Encyclopedia of Computer Science and Technology, Vol. 37, edited by Allen Kent, James G. Williams, and Carolyn M. Hall, published by Marcel Dekker, Inc., New York, .Supplement 22, 1997, pp. 157-174.

Integrating Statistical and Non-Statistical Evidence Using Belief Functions

by

Rajendra P. Srivastava*
Ernst & Young Professor
Director, Ernst & Young Center for Auditing Research and Advanced Technology
Division of Accounting and Information Systems
School of Business, The University of Kansas
Lawrence, KS 66045

for

Encyclopedia of Computer Science and Technology

June 1995

*The author would like to thank Peter R. Gillett for his valuable comments on the article.

Integrating Statistical and Non-Statistical Evidence Using Belief Functions

ABSTRACT

This article demonstrates how one can integrate statistical evidence with non-statistical evidence, especially in auditing, in an objective way using belief functions. In particular, the article focuses on: (1) determining the sample size in a statistical test for a desired level of belief; (2) determining the achieved level of belief for a given sample result; and (3) integrating statistical and non-statistical items of evidence. We describe the integration process for two different sampling techniques: variable sampling and attribute sampling. Numerical examples from auditing are used for illustration purposes. The article uses Srivastava and Shafer (1994) for variable sampling, and Srivastava and Gillett (1995) for attribute sampling.

Key Words: Audit Judgment, Belief Functions, Non-Statistical Evidence, Statistical Evidence, Variable Sampling, Attribute Sampling

Integrating Statistical and Non-Statistical Evidence Using Belief Functions

TABLE OF CONTENTS

Integrating Statistical and Non-Statistical Evidence Using Belief Functions

1. INTRODUCTION

The main purpose of this article is to discuss a method for integrating statistical and non-statistical items of evidence using belief functions. In particular, the article focuses on (1) determining the sample size in a statistical test for a desired level of belief; (2) determining the achieved level of belief for a given sample result; and (3) integrating statistical and non-statistical items of evidence. There are three issues in the integration process. First, we need to express the strength of non-statistical evidence in terms of belief functions. This process is relatively straightforward, and we will discuss it in Section 2 in detail. Second, we need to express the strength of statistical evidence in terms of belief functions. This has been discussed by Shafer (1976) and further elaborated by Srivastava and Shafer (1994) for variable sampling. We will use the Srivastava and Shafer approach to express the strength of statistical evidence in terms belief functions for variable sampling, and the Srivastava and Gillett (1995) approach for attribute sampling. The third issue deals with the integration process. Once the strengths of various items of evidence are expressed in terms of belief functions, it becomes a trivial exercise to combine all the evidence using Dempster's rule (see Section 2 for details).

In this article, we use examples from auditing to illustrate the concepts. The main goal of the auditor in an audit[1] of the financial statements of a company is to give an opinion on whether the financial statements present fairly the financial position of the company. In order to form an opinion, the auditor collects, evaluates, and integrates various items of evidence related to various accounts on the balance sheet and on the income statement. In general, the audit evidence consists of items of evidence that are statistical or non-statistical in nature. Moreover, some items of evidence bear on individual accounts or audit objectives of the account while others bear on the entire financial statements (e.g., see Srivastava and Shafer 1992, and Shafer and Srivastava 1990).

An example of non-statistical evidence is the level of assurance provided by analytical procedures (e.g., ratio analyses, comparison of the current account balance with the previous years' balances, etc.)

that the inventory account balance is not materially misstated. The auditor evaluates this evidence in light of what he or she knows about the strength of the analytical procedures and relevance of the data used and makes a judgment about the level of assurance it provides for the assertion that the inventory account balance is not materially misstated. An example of statistical evidence is the physical count and valuation of a sample of randomly selected inventory items for determining the total value of the inventory. The auditor integrates the level of assurance obtained from the non-statistical evidence with the level of assurance obtained from the statistical evidence to determine the total assurance that the inventory account is not materially misstated. Our objective in this article is to discuss the belief-function approach for such an integration. Srivastava and Shafer (1994) contend that such an objective way to combine statistical and non-statistical evidence will make the audit process more efficient.

The remainder of this article is divided into five sections. In Section 2, we discuss the basics of belief functions and Dempster's rule of combination. In Section 3, we outline the standard statistical approach to sampling for audit decisions using the mean-per-unit method. Also, we discuss the relationship between a given level of confidence in a decision interval with the corresponding level of belief in the interval. We derive a formula for the sample size for a desired level of belief in the interval, and show how to determine the achieved level of belief for a given sample result. In Section 4, we illustrate how to integrate non-statistical evidence with statistical evidence under variable sampling. In Section 5, we discuss the integration of non-statistical evidence with statistical evidence under attribute sampling. Finally, in Section 6, we provide a summary of our results.

2. THE BASICS OF BELIEF FUNCTIONS

The current primary form of belief functions,[2] known as Dempster-Shafer theory of belief functions, is the work of Dempster in the 1960's and Shafer in the 1970's. In this section, we describe the basics of belief functions. For comprehensive coverage, see Shafer (1976).

The main difference between probability theory and belief functions is in the method of assignment of uncertainty to a set of mutually exclusive and exhaustive states or assertions of interest. We will call this set the *frame* and represent it by the symbol Θ. In probability theory, we assign uncertainty to each individual element of the frame and call the assigned value the probability of occurrence of the element. All these probabilities must add to one. Consider the following auditing example. The inventory account balance is not materially misstated, 'a', and it is materially misstated, '~a', are the two states representing a mutually exclusive and exhaustive set. Here the frame consists of the two elements: $\Theta = \{a, \sim a\}$. In probability theory, we assign probability to each element of the frame[3], i.e., $P(a) = 0$, and $P(\sim a) = 0$. Also, we know that $P(a) + P(\sim a) = 1$.

