A}\left(\mathrm{e}^{x-\lambda}\left(\frac{\lambda}{x}\right)^x\right).$$

Thus we can define the belief function for an interval A as

$$\mathrm{Bel}_x(A) = 1 - \sup_{\lambda\notin A}\left(\mathrm{e}^{x-\lambda}\left(\frac{\lambda}{x}\right)^x\right). \tag{13}$$

Suppose that the interval A is $[\lambda_\mathrm{L}, \lambda_\mathrm{U}]$. We have already seen that $f(A)$ reaches its maximum value at x; hence, *provided that* $x \in [\lambda_\mathrm{L}, \lambda_\mathrm{U}]$, the supremum outside $[\lambda_\mathrm{L}, \lambda_\mathrm{U}]$ is reached at whichever of the endpoints of the tails has the higher value; i.e.

$$\mathrm{Bel}_x(A) = 1 - \max\left(\mathrm{e}^{x-\lambda_\mathrm{L}}\left(\frac{\lambda_\mathrm{L}}{x}\right)^x, \mathrm{e}^{x-\lambda_\mathrm{U}}\left(\frac{\lambda_\mathrm{U}}{x}\right)^x\right). \tag{14}$$

When $x \notin [\lambda_\mathrm{L}, \lambda_\mathrm{U}]$, Eq. (14) yields $\mathrm{Bel}_x([\lambda_\mathrm{L}, \lambda_\mathrm{U}]) = 0$.

When the Poisson distribution is used for MUS, bounds are estimated separately for overstatements and understatements, so that, for example, $[\lambda_\mathrm{L}, \lambda_\mathrm{U}]$ will usually be $[0, \lambda_\mathrm{U}]$, and Eq. (14) reduces to

$$\mathrm{Bel}_x(A) = 1 - \mathrm{e}^{x-\lambda_\mathrm{U}}\left(\frac{\lambda_\mathrm{U}}{x}\right)^x. \tag{15}$$

4. A belief-function approach to MUS

As described in Section 3, belief in an interval $[\lambda_\mathrm{L}, \lambda_\mathrm{U}]$ for the value of the parameter in a Poisson distribution may be given by

$$\mathrm{Bel}_x(A) = 1 - \max\left(\mathrm{e}^{x-\lambda_\mathrm{L}}\left(\frac{\lambda_\mathrm{L}}{x}\right)^x, \mathrm{e}^{x-\lambda_\mathrm{U}}\left(\frac{\lambda_\mathrm{U}}{x}\right)^x\right).$$

This provides us with a means of implementing MUS within a belief-function framework, based on the use of the Poisson distribution. The derivation of similar results for the Binomial distribution is straightforward, but is not given here. [5] As a matter of notational convenience, the subscript on the belief function Bel will be omitted in what follows.

Fig. 1 illustrates belief in the interval [1,6] given three observed errors. The belief in this interval, as described in Section 3, is equal to the complement of the plausibility of the regions outside the interval; i.e., (1 – the maximum value of the renormalized likelihood function outside the interval). However, the maximum values in the two tails are achieved at the critical values 1 and 6, and the corresponding values are 0.2737 and 0.3983. Belief in the interval [1,6] is therefore the complement of the larger value, $1 - 0.3983 = 0.6017$.

4.1. Sample size determination

As in the statistical approach described earlier, suppose that the auditor wishes to calculate the sample size n, in order to obtain a belief of at least b that an account balance with a recorded book value of B is not materially misstated (i.e., any misstatement falls within the interval $[-T, T]$), when up to k errors will be accepted. What is required is the smallest value for n such that the auditor has a belief b that the interval $[0, \mathrm{UB_O}]$ for overstatements and the interval $[0, \mathrm{UB_U}]$ for understatements are both contained within the interval

[5] It may be shown that the normalized likelihood function for the Binomial distribution is given by $f(p) = (np/k)^k((n - np)/(n - k))^{n-k}$.