2.1. m-values

In belief functions, uncertainty is not only assigned to the single elements of the frame but also to all other proper subsets of the frame and to the entire frame. We call these uncertainties m-values. Similar to probabilities, all these m-values must add to one. For the example considered above[4], we will have $m(\{a\}) = 0$, $m(\{\sim a\}) = 0$, and $m(\{a, \sim a\}) = 0$, and, $m(\{a\}) + m(\{\sim a\}) + m(\{a, \sim a\}) = 1$. Let us assume that the auditor has performed analytical procedures relevant to the inventory account balance and finds no significant difference between the recorded value and the predicted value. Based on this finding, the auditor feels that the recorded value appears reasonable and is not materially misstated. However, he or she does not want to put too much weight on this evidence. Assume that the auditor assigns a small level of assurance, say 0.3, that the inventory account is not materially misstated. We can express this feeling in terms of m-values as: $m(\{a\}) = 0.3$, $m(\{\sim a\}) = 0$, and $m(\{a, \sim a\}) = 0.7$. The belief function interpretation of these m-values is that the auditor has 0.3 level of support for 'a', no support for '~a', and 0.7 level of support remains uncommitted, which represents a measure of ignorance.

However, if we had to express the above feelings in terms of probabilities, then we get into problems, because we will assign $P(a) = 0.3$ and $P(\sim a) = 0.7$. This which implies that there is a 70 percent chance that the inventory account is materially misstated, which we know is not what the auditor is trying to say. The auditor has obtained no evidence supporting a belief that the inventory account is materially misstated. Thus, we can use m-values to better express the basic judgment about the level of support or assurance the auditor obtains from an item of evidence for an assertion.

2.2. Belief Functions

The belief function, Bel(A) for a subset A of elements of the frame represents the total belief in A. This belief will be more than m(A). Actually, Bel(A) is equal to m(A) plus sum of all the m-values for the sets of elements that are contained in A. In terms of symbols:

$$\mathrm{Bel(A)} = \sum_{\mathrm{B} \subseteq \mathrm{A}} \mathrm{m(B)}\,. \tag{1}$$

By definition, belief in the empty set is zero.

Let us consider an example to illustrate the definition of belief functions. Suppose you have a friend who lives on the East Coast in the New Jersey area. The only contact you have with him is through greeting cards that he sends you periodically with no return address. You want to find the belief that your friend lives in New Jersey. After looking through all the cards you have received over the years, you can identify the following post-office seals marked on the cards: 10% of the cards are marked North Brunswick, 25% East Brunswick, 10% Philadelphia, and 15% Newark. Thirty percent of the cards have only the Brunswick part legible which means you cannot determine from what part of Brunswick the card was mailed. For the remaining 10 percent, nothing is legible on the seals. These numbers can be interpreted as non-zero m-values for different subsets of the frame that your friend lives somewhere on the East Coast near New Jersey. Based on just this evidence, you wish to form your total belief that the friend lives in New Jersey. This belief will be the sum of the m-values that he lives in North Brunswick, East Brunswick, Brunswick, and Newark. For this example, the belief is 0.80 (10% North Brunswick, 25% East Brunswick, 15% Newark, 30% Brunswick area). Similarly, the belief that the friend lives in Brunswick, which includes North Brunswick and East Brunswick, will be 0.65 (10% North Brunswick, 25% East Brunswick, 30% Brunswick).

Going back to our auditing example of analytical procedures, the auditor's assessment of the level of support in terms of m-values was: m({a}) = 0.3, m({~a}) = 0, and m({a, ~a}) = 0.7. Based on analytical procedures alone, the belief that the account is not materially misstated is 0.3 (i.e., Bel({a}) = 0.3) and there is no support for the belief that the account is materially misstated (Bel({~a}) = 0). In general, a zero belief in the belief-function formalism means that there is no evidence to support the proposition. In other words, a zero belief in a proposition represents lack of evidence. In contrast, a zero probability in probability theory means that the proposition cannot be true, which represents an impossibility. Also, one finds that beliefs for 'a' and '~a' do not necessarily add to one, i.e., Bel({a}) + Bel({~a}) = 1, whereas in probability it is always true that P(a) + P(~a) = 1.

Belief functions differ from probabilities in representing ignorance. In probability theory, we represent ignorance by assigning equal probability to all the outcomes or elements of the frame. In the belief-function framework, we represent ignorance by assigning an m-value of one to the entire frame and an m-value of zero to all its proper subsets. The belief-function formalism becomes the Bayesian formalism when non-zero m-values exist only for single elements of the frame. In such a case, m-values become probabilities, i.e., m({a_i}) = P(a_i), and Dempster's rule in the belief-function formalism becomes Bayes' rule in the probability theory (Shafer 1976).

2.3. Plausibility Functions

The plausibility function, Pl(.), is an interesting function, especially for auditing as discussed by Srivastava and Shafer (1992). It presents the non-frequentist's interpretation of audit risk. By definition, the plausibility of A is equal to one minus the belief in ~A, i.e., Pl(A) = 1 - Bel(~A) where ~A represents the set of elements that are not in A. Intuitively, the plausibility of A is the degree to which A is plausible given the evidence. In other words, Pl(A) is the degree to which we do not assign belief to its negation, ~A.

In our example of analytical procedures, we have Bel({a}) = 0.3, Bel({~a}) = 0. These values yield the following plausibility values: Pl({a}) = 1, and Pl({~a}) = 0.7. Pl({a}) = 1 indicates that 'a' is maximally plausible since we have no evidence against it. However, Pl({~a}) = 0.7 indicates that if we had no other items of evidence to consider, then the maximum possible risk that the inventory account is materially misstated would be 0.7, even though we have no evidence that the account is materially misstated (Bel({~a}) = 0). The plausibility function for the assertion that the account is materially misstated is the belief-function interpretation of the audit risk associated with the evidence. In the present example, one can interpret Pl({~a}) = 0.7 as the analytical procedure risk in the audit risk model of SAS 47.

2.4. Dempster's Rule of Combination

Dempster's rule (see, e.g., Shafer 1976) is used to combine various independent items of evidence in the belief-function framework. In the case of two items of evidence bearing on a frame Θ with m_1 and m_2, respectively, being the two sets of m-values associated with each item of evidence, Dempster's rule yields the following set for the combined m-values:

$$m(A) = K^{-1}\Sigma\{m_1(B_1)m_2(B_2)|B_1 \cap B_2 = A, A? \varnothing\} \quad (2)$$

where m(A) is the resultant m-value for the subset A of the frame Θ, and K is the renormalization constant given as:

$$K = 1 - \Sigma\{m_1(B_1)m_2(B_2)|B_1 \cap B_2 = \varnothing\}. \quad (3)$$

The second term in K represents the conflict between the two items of evidence. When the two items of evidence totally conflict with each other, that is, when K = 0, these items of evidence are not combinable. For Dempster's rule with more than two items of evidence, see Shafer (1976).

3. AUDIT DECISIONS WITH STATISTICAL EVIDENCE USING VARIABLE SAMPLING

As mentioned earlier, the auditor uses statistical evidence on every engagement to determine whether an account balance is fairly stated, i.e., it is not materially misstated. For example, in determining whether the inventory account balance is not materially misstated, the auditor selects a sample of inventory items, physically counts them and determines their values from which he or she projects the total value of the inventory. In general, there are two types of decisions the auditor makes. First, the auditor determines the extent of the statistical test, i.e., the sample size, n, for a desired level of confidence. Second, the auditor evaluates the sample results to determine whether to accept or reject the recorded account balance to be fairly stated. There are several statistical approaches, such as, mean-per-unit estimator, difference estimator, ratio estimator, etc., used in auditing for determining whether the recorded account balance is not materially misstated (e.g., see Bailey 1981; and Arens and

Loebbecke 1981). We will use the mean-per-unit estimator to illustrate the decision process and compare it with the belief-function approach.

3.1. Standard Statistical Approach

Consider the audit of an inventory account. We know that the population of the values of inventory items is not usually normally distributed. However, based on the *central limit* theorem, we can assume the sample mean to be normally distributed with the unknown true mean μ_o, provided a sufficiently large sample is taken. For simplicity, we assume that we know the standard deviation σ of the population.

The true, but unknown, mean μ_o is the mean that the auditor would find if he or she were able to audit the entire population accurately. However, because of resource limitations and usually a large population size, the auditor is not able to perform the audit procedure on the entire population. Rather, he or she performs the audit procedure on a sample of inventory items and estimates the true mean μ_o from the sample results. The auditor accepts the recorded mean, μ_r, to be fairly stated if it is within the maximum tolerable error, TE, per item (the maximum error regarded as material) from the true mean μ_o, and rejects it otherwise as being materially misstated. In terms of hypothesis testing, the null hypothesis is that μ_o is in the interval:

$$\mu_r - TE = \mu_o = \mu_r + TE, \tag{4}$$

and the alternative hypothesis is that μ_o falls outside the interval (4)—the difference between μ_o and μ_r is more than TE.

3.1.1. Sample Size Determination

The standard statistical approach to the above hypotheses testing (Bailey 1981, Roberts 1978) uses the following relationship to determine the sample size n:

$$\left(Z_{\alpha/2} + Z_\beta\right)\frac{\sigma}{\sqrt{n}} = TE,$$

where $Z_{\alpha/2}$ and Z_β are standard normal deviates. In other words,

$$n = \left[\frac{(Z_{\alpha/2} + Z_{\beta})\sigma}{TE} \right]^2 . \tag{5}$$

The auditor, using a sample size n as given in (5), expects to maintain his or her Type-I error at α level and Type-II error at β level. As can be seen from (5), the sample size increases as we decrease Type-I and Type-II risks. Also, it increases as the standard deviation, σ, increases and TE decreases.

3.1.2 Evaluation of Sample Results

Let us consider that the auditor has performed the test and wants to determine whether to accept or reject the recorded mean as fairly stated. The auditor accepts the recorded mean if the audited mean (the sample mean) $\bar{y}$ falls in the following region:

$$\mu_r - A = \bar{y} = \mu_r + A, \tag{6}$$

otherwise he or she rejects it, where the precision A is given by

$$A = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = TE - Z_{\beta} \frac{\sigma}{\sqrt{n}}. \tag{7}$$

The acceptance decision based on (6) and (7) means that the probability of accepting (4) will be (1- α) if $\mu_r = \mu_o$ and approximately β if $|\mu_r - \mu_o| = TE$. As stated earlier, α is the probability of Type I error, the probability of rejecting the null hypothesis if $\mu_r = \mu_o$, and β is the maximum probability of Type II error, the probability of accepting the null hypothesis if it is just barely false (i.e., $\mu_r = \mu_o + TE$, or $\mu_r = \mu_o - TE$).

In reality, the auditor very rarely knows the standard deviation of the population, σ. Thus, he or she must use an estimated standard deviation, S, for computing the precision. Since the auditor is more concerned with incorrect acceptance, he or she determines the precision from (7) based on a desired level of β-risk (AICPA 1981, 1983a, 1983b):

$$A = TE - Z_{\beta} \frac{S}{\sqrt{n}} . \tag{8}$$

The effective level of risk of incorrect rejection (α-risk or Type I error) can be obtained from $Z_{\alpha/2} = \frac{TE}{S/\sqrt{n}} - Z_{\beta}$ (see Equation 7).

3.2. Belief-Function Approach

In this section, we discuss (1) the relationship between the confidence level in a decision interval in a statistical test and the level of belief in the interval; (2) the desired sample size for a statistical test for a given level of belief in the decision interval; and (3) the evaluation of sample results.