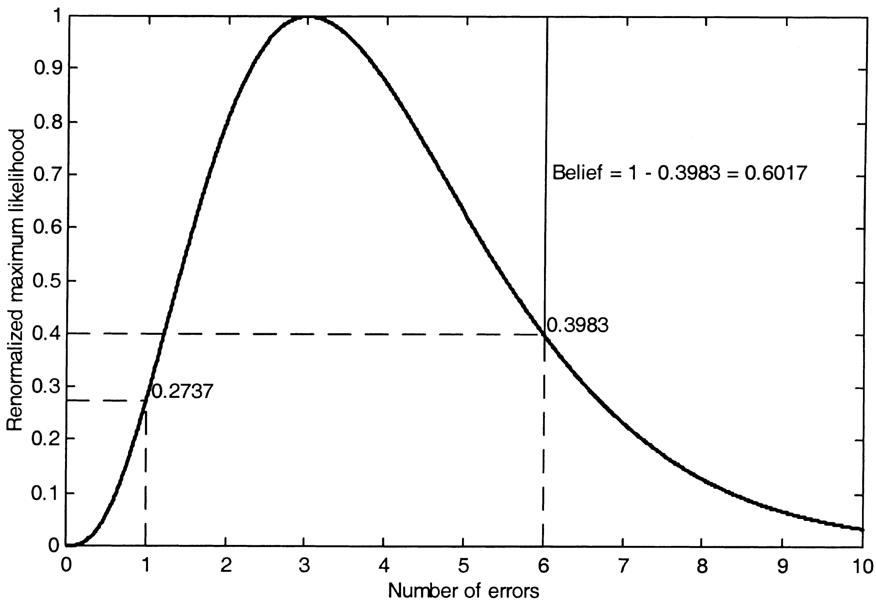

Fig. 1. Belief in intervals based on three observed errors.

[0, *T*]. As usual, MUS samples are planned based on acceptable overstatements, assuming 100% errors. The sample size may be found by solving

$$\mathrm{Bel}([\lambda_\mathrm{L}, \lambda_\mathrm{U}]) = 1 - \mathrm{e}^{k-\lambda_\mathrm{U}}\left(\frac{\lambda_\mathrm{U}}{k}\right)^k = b, \tag{16}$$

for λ_U. This gives the largest λ_U such that if up to k 100% errors are acceptable, $\mathrm{Bel}([\lambda_\mathrm{L}, \lambda_\mathrm{U}]) = b$. This is equivalent to a belief b in the interval for misstatements of $[(\lambda_\mathrm{L}/n)B, (\lambda_\mathrm{U}/n)B]$; however, we want this interval to be contained in the interval $[0, T]$, and in the limit this is achieved when $(\lambda_\mathrm{U}/n)B = T$. Thus we will have a belief of at least b in the interval $[0, T]$ for misstatements (i.e., $\mathrm{Bel}([0, T]) \geqslant b$) when $n = B\lambda_\mathrm{U}/T$.

When no errors are to be accepted, Eq. (16) may be solved explicitly to give $\lambda_U = -\ln(1-b)$, and hence

$$n = \frac{-T\ln(1-b)}{B},$$

as for the statistical procedure, and we see that in this case $b = 1 - \beta$. Determining the desired level of belief b plays the same role in the belief-function framework as controlling beta risk in the probability framework, and when no errors are anticipated the two methods give the same results.

However, when errors are to be accepted, the solution of (16) is again tedious, and Table 4 may be used for this purpose. As in Table 1, values of λ_U may be found in the columns headed UEL; the table is accessed in the row for the appropriate expected number of errors, in the column for belief b.

Consider the earlier example of a population with a recorded book value of \$5,000,000 that the auditor wishes to audit using MUS with a tolerable misstatement of \$500,000 and belief of at least 95%. Suppose that the auditor is willing to accept up to 3 of the selected dollars being in error. From Table 4, $\lambda_U = 9.432$ for three errors and a belief of 95%. The required sample size is therefore $\frac{5{,}000{,}000 \times 9.432}{500{,}000} = 94.32 \approx 95$. Of course, a 95% belief in the interval $[0, T]$, with no belief outside the interval, makes stronger epistemological demands than 95% confidence (with a 5% risk). Intuitively, then, it is reasonable that the necessary sample size be larger. The issue of what level of belief is appropriate for an auditor who would have been satisfied with a 5% beta risk is not addressed in this paper: presumably, it may well be less than 95%. Canons for acceptable levels of belief will need to be developed by audit practitioners in light of experience using belief functions.

As before, when the auditor has assessed an expected level of misstatements by value rather than an expected number of errors, we may proceed iteratively, recalling from Eq. (3) that $E = B \times k/n = Bk \times T/B\lambda = kT/\lambda$. For $k = 0, 1, 2, \ldots$, Eq. (16) may be solved for λ_U, and the result used to compute E from Eq. (3) until the required level of expected error is reached.

In the example considered earlier, suppose that the expected aggregate error was \$159,033.