3.2.1. Relationship Between Confidence Level and Level of Belief

Srivastava and Shafer (1994) use a "likelihood method" for determining the level of belief from statistical evidence as originally proposed by Shafer (1976). In this approach, the level of confidence in the likelihood interval is defined to be the level of belief in the interval. For example, a 100x% likelihood interval provides x level of belief that the true audited mean μ_o lies in the interval. Srivastava and Shafer (1994) determine the 100x% likelihood interval for μ_o when the sample audited mean is $\overline{y}$:

$$\left[\overline{y} - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)},\ \overline{y} + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)}\right]. \tag{9}$$

Thus the belief, say Bel_2, from the statistical evidence that μ_o lies in the above interval is x:

$$Bel_2\left([\overline{y} - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)} \leq \mu_o \leq \overline{y} + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)}]\right) = x. \tag{10}$$

In general, the recorded account balance is fairly stated (fs, i.e., not materially misstated) when the true mean, μ_o, lies in the interval:

$$\mu_r - TE = \mu_o = \mu_r + TE. \tag{11}$$

In other words, the recorded account balance is not materially misstated when the recorded mean is within the tolerable error of the true (audited) mean, i.e., $|\mu_r - \mu_o| = TE$. However, since we do not know the true mean, we need to express the above condition in terms of the sample audited mean, $\overline{y}$.

In order for the degree of belief in the condition (11) (i.e., the account balance is not materially misstated) to be at least x, the interval in (9) must be contained in the interval $[\mu_r - TE, \mu_r + TE]$. This requirement yields the following interval (see Srivastava and Shafer 1994, p. 522):

$$\mu_r - TE + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)} \leq \overline{y} \leq \mu_r + TE - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)}. \tag{12}$$

Thus, if the condition in (12) were true then we will have the belief that the account balance is not materially misstated to be at least equal to x. However, there is some level of uncertainty in achieving (12).

Thus, for planning purposes, Srivastava and Shafer (1994) require that the condition in (12) be met with probability $(1 - \alpha)$ when $\mu_r = \mu_o$. This requirement can be written as:

$$P\left(\mu_o - TE + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)} \leq \bar{y} \leq \mu_o + TE - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1-x)}\right) = (1-\alpha),$$

or

$$P\left(-\frac{TE}{\sigma/\sqrt{n}} + \sqrt{-2\log_e(1-x)} \leq \frac{\bar{y} - \mu_o}{\sigma/\sqrt{n}} \leq \frac{TE}{\sigma/\sqrt{n}} - \sqrt{-2\log_e(1-x)}\right) = (1-\alpha). \quad (13)$$

The above condition requires that:

$$\frac{TE}{\sigma/\sqrt{n}} - \sqrt{-2\log_e(1-x)} = Z_{\alpha/2},$$

or

$$n = \frac{\sigma^2}{TE^2}\left[Z_{\alpha/2} + \sqrt{-2\log_e(1-x)}\right]^2. \quad (14)$$

Equation (13) suggests that failing to obtain the desired degree of belief of at least x from the statistical evidence is equivalent to rejecting the null hypothesis with probability α — the significance level of the test. Similarly, we can relate the minimum power of the test $(1 - \beta)$ to the degree of belief of at least x in the interval as follows. The probability of rejecting if the null hypothesis is just barely materially false—the minimum power of the test—is the probability of (12) failing when μ_o is equal to μ_r + TE, or

$$1 - P\left(-2Z_{\alpha/2} - \sqrt{-2\log_e(1-x)} \leq \frac{\bar{y} - (\mu_r + TE)}{\sigma/\sqrt{n}} \leq -\sqrt{-2\log_e(1-x)}\right)$$

$$\geq 1 - P\left(\frac{\bar{y} - (\mu_r + TE)}{\sigma/\sqrt{n}} \leq -\sqrt{-2\log_e(1-x)}\right) = 1 - \beta, \quad (15)$$

where β is Type II error and the corresponding normal deviate Z_β can be written as:

$$Z_\beta = \sqrt{-2\log_e(1-x)}. \quad (16)$$

Using (16), we can now write the desired belief x that the condition (12) is true with probability $(1-\alpha)$ and power $(1-\beta)$ as:

$$x = 1 - \exp(-\tfrac{1}{2}Z_\beta^2). \quad (17)$$

3.2.2. Sample Size Determination

From (14) we see that, if our objective is to obtain a degree of belief x that there is no material error, and if we want to achieve this objective with a probability of $(1-\alpha)$ when there is no error at all, then we must use the following relationship to determine the sample size:

$$n = \frac{\sigma^2}{TE^2}\left[Z_{\alpha/2} + \sqrt{-2\log_e(1-x)}\,\right]^2. \tag{18}$$

The minimum power of this test is (1- β) where β is given by

$$\beta = P\left(\frac{\bar{y} - (\mu_r + TE)}{\sigma/\sqrt{n}} \leq -\sqrt{-2\log_e(1-x)}\right). \tag{19}$$

It is noteworthy that the sample size formula (18) for the belief-function approach is the same as (5) for the standard statistical approach. The only difference is that (18) provides us the desired level of belief in the condition (12) that the account balance is not materially misstated which then can be combined with the beliefs from the non-statistical evidence to obtain the overall belief. Srivastava and Shafer (1994) point out that the belief function approach is relatively more efficient because it aggregates objectively both the statistical and non-statistical items of evidence.

Let us consider a specific example. Suppose the auditor wants to determine the sample size to conduct an audit of the inventory account. There are 5,000 items in the stock (N = 5,000). The total recorded value of the inventory is \$2,500,000. The recorded mean μ_r = \$500. The estimated standard deviation σ = \$75, and the tolerable error TE = \$25 per item. Suppose the auditor plans to conduct the audit at 15% risk of incorrect rejection (i.e., α = 15%) and plans to achieve 70% belief in the decision interval. From (18), the sample size for this interval is:

$$n = \frac{75^2}{25^2}\left[Z_{0.075} + \sqrt{-2\log_e(1-0.7)}\,\right]^2$$

$$= 9\left[1.44 + \sqrt{-2\log_e(0.3)}\,\right]^2 = 81.$$

It is interesting to note that the standard statistical approach yields the same sample size as the belief-function approach provided we use the same level of risk of incorrect rejection and the corresponding power related to the desired belief (16).

3.2.3. Evaluation of Sample Results

Let us assume that the auditor has performed the statistical test and obtained the sample mean, $\bar{y}$, and the standard error of the mean, $S/\sqrt{n}$. As discussed earlier, the recorded mean is not materially misstated if the sample mean $\bar{y}$ meets the condition (12). However, the level of belief in the interval depends on the value of $\bar{y}$. The *maximum* belief x that the recorded mean is not materially misstated when the sample mean meets the condition (12) is obtained by solving the following equations[5]:

$$\mu_r + TE - \bar{y} = \frac{S}{\sqrt{n}}\sqrt{-2\log_e(1-x)} \quad \text{for} \quad \mu_r \le \bar{y} \le \mu_r + TE, \tag{20}$$

and

$$\bar{y} - (\mu_r - TE) = \frac{S}{\sqrt{n}}\sqrt{-2\log_e(1-x)} \quad \text{for} \quad \mu_r - TE \le \bar{y} \le \mu_r. \tag{21}$$

The above expressions can be combined in a single equation:

$$\mathrm{Bel}_2(fs) = x = 1 - \exp\left(-\frac{n}{2S^2}(TE - |\bar{y} - \mu_r|)^2\right), \text{ for } \mu_r - TE \le \bar{y} \le \mu_r + TE, \tag{22}$$

where, fs stands for "the account balance is fairly stated." The belief that the recorded mean is not fairly stated (not fs), i.e., materially misstated, when the observed sample mean lies in the interval B = [μ_r - TE, μ_r+ TE] is zero as discussed by Srivastava and Shafer (1994), i.e.,

$$\mathrm{Bel}_2(\text{not } fs) = 0, \text{ for } \mu_r - TE \le \bar{y} \le \mu_r + TE. \tag{23}$$

Similarly, we obtain the following beliefs when the observed mean $\bar{y}$ falls outside the interval B, i.e., for $\bar{y} \ge \mu_r + TE$ or $\bar{y} \le \mu_r - TE$

$$\mathrm{Bel}_2(fs) = 0, \tag{24}$$

$$\mathrm{Bel}_2(\text{not } fs) = 1 - \exp\left(-\frac{n(\bar{y} - \mu_r - TE)^2}{2S^2}\right), \text{ for } \bar{y} \ge \mu_r + TE, \tag{25}$$

and

$$\mathrm{Bel}_2(\text{not } fs) = 1 - \exp\left(-\frac{n(\mu_r - TE - \bar{y})^2}{2S^2}\right), \text{ for } \bar{y} \le \mu_r - TE. \tag{26}$$

Equations (22-23) provide beliefs that the recorded account balance is not materially misstated (fs), and materially misstated (not fs), respectively, when the observed mean $\bar{y}$ falls in the interval B. When the observed mean falls outside of the interval B, the respective beliefs that the recorded balance is fairly stated, and not fairly stated are given by (24-26). These results are interesting and intuitive as discussed below.

Figure 1 here

Figure 1 represents the graphs of the belief that the account balance is fairly stated, $Bel_2(fs)$, and the belief that it is not fairly stated, $Bel_2(\text{not fs})$, as a function of the sample mean $\bar{y}$. The following values are used for the graphs: μ_r =\$500, TE =\$25 per item, n = 100, and three values of the standard error of the mean ($S/\sqrt{n}$ = \$5, \$10, and \$15). In this case, the interval B = [\$475, \$525]. As we can see from (22-26), $Bel_2(fs)$ and $Bel_2(\text{not fs})$ depend on the difference between the observed mean error ($\mu_r - \bar{y}$) and the tolerable error TE. As this difference increases, $Bel_2(fs)$ decreases and become zero at the either end of the interval B and remain zero outside the interval (see Panel A). Also, we find that $Bel_2(fs)$ peaks at $\bar{y} = \mu_r$. Similarly, from Panel B, we observe that $Bel_2(\text{not fs}) = 0$ within the interval B, but outside of the interval, it gradually increases with the increase in the mean error and reaches unity at a large value of ($|\mu_r - \bar{y}|$ - TE). Also, as evident from Figure 1, both $Bel_2(fs)$ and $Bel_2(\text{not fs})$ increase with the decrease in the standard error of the mean for a given sample mean $\bar{y}$.

The above results are intuitive. For example, we expect the belief that the account balance is fairly stated to be maximum when the observed (i.e., audited) mean is equal to the recorded mean. Also, we expect this belief to decrease as the mean error ($\mu_r - \bar{y}$) increases. It should be noted that $Bel_2(fs) = 0$ and $Bel_2(\text{not fs}) = 0$ at the end points of the interval B. This implies that if the sample mean falls right at one of the end points of the interval B then we are ignorant about whether the account balance is materially misstated or not materially misstated. The corresponding Type II error (β-risk) from (19) becomes 0.5. In probability theory, this means that the probability that the account is materially misstated is 50% and that it is not materially misstated is 50%, i.e., we are completely ignorant about the state of the account balance. This is equivalent to the result obtained through the belief function approach.

4. INTEGRATING NON-STATISTICAL AND STATISTICAL EVIDENCE FOR AUDIT DECISIONS UNDER VARIABLE SAMPLING

In general, the auditor integrates statistical and non-statistical evidence at two different stages in an audit. The first stage is when the auditor is planning an audit. The second stage is when the auditor is evaluating the audit.

4.1. Planning an Audit

Let us consider, for simplicity, the earlier example of the audit of an inventory account. Assume that the auditor has already collected all the non-statistical evidence related to the inventory account balance by performing such procedures as: (1) understanding the economic environment for the risk of inventory obsolescence, and (2) analytical procedures. Also, assume that the auditor's overall assessment of belief from all such items of evidence that the account balance is fairly stated is 0.7, i.e., $Bel_1(fs) = 0.7$, and that the account is materially misstated is zero, i.e., $Bel_1(\text{not fs}) = 0$. As discussed in Srivastava and Shafer (1992), the above beliefs could be considered to be either the overall subjective assessment of the auditor or the result of combining the individual assessment of belief from each item of evidence using Dempster's rule. Since we need the basic probability assignment functions or m-values in Dempster's rule, we need to express the above belief in terms of m-values. In the present case, it is straightforward: $m_1(fs) = 0.7$, $m_1(\text{not fs}) = 0$, and $m_1(\{fs, \text{not fs}\}) = 0.3$.