For $k = 0$, Table 4 gives $\lambda_U = 2.996$, but yields a value for E of 0. For $k = 1$, Table 4 gives $\lambda_U = 5.744$, giving a value for E of \$87,047. Similarly, for $k = 2$, Table 4 gives $\lambda_U = 7.689$, giving a value for E of \$130,056. Finally, for $k = 3$, Table 4 gives $\lambda_U = 9.432$, giving a value for E of \$159,033.

If the expected error falls between k and k+1 errors, giving rise to λ_k and λ_{k+1}, respectively, producing expected errors of E_k and E_{k+1}, a conservative procedure would again be to use a sample size based on λ_{k+1}, but linear interpolation may be used.

In the above example, if the auditor expected aggregate error of \$141,647, then interpolation would give

$$\lambda = 7.689 + \frac{141{,}647 - 130{,}056}{159{,}033 - 130{,}056} \times (9.432 - 7.689) = 8.386$$

and

$$n = \frac{5{,}000{,}000 \times 8.386}{500{,}000} = 83.86 \approx 84.$$

Table 4
UEL is $B_k(b)$ such that $1 - e^{k-B_k}(B_k/k)^k = b$, where k = number of errors; PGW = $B_k(b) - B_{k-1}(b) - 1$

Number of errors	Belief b in the interval $[\lambda_L, \lambda_U]$													
	75%		80%		85%		90%		95%		98%		99%	
	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW	UEL	PGW
0	1.386		1.609		1.897		2.303		2.996		3.689		4.605	
1	3.693	1.306	3.994	1.385	4.372	1.475	4.890	1.587	5.744	1.748	6.572	1.883	7.638	2.033
2	5.357	0.664	5.706	0.712	6.141	0.768	6.729	0.839	7.689	0.945	8.608	1.036	9.779	1.141
3	6.873	0.517	7.261	0.555	7.741	0.600	8.387	0.657	9.432	0.743	10.426	0.818	11.684	0.905
4	8.312	0.438	8.732	0.471	9.251	0.510	9.946	0.559	11.066	0.634	12.125	0.699	13.458	0.774
5	9.699	0.388	10.149	0.417	10.702	0.451	11.442	0.496	12.628	0.562	13.745	0.620	15.147	0.689
6	11.051	0.351	11.527	0.378	12.111	0.409	12.891	0.450	14.139	0.510	15.309	0.564	16.773	0.626
7	12.374	0.324	12.875	0.348	13.489	0.377	14.306	0.415	15.610	0.471	16.829	0.520	18.352	0.579
8	13.676	0.302	14.200	0.324	14.840	0.352	15.693	0.387	17.049	0.439	18.315	0.486	19.892	0.540
9	14.959	0.283	15.505	0.305	16.171	0.331	17.056	0.364	18.462	0.413	19.773	0.457	21.401	0.509
10	16.228	0.268	16.794	0.289	17.484	0.313	18.401	0.344	19.854	0.392	21.206	0.433	22.884	0.482
11	17.483	0.255	18.068	0.275	18.782	0.298	19.728	0.328	21.227	0.373	22.619	0.413	24.343	0.460
12	18.727	0.244	19.331	0.263	20.067	0.285	21.042	0.314	22.583	0.357	24.013	0.395	25.783	0.440
13	19.961	0.234	20.583	0.252	21.341	0.274	22.343	0.301	23.926	0.342	25.392	0.379	27.205	0.422
14	21.187	0.225	21.826	0.243	22.604	0.263	23.633	0.290	25.256	0.330	26.758	0.365	28.612	0.407
15	22.404	0.218	23.061	0.234	23.858	0.254	24.912	0.280	26.574	0.318	28.110	0.352	30.005	0.393
16	23.615	0.211	24.287	0.227	25.104	0.246	26.183	0.271	27.882	0.308	29.451	0.341	31.385	0.380
17	24.819	0.204	25.507	0.220	26.343	0.238	27.445	0.262	29.181	0.299	30.782	0.331	32.754	0.369
18	26.017	0.198	26.721	0.213	27.574	0.232	28.700	0.255	30.471	0.290	32.104	0.321	34.112	0.358
19	27.210	0.193	27.928	0.208	28.800	0.225	29.948	0.248	31.753	0.282	33.416	0.313	35.461	0.349