Suppose the auditor needs an overall belief of 0.95 in order to assert that the inventory balance is fairly stated. Since the auditor has already obtained 0.7 degree of belief from the non-statistical evidence, as assumed earlier, the level of belief needed from the statistical evidence to achieve a total belief of 0.95 is determined in the following way. Suppose x represent the degree of belief that the account balance is fairly stated ($Bel_2(fs) = x$) based on the statistical evidence, and there is no belief that the account balance is not fairly stated (i.e., $Bel_2(\text{not fs}) = 0$). These beliefs can be expressed in terms of m-values as: $m_2(fs) = x$, $m_2(\text{not fs}) = 0$, $m_2(\{fs, \text{not fs}\}) = 1 - x$. When these m-values are combined with the m-values obtained from the non-statistical evidence using Dempster's rule, one obtains the following m-values :

$$m(fs) = m_1(fs)m_2(fs) + m_1(fs)m_2(\{fs, \text{not fs}\}) + m_1(\{fs, \text{not fs}\})m_2(fs)$$
$$= 0.7\,x + 0.7\,(1 - x) + 0.3\,x = 0.7 + 0.3\,x.$$

Since the auditor wants $m(fs) = 0.95$, the above equation yields $x = 0.8333$. Thus, the desired level of belief from the statistical evidence that the account balance is not materially misstated is 0.8333 and the belief that it is materially misstated is zero (i.e., $Bel_2(fs) = 0.8333$ and $Bel_2(\text{not fs}) = 0$). Now, the auditor can use (18) to determine the extent of statistical testing, i.e., the sample size n, for the desired level of belief. For TE = \$25, α-risk = 15% (i.e., $Z_{\alpha/2} = 1.44$), σ = \$75, and x = 0.8333, we obtain n = 100 from (18).

4.2. Evaluation of Audit

Continuing with the previous example, assume that the auditor has randomly selected 100 items (n = 100) from the inventory population and performed the relevant statistical procedure. Assume that the audited sample mean of the inventory value per unit is \$486, i.e., $\overline{y}$ = \$486, the observed standard deviation is \$85, and the recorded mean, μ_r = \$500, as assumed earlier. Thus, for TE =\$25 per item, the interval B = [\$475, \$525]. The auditor would like to know whether the sample result along with the non-statistical evidence about the inventory valuation is good enough to accept the account balance to be fairly stated. In other words, the auditor would like to know the level of belief that the account is fairly stated from the statistical evidence. In the present example, we obtain the following beliefs using (22) and (23):

$$\text{Bel}_2(\text{fs}) = 1 - \exp\left(- \frac{100}{2(85)^2}(25 - |486 - 500|)^2\right) = 0.56715,$$

and

$$\text{Bel}_2(\text{not fs}) = 0.$$

This implies that

$$m_2(\text{fs}) = 0.56715,\ m_2(\text{not fs}) = 0, \text{ and } m_2(\{\text{fs, not fs}\}) = 0.43285.$$

Obviously, the target belief of 0.8333 that the account is fairly stated is not available from the statistical evidence. Thus, when the above evidence is combined with the non-statistical evidence (with m_1(fs) = 0.7, m_1(not fs) = 0, m_1({fs, not fs}) = 0.3), the overall belief[6] in 'fs' is Bel(fs) = 0.87015. This is less than the desired value of 0.95. Therefore, the auditor cannot assert that the account is fairly stated. There are several options available to the auditor at this point, including: (1) The auditor can increase the sample size, perform the audit procedure on the additional sample items, analyze the sample results, and determine whether the account balance is fairly stated. (2) The auditor can make an adjustment to the account balance and thus increase the belief to the desired level that the account is not materially misstated. (3) The auditor can reject the account balance as fairly stated and refuse to give an unqualified opinion on the account.

Under the first alternative, the auditor increases the sample size to achieve the desired level of belief of 0.8333 from the statistical evidence. For this purpose, the auditor should use the new estimated standard deviation of $85, and (18) to determine the new sample size. The result is[7] n = 129, compared to n = 100, the initial sample size. The auditor will now select 29 additional items from the inventory stock and perform the audit procedure. Suppose the auditor has done that and obtained a new mean of $489, i.e., $\overline{y}$ = $489 and a new standard deviation of $83, i.e., S = $83. Using (22-23), the auditor obtains[8] a new belief from the statistical evidence of 0.8404 that the account is fairly stated and no belief that the account is not fairly stated. The belief of 0.8404 in 'fs' is greater than the desired value of 0.8333, and thus when the auditor combines the two items of evidence, the statistical evidence and the non-statistical evidence, the overall belief that the account is fairly stated becomes[9] 0.9521 which is above the overall desired value of 0.95. Now, the auditor would feel comfortable accepting the account to be fairly stated since the combined belief in 'fs' is above the desired level.

Under the second alternative, the auditor proposes an adjustment to the account such that the level of belief from the statistical evidence that the account is not materially misstated is at least 0.8333, as required. In the present case, this objective is achieved by reducing the recorded mean to $494.91 from $500.00 (use (21) to compute the adjusted recorded mean). This suggests that the auditor should propose a reduction of the recorded mean by $5.09. The adjusted mean will yield the desired belief of 0.95 for the overall belief in 'fs.'

The last alternative is straightforward. The auditor issues either a "qualified opinion" or an "adverse opinion" depending on the severity of the misstatement (see, e.g., Arens and Loebbecke 1994).