20	28.398	0.188	29.131	0.202	30.019	0.220	31.190	0.242	33.029	0.275	34.721	0.305	36.801	0.340
21	29.581	0.183	30.328	0.197	31.233	0.214	32.425	0.236	34.297	0.268	36.019	0.297	38.133	0.332
22	30.760	0.179	31.521	0.193	32.443	0.209	33.656	0.230	35.559	0.262	37.309	0.291	39.457	0.324
23	31.935	0.175	32.709	0.188	33.647	0.204	34.881	0.225	36.816	0.256	38.593	0.284	40.774	0.317
24	33.107	0.171	33.894	0.184	34.847	0.200	36.101	0.220	38.066	0.251	39.871	0.278	42.084	0.310
25	34.274	0.168	35.074	0.181	36.043	0.196	37.317	0.216	39.312	0.246	41.144	0.272	43.388	0.304
26	35.439	0.164	36.251	0.177	37.235	0.192	38.528	0.212	40.553	0.241	42.411	0.267	44.686	0.298
27	36.600	0.161	37.425	0.174	38.424	0.189	39.736	0.208	41.790	0.236	43.673	0.262	45.978	0.292
28	37.758	0.158	38.596	0.171	39.609	0.185	40.940	0.204	43.022	0.232	44.930	0.257	47.265	0.287
29	38.914	0.156	39.763	0.168	40.791	0.182	42.140	0.200	44.250	0.228	46.183	0.253	48.547	0.282
30	40.067	0.153	40.928	0.165	41.969	0.179	43.337	0.197	45.474	0.224	47.432	0.249	49.824	0.277
35	45.796	0.146	46.713	0.157	47.822	0.170	49.275	0.188	51.543	0.214	53.617	0.237	56.147	0.265
40	51.475	0.136	52.444	0.146	53.615	0.159	55.149	0.175	57.539	0.199	59.721	0.221	62.379	0.246
45	57.113	0.128	58.131	0.137	59.361	0.149	60.970	0.164	63.475	0.187	65.759	0.208	68.538	0.232
50	62.716	0.121	63.781	0.130	65.066	0.141	66.747	0.155	69.360	0.177	71.740	0.196	74.633	0.219
55	68.290	0.115	69.399	0.124	70.737	0.134	72.486	0.148	75.203	0.168	77.675	0.187	80.676	0.209
60	73.838	0.110	74.990	0.118	76.379	0.128	78.192	0.141	81.008	0.161	83.567	0.179	86.673	0.199
65	79.364	0.105	80.557	0.113	81.994	0.123	83.869	0.135	86.780	0.154	89.424	0.171	92.629	0.191
70	84.870	0.101	86.102	0.109	87.586	0.118	89.521	0.130	92.523	0.149	95.248	0.165	98.550	0.184
75	90.359	0.098	91.629	0.105	93.157	0.114	95.150	0.126	98.241	0.143	101.044	0.159	104.438	0.178
80	95.831	0.094	97.138	0.102	98.709	0.111	100.759	0.122	103.934	0.139	106.813	0.154	110.298	0.172
85	101.289	0.092	102.631	0.099	104.245	0.107	106.349	0.118	109.607	0.135	112.559	0.149	116.130	0.167
90	106.734	0.089	108.110	0.096	109.765	0.104	111.922	0.115	115.260	0.131	118.284	0.145	121.940	0.162
95	112.167	0.087	113.576	0.093	115.271	0.101	117.479	0.111	120.895	0.127	123.989	0.141	127.727	0.157
100	117.588	0.084	119.030	0.091	120.764	0.099	123.021	0.109	126.514	0.124	129.675	0.137	133.493	0.153

This calculation for \$141,647 expected aggregate error interpolates based on 2.4 errors; in the belief-function case, Eq. (16) may be solved directly to give a more accurate estimate for λ of 8.403, leading to a sample size of

$$n = \frac{5{,}000{,}000 \times 8.403}{500{,}000} = 84.03 \approx 85.$$

4.1.1. Controlling alpha risk

The methods described above for determining sample sizes for MUS do not explicitly control the risk of incorrect rejection. Suppose that the auditor determines some low proportion of error, μ, such that it would be undesirable to reject a population with this error too often. Suppose, in fact, that the auditor wishes to have a belief a that the interval $[\lambda_L, \lambda_U]$ marginally contains μn. Then the sample size is determined by finding the smallest value n such that

$$1 - e^{k-\lambda_L}\left(\frac{\lambda_L}{k}\right)^k \geqslant a, \qquad \text{where } \lambda_L = \mu n \tag{17}$$

and simultaneously

$$1 - e^{k-\lambda_U}\left(\frac{\lambda_U}{k}\right)^k \leqslant b, \qquad \text{where } \lambda_U = \frac{nT}{B}. \tag{18}$$

Eq. (17) finds the smallest value of λ_L for which the belief is not more than a, if λ_L were any smaller, the belief would be less than a. At the same time, Eq. (18) finds the largest value of λ_U for which belief does not exceed b. Tables 5 and 6 are provided to facilitate the necessary calculations.