5. INTEGRATING NON-STATISTICAL AND STATISTICAL EVIDENCE FOR AUDIT DECISIONS UNDER ATTRIBUTE SAMPLING

Attribute sampling is a statistical technique where one predicts the presence or absence of an attribute in the population based on the sample results. Attribute sampling is used by auditors to determine whether a control procedure is being followed when a transaction is processed and how effective the control is. The more effective the controls, the less detailed tests of account balances the auditor may need to perform. Because tests of details, such as inventory valuation discussed in Section 3, are more costly, the auditor would like to depend heavily on the controls if they are effective.

In general, the auditor collects both non-statistical and statistical types of evidence for determining the effectiveness of controls. An example of a non-statistical item of evidence for testing the control "credit sales are properly approved by the credit manager before goods are shipped," is the information related to the control environment such as management philosophy towards the importance of controls; credit manager's integrity, honesty and competence; factors dealing with economic environment; etc. In fact, *Statements on Auditing Standards No. 55: Consideration of the Internal Control Structure in a Financial Statement Audit* (AICPA 1988) requires auditors to collect and evaluate such pieces of information along with developing understanding of the client's accounting system before performing statistical tests for control effectiveness. An example of a statistical test (attribute sampling) for the control mentioned above is to check on a sample basis whether the credit manager has really been approving the sales before the goods are shipped by looking for the presence or absence of the credit manager's signature on the processed sales orders.

The extent of testing for attribute sampling for the effectiveness of a control depends on the level of belief the auditor obtains from the non-statistical evidence. For example, if the overall feeling of the auditor is good towards the factors mentioned earlier under the non-statistical evidence, the auditor will perform less extensive attribute sampling for the effectiveness of the control. However, currently, we do not have any approach that formalizes the above process. Recently, Srivastava and Gillett (1995) have proposed a belief-function approach for attribute sampling similar to the one discussed in Section 3 for variable sampling. In the belief-function approach, the auditor first should estimate the level of belief for the effectiveness of a control based on the non-statistical evidence and then combine that with the belief obtained from the attribute sampling to obtain the overall belief that the control is effective.

Like variable sampling, there are three main issues in integrating the non-statistical evidence with the statistical evidence in attribute sampling. First, the auditor needs to make an estimate of the belief he or she obtains from the non-statistical evidence. This estimate could be based on an overall subjective judgment of the auditor or on several individual belief estimates related to individual pieces of information which are combined using Dempster's rule. Second, the auditor needs to determine the sample size n for attribute sampling for a given belief 'x' that the control is effective. The value of 'x' depends on the degree of the overall belief desired by the auditor for the control effectiveness and the level of belief obtained from the non-statistical evidence. Third, the auditor has to determine the level of belief that the control is effective based on the sample results. If the achieved belief is less than the planned belief of 'x' then the auditor has to take certain actions, including: (1) think of consequences of the weakness, and see if there are any compensating controls that need consideration, (2) reduce the planned level of reliance on the control, (3) extend the sample size to obtain the planned level of belief.

5.1 Sample Size Determination

In attribute sampling, the auditor would be interested in determining the belief that the true population occurrence rate of the attribute being tested lies in an acceptable interval B = [p1, p2] where p1 and p2 are the lower and upper ends of the interval. Usually, the auditor sets p1 =0 and p2 = TOR (tolerable occurrence rate). For a binomial distribution, Srivastava and Gillett (1995) derive the following relationship for the sample size calculation for a desired level of belief 'x' that the true population occurrence rate lies in B:

$$x = 1 - \left(\frac{n.TOR}{k}\right)^{k}\left(\frac{n - n.TOR}{n - k}\right)^{n - k}, \qquad (27)$$

where k is the estimated sample occurrence number. As Srivastava and Gillett show, the sample size increases as the belief in the interval B increases. Also, the sample size increases as the value of k increases, i.e., as the number of occurrence expected in the sample increases. As it is evident, one can not solve (27) for n analytically. However, one can write a computer program to determine the value of n for the desired values of k and TOR. This is discussed further below.

Let us consider a numerical example to illustrate the sample size calculation. Consider that the auditor is testing the effectiveness of a control through attribute sampling and the attribute being tested is "the sales order is not properly approved by the credit manager before shipment." Based on the non-statistical evidence, the auditor has a belief of, say 0.7, that the control is operating effectively. For a total belief of, say, 0.94 that the control is operating effectively, the auditor will desire a belief[10] of 0.8 from the attribute sampling. Suppose the control is considered effective if the population occurrence rate falls in the interval B =[0, 0.1]. Also, assume that 0.8 level of belief is desired for k = 1 (i.e., with no more than one occurrence in the sample). Thus, from (27), a sample of 39 is obtained for k = 1, and a belief of 0.8 that the population occurrence rate is in the interval B = [0, 0.1].

5.2. Evaluation of Sample Results

Continuing with our auditing example discussed earlier, assume that the auditor has performed the attribute sampling with a sample size of 40 (i.e., n = 40) and has obtained one occurrence of the attribute (k = 1). For a tolerable occurrence rate of 0.1 (i.e., TOR = 0.1, and B = [0, 0.1]), the level of belief from (27) that the control is operating effectively is 0.8237. This is more than the planned level of belief, 0.8. However, if the auditor found two occurrences in the sample of 40 then the achieved level of belief calculated from (27) would be 0.4874, a significantly lower belief than planned. If the result is not acceptable then there are several options available to the auditor, as mentioned earlier in this section.

6. SUMMARY AND CONCLUSION

We have demonstrated how to integrate statistical and non-statistical evidence using belief functions. We have used auditing examples to illustrate the process. The main reason for choosing auditing examples is that it is quite common in auditing to come across both types of evidence. Also, auditors do use both types of evidence in their analyses. We have discussed the case of variable sampling in detail. However, only a brief discussion is presented for attribute sampling. We have derived a formula for the sample size needed for a desired level of belief in each case. Numerical examples are used to illustrate both the sample size determination and the sample result evaluation. It is interesting to note that the sample size formulae are similar to the formulae used in the standard statistical approach. As expected, with all the factors held fixed, the sample size increases with the increase in the desired level of belief. As discussed by Srivastava and Shafer (1992), the belief-function approach, in general, should provide a more efficient audit because it allows the auditor to integrate all the statistical and non-statistical evidence in an objective way. If the belief from the non-statistical evidence is large then the auditor needs only a small amount of belief from the statistical evidence to achieve the desired level of overall belief. It should be emphasized that belief functions provide a more intuitive representation of auditor's ignorance (lack of evidence) than Bayesian probabilities.