Fig. 2 illustrates a belief 0.90 that rejected populations contain $\mu n = 0.632$ errors, and a belief 0.95 that accepted populations contain no more than $nT/B = 9.432$ errors, given three observed errors.

Suppose, for example, that the auditor wishes to audit a balance of \$5,000,000 with a 90% belief that there are at least 0.5% errors, and a 95% belief that the total misstatement does not exceed \$500,000. If the auditor

Table 5
$\lambda_L(a,k)$ such that $1 - e^{k-\lambda_L}(\lambda_L/k)^k = a$, where k = number of errors

Number of errors	Belief a in the interval (λ, ∞)				
	0.50	0.75	0.90	0.95	0.99
0	–	–	–	–	–
1	0.232	0.102	0.038	0.019	0.004
2	0.761	0.464	0.266	0.180	0.076
3	1.394	0.956	0.632	0.477	0.259
4	2.083	1.522	1.085	0.863	0.531
5	2.808	2.138	1.597	1.314	0.872

Table 6
$\lambda_U(b,k)$ such that $1 - e^{k-\lambda_U}(\lambda_U/k)^k = b$, where k = number of errors

Number of errors	Belief b in the interval $[0,\lambda]$				
	0.50	0.75	0.90	0.95	0.99
0	0.693	1.386	2.303	2.996	4.605
1	2.678	3.693	4.890	5.744	7.638
2	4.156	5.357	6.729	7.689	9.779
3	5.525	6.873	8.387	9.432	11.684
4	6.838	8.312	9.946	11.066	13.458
5	8.114	9.699	11.442	12.628	15.147

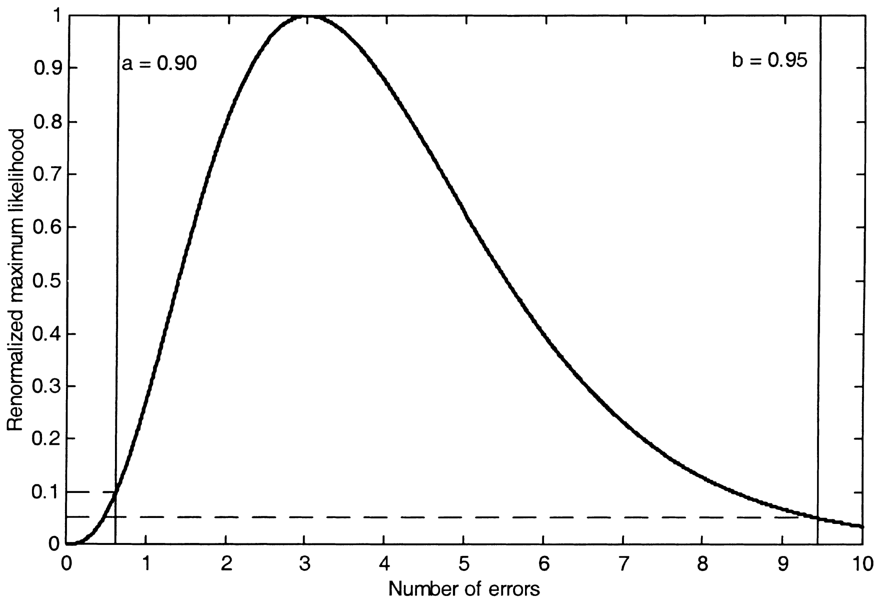

Fig. 2. Beliefs a and b in intervals based on three observed errors.

accepts the population only when there is one error or less, then from Tables 5 and 6 we see that

$$\lambda_L = 0.005n \leqslant 0.038 \quad \text{or} \quad n \leqslant 7.6$$

and

$$\lambda_U = 0.1n \geqslant 5.744 \quad \text{or} \quad n \geqslant 57.44.$$

This is clearly inconsistent. If the auditor accepts the population when there are two errors or less, we see from Tables 5 and 6 that

$$\lambda_{\mathrm{L}} = 0.005n \leqslant 0.266 \quad \text{or} \quad n \leqslant 53.2,$$

$$\lambda_{\mathrm{U}} = 0.1n \geqslant 7.689 \quad \text{or} \quad n \geqslant 76.89.$$

This is still not consistent. However, if the auditor is willing to accept up to three errors, we see that

$$\lambda_{\mathrm{L}} = 0.005n \leqslant 0.632 \quad \text{or} \quad n \leqslant 126.4,$$

$$\lambda_{\mathrm{U}} = 0.1n \geqslant 9.432 \quad \text{or} \quad n \geqslant 94.32,$$

so that the sample size of 95 determined earlier will certainly be sufficient to achieve the desired beliefs. Note that Table 5 cannot provide a lower limit when the expected number of errors is zero.

4.2. Sample evaluation

Within the belief-function framework, MUS samples may be evaluated as in the statistical case, but using Table 4. If no differences are discovered, then the UEL (for both overstatements and understatements) may be determined by

$$\mathrm{UEL} = \frac{B \times (-\ln(1-b))}{n} \times 100\%. \tag{19}$$

For a number of levels of belief b, the value of $-\ln(1-b)$ is given in the first row of Table 4, in the UEL columns.

As before, when audit differences are found, overstatements and understatements must be evaluated separately. Suppose that k overstatements are found, and that the values of these differences are $d_1, \ldots, d_k$. Suppose also that the taintings, sorted in descending order, are $t_1, \ldots, t_k$. Then the gross UEL for overstatements is given by

$$\begin{aligned} \mathrm{UEL_O} = {} & \frac{B \times B_0}{n} \times 100\% + \frac{B \times (B_1 - B_0)}{n} \times t_1 \\ & + \cdots + \frac{B \times (B_k - B_{k-1})}{n} \times t_k, \end{aligned} \tag{20}$$

where B_i is given by

$$\mathrm{e}^{k-B_i} \left(\frac{B_i}{k} \right)^k = 1 - b.$$

For a number of levels of belief b, the value of B_i is given in the UEL columns of Table 4, and the value of $B_i - B_{i-1}$ can then be calculated.

Suppose, as before, that an auditor is auditing a population with a recorded book value of \$5,000,000 with a tolerable misstatement of \$500,000, and desires a 95% belief in the interval [0, 500,000]; suppose also that based on a sample of 78 the audited values for recorded amounts of \$1,000, \$2,000 and \$3,000 are \$500, \$1,400, \$2,400, respectively. Then the taintings in descending order are 0.5, 0.3 and 0.2. From Table 4, the values of $B_i - B_{i-1}$ for 1, 2 and 3 errors, respectively, are 2.748, 1.945 and 1.743. Thus Eq. (20) gives

$$\begin{aligned} \mathrm{UEL_O} &= \frac{5{,}000{,}000 \times 2.996}{78} \times 100\% + \frac{5{,}000{,}000 \times 2.748}{78} \times 0.5 \\ &\quad + \frac{5{,}000{,}000 \times 1.945}{78} \times 0.3 + \frac{5{,}000{,}000 \times 1.743}{78} \times 0.2 \\ &= \$339{,}878. \end{aligned}$$

Note that, consistent with the view that a belief of 95% is a stronger claim than 95% confidence, the UEL is larger than for the probabilistic analysis given earlier. Table 4 also provides belief-function equivalents of the precision gap wideners.

When both k overstatements and l understatements are found, the UEL for overstatements may be reduced as before to take account of the understatements by subtracting from UEL the value of

$$\mathrm{MLE_U} = \frac{1}{n} \times \sum_{1}^{l} s_l \times B,$$

where $s_1, \ldots, s_l$ are the understatement taintings for the l understatements. The auditor then has a belief b that misstatements fall in the interval $[0, \mathrm{UB_O}]$, where $\mathrm{UB_O} = \mathrm{UEL_O} - \mathrm{MLE_U}$.

Suppose in the example above that recorded amounts of \$1,000 and \$800 had audited values of \$1,100 and \$1,000, respectively, giving rise to tainting of 0.10 and 0.25. Then, based on Eq. (10)

$$\mathrm{MLE_U} = \frac{1}{78} \times (0.10 + 0.25) \times 5{,}000{,}000 = \$22{,}436.$$

Thus the net upper bound for overstatements given by

$\mathrm{UEL_O} - \mathrm{MLE_U}$ is \$339,878 – \$22,436 = \$317,442.

Although this method of recognizing the net effect on $\mathrm{UEL_O}$ of understatements discovered in the sample is intuitively appealing here, as in the statistical case, it is not based on a theoretical analysis, and there are at present no empirical studies supporting its use in the belief-function case. The audit conclusion from this sampling procedure is, therefore, that the auditor has a belief at least 0.95 that the overstatements fall in the interval [0, 317, 442]; this is within the original desired interval, based on planning materiality.

An upper bound for understatements may similarly be obtained, giving the result that the auditor has a belief at least 0.95 that the understatements fall in the interval [0, 184, 455]. Since both the overstatement and understatement intervals are contained within the tolerable interval for misstatements of [0, 500,000], the auditor has a belief at least 0.95 that the account is not materially misstated.

In general, provided that the intervals $[0, \mathrm{UB_O}]$ and $[0, \mathrm{UB_U}]$ both fall within the tolerable misstatement interval $[0, T]$, the auditor will have a belief b that the account is not materially misstated, and a zero belief that the account is misstated. If either the overstatement or the understatement interval is not contained within $[0, T]$, however, the auditor may wish to perform additional audit procedures and accept a lower level of belief, to increase the sample size and re-evaluate the sample, or to propose an adjustment based on the errors discovered, so that the revised intervals $[0, \mathrm{UB_O}]$ and $[0, \mathrm{UB_U}]$ both become acceptable.

As in the probability theory case, each additional error increases the UEL, and each error contributes a greater amount to the increased UEL than its own value. The PGW columns of Table 4 indicate by how much the precision gap (between the upper error level used in the belief interval and the most likely error) is increased by each successive error.

5. Summary and conclusions

This paper has described several methods of MUS, and demonstrated how these methods may be used in the context of a belief-function framework. Formulae and Tables for planning and evaluating MUS within belief functions are provided. The methods are illustrated with reference to an auditing example, showing how a sample size is calculated to obtain a given level of belief in a planned interval for misstatements $[0, T]$, and how sample results may be evaluated to give a desired level of belief in an achieved misstatement interval $[0, \mathrm{UB}]$. Consistent with intuition, sample sizes for a given level of belief are somewhat larger than (though comparable with) those for the same numerical level of statistical confidence. For a fixed sample size, higher upper bounds are obtained using beliefs than using the same level of statistical confidence. While the tables are correspondingly different, application of the method is procedurally identical to the normal statistical method. Statistical evidence in the form of belief in intervals may be integrated with non-statistical evidence using Dempster's Rule for the combination of beliefs in the normal way.

Future simulation work is desirable to support the use of the MLE adjustment to UEL in the belief-function case. Although details are not given of extensions to the use of the Binomial distribution in MUS, this extension is straightforward; it may be best considered, however, in the context of an extension of this paper to cover the cell-selection methods that some authors favor [30,31]. Finally, future research is needed on the use of alternatives to the Stringer bound within the belief-function framework.

Acknowledgements

Support for this research was provided by the Ronald G. Harper Doctoral Fellowship and by the Ernst and Young Center for Auditing Research and Advanced Technology. Comments provided by Rajendra P. Srivastava, Keith Harrison, Margaret Reed, Paul van Batenburg and Glenn Shafer are gratefully acknowledged.

References

[1] G. Shafer, R.P. Srivastava, The Bayesian and belief-function formalisms: a general perspective for auditing, Auditing: A Journal of Practice and Theory 9 (1990) 110–137.
[2] American Accounting Association, A Statement of Basic Auditing Concepts, American Accounting Association, Sarasota, FL, 1973.
[3] A.A. Arens, J.K. Loebbecke, Auditing: An Integrated Approach, sixth ed., Prentice-Hall, Englewood Cliffs, NJ, 1994.
[4] American Institute of Certified Public Accountants, Statement on Auditing Standards No. 31, Evidential Matter, 1980.
[5] American Institute of Certified Public Accountants, Statement on Auditing Standards No. 39, Audit Sampling, 1981.
[6] American Institute of Certified Public Accountants, Statement on Auditing Standards No. 47, Audit Risk and Materiality in Conducting an Audit, 1983.
[7] B.E. Cushing, J.K. Loebbecke, Analytical approaches to audit risk: a survey and analysis, Auditing: A Journal of Practice and Theory 3 (2) (1983) 23–41.
[8] W.R. Kinney, A note on compounding probabilities in auditing, Auditing: A Journal of Practice and Theory 2 (2) (1983) 13–22.
[9] J. Jiambalvo, W. Waller, Decomposition and assessments of audit risk, Auditing: A Journal of Practice and Theory 3 (2) (1984) 80–88.
[10] D.A. Leslie, An analysis of the audit framework focusing on inherent risk and the role of statistical sampling in compliance testing, in: Research on Auditing Symposium VII, University of Kansas, 1984, pp. 89–125.
[11] W.R. Kinney, Achieved audit risk and the audit outcome space, Auditing: A Journal of Practice and Theory 8 (1989) 67–84.
[12] J.A. Sennetti, Toward a more consistent model for audit risk, Auditing: A Journal of Practice and Theory 9 (1990) 103–112.
[13] R.P. Srivastava, G. Shafer, Belief-function formulas for audit risk, The Accounting Review 67 (1992) 249–283.
[14] R.P. Srivastava, Belief functions and audit decisions, Auditor's Report 17 (1993) 8–12.
[15] R.P. Srivastava, The belief-function approach to aggregating audit evidence, International Journal of Intelligent Systems 10 (1995) 329–356.

[16] R.P. Srivastava, A general scheme for aggregating evidence in auditing: propagation of beliefs in networks, in: M.A. Vasarhelyi (Ed.), Artificial Intelligence in Accounting and Auditing, vol. 3, Markus Wiener Publishers, Princeton, 1995.
[17] R.P. Srivastava, et al., Belief propagation in networks for auditing, Presented at Audit Judgment Symposium, University of Southern California, 1990.
[18] A.D. Akresh, J.K. Loebbecke, Audit approaches and techniques, in: A.R. Abdel-khalik, I. Solomon (Eds.), Research Opportunities in Auditing: The Second Decade, American Accounting Association, Sarasota, FL, 1988.
[19] P.R. Gillett, Automated dynamic audit programme tailoring: an expert system approach, Auditing: A Journal of Practice and Theory 12 (1993) 173–189.
[20] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[21] R.P. Srivastava, G. Shafer, Integrating statistical and non-statistical audit evidence using belief functions: a case of variable sampling, International Journal of Intelligent Systems 9 (1994) 519–539.
[22] R.S. Kaplan, Sample size computations for dollar unit sampling, Journal of Accounting Research (1975) 238–258.
[23] W.G. Cochran, Sampling Techniques, Wiley, New York, 1977.
[24] A. Bailey, Statistical Auditing: Review, Concepts and Problems, Harcourt Brace Jovanovich, New York, 1981.
[25] D.M. Guy, et al., Audit Sampling, third ed., Wiley, New York, 1993.
[26] W.E. Deming, Sample Design in Business Research, Wiley, New York, 1960.
[27] A. van Heerden, Statistical sampling as a means of auditing, Maandblad voor Accountancy en Bedrijfshuishoudkunde 35 (1961) 453–475.
[28] K.W. Stringer, Practical aspects of statistical sampling in auditing, in: Proceedings of the Business and Economic Statistics Section, American Statistical Association, 1963, pp. 405–411.
[29] R.J. Anderson, A.D. Teitlebaum, Dollar-unit sampling: a solution to the audit sampling dilemma, CA Magazine, 1973.
[30] A.D. Teitlebaum, Dollar-unit sampling in auditing, Presented at National Meeting of the American Statistical Association, Montreal, 1973.
[31] D.A. Leslie, et al., Dollar Unit Sampling: a Practical Guide for Auditors, Copp Clark Pitman, Toronto, 1979.
[32] American Institute of Certified Public Accountants, Audit and Accounting Guide, Audit Sampling, 1983.
[33] G. Pap, M.C.A. van Zuijlen, On the asymptotic behavior of the stringer bound, Statistica Neerlandica, Forthcoming, 1996.
[34] L. Dworin, R.A. Grimlund, Dollar unit sampling for accounts receivable and inventory, The Accounting Review 59 (1984) 218–241.
[35] W.L. Felix, et al., Arthur Andersen's new monetary unit sampling approach, Auditing: A Journal of Practice and Theory 9 (3) (1990) 1–16.
[36] K. Tsui, et al., Multinomial-Dirichlet bounds for dollar-unit sampling in auditing, The Accounting Review (1985) 76–96.
[37] M. Menzefricke, W. Smieliauskas, A simulation study of the performance of parametric dollar unit sampling statistical procedures, Journal of Accounting Research (1984) 588–604.
[38] J. Neter, J. Godfrey, Robust Bayesian bounds for monetary-unit sampling in auditing, Journal of Royal Statistical Society (1985) 157–168.
[39] A. Lucassen, et al., Modifications of the Stringer-bound: a simulation study on the performance of audit sampling evaluation methods, Kwantitatieve Methoden 51 (1996) 17–25.