FOOTNOTES

1. In general, there are three types of audit: audit of financial statements, compliance audit, and operational audit. The basic audit process in all three cases is the same, and involves collection, evaluation and aggregation of evidence related to the audit objective(s).

2. Most of this section's material is taken from Srivastava (1993).

3. In the case of n elements in the frame, we will have $P(a_i) = 0$, and $\sum_{i=1}^{n} P(a_i) = 1$, where a_i represents the ith element of the frame.

4. For a frame of n elements, we will have, in general, m-values for each of the individual elements, each set of two elements, each set of three elements, and so on, to the m-value for the entire frame. All such m-values add to one, i.e., $\sum_{A \subseteq \Theta} m(A) = 1$, where A represents all the proper subsets of the frame Θ. The m-value for the empty set is zero.

5. The 100x% likelihood interval for μ_o is (see Equation 9):

$$\left[\bar{y} - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1 - x)},\ \bar{y} + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1 - x)}\right].$$

The condition that the account balance is not materially misstated is obtained by requiring that the likelihood interval be contained in the interval $[\mu_r - TE, \mu_r + TE]$, that is, the upper boundary of the likelihood interval be less than $(\mu_r + TE)$ and the lower boundary be greater than $(\mu_r - TE)$. This requirement gives the following condition for the recorded account balance to be not materially misstated (i.e., fairly stated):

$$\text{fs: } \mu_r - TE + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1- x)} \leq \bar{y} \leq \mu_r + TE - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1- x)}.$$

The condition that the account balance is materially misstated is (i.e., not fairly stated):

$$\text{not fs: } \bar{y}\ \text{Š}\ \mu_r - TE + \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1- x)} \text{ or } \bar{y} \bullet \mu_r + TE - \frac{\sigma}{\sqrt{n}}\sqrt{-2\log_e(1- x)}$$

6. The two items of evidence provide the following m-values: $m_1(fs) = 0.7$, $m_1(\text{not } fs) = 0$, and $m_1(\{fs, \text{not } fs\}) = 0.3$, and $m_2(fs) = 0.56715$, $m_2(\text{not } fs) = 0$, and $m_2(\{fs, \text{not } fs\}) = 0.43285$. Using Dempster's rule (Equations 2 and 3), we obtain K = 1 and

$$m(fs) = m_1(fs)m_2(fs) + m_1(fs)m_2(\{fs, \text{not } fs\}) + m_1(\{fs, \text{not } fs\})m_2(fs)$$

$$= 0.7x0.56715 + 0.7x0.43285 + 0.3x0.56715 = 0.87015,$$

$m(\text{not } fs) = 0$, and $m(\{fs, \text{not } fs\}) = m_1(\{fs, \text{not } fs\})m_2(\{fs, \text{not } fs\}) = 0.3x0.43285 = 0.12985$.

7. $n = \frac{85^2}{25^2}\left[1.44 + \sqrt{-2\log_e(1 - 0.8333)}\right]^2 = 129.$

8. $Bel_2(fs) = 1 - \exp\left(-\frac{129}{2(83)^2}(25 - |489 - 500|)^2\right) = 0.8404$, $Bel_2[\text{not } fs] = 0$.

9. $m(fs) = m_1(fs)m_2(fs) + m_1(fs)m_2(\{fs, \text{not } fs\}) + m_1(\{fs, \text{not } fs\})m_2(fs)$

$= 0.7 \times 0.8404 + 0.7 \times 0.1596 + 0.3 \times 0.8404 = 0.9521$.

10. The two sets of m-values, one from the statistical evidence and the other from the non-statistical evidence, when combined using Dempster's rule, yield a total belief of 0.94 that the control is effective.

REFERENCES

American Institute of Certified Public Accountants. 1981. *Statements on Auditing Standards, No. 39: Audit Sampling.* New York: AICPA.

American Institute of Certified Public Accountants. 1983a. *Audit and Accounting Guide: Audit Sampling*. New York: AICPA.

American Institute of Certified Public Accountants. 1983b. *Statements on Auditing Standards, No. 47: Audit Risk and Materiality in Conducting an Audit.* New York: AICPA.

American Institute of Certified Public Accountants. 1988. *Statement on Auditing Standards, No. 55: Consideration of the Internal Control Structure in a Financial Statement Audit*. New York: AICPA.

Arens, A. A. and Loebbecke, J. K. 1981. *Applications of Statistical Sampling to Auditing,* Englewood Cliffs, NJ: Prentice-Hall.

Arens, A. A. and Loebbecke, J. K. 1994. *AUDITING: An Integrated Approach*. Englewood Cliffs, NJ: Prentice-Hall.

Bailey, Jr., A. D. 1981. *Statistical Auditing: Review, Concepts, and Problems.* Harcourt Brace Jovanovich.

Roberts, D. M. 1978. *Statistical Auditing*. American Institute of Certified Public Accountants. New York.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.

Shafer, G. R. and R. P. Srivastava. 1990. The Bayesian and Belief-Function Formalisms: A General Perspective for Auditing. *Auditing: A Journal of Practice and Theory*, (Supplement): 110-48.

Srivastava, R. P. 1993. Belief Functions and Audit Decisions. *Auditors Report*, Vol. 17, No. 1, Fall: 8-12.

Srivastava, R. P. and G. Shafer. 1994. Integrating Statistical and Non-Statistical Audit Evidence Using Belief Functions: A Case of Variable Sampling. *International Journal of Intelligent Systems,* Vol. 9: 519-539.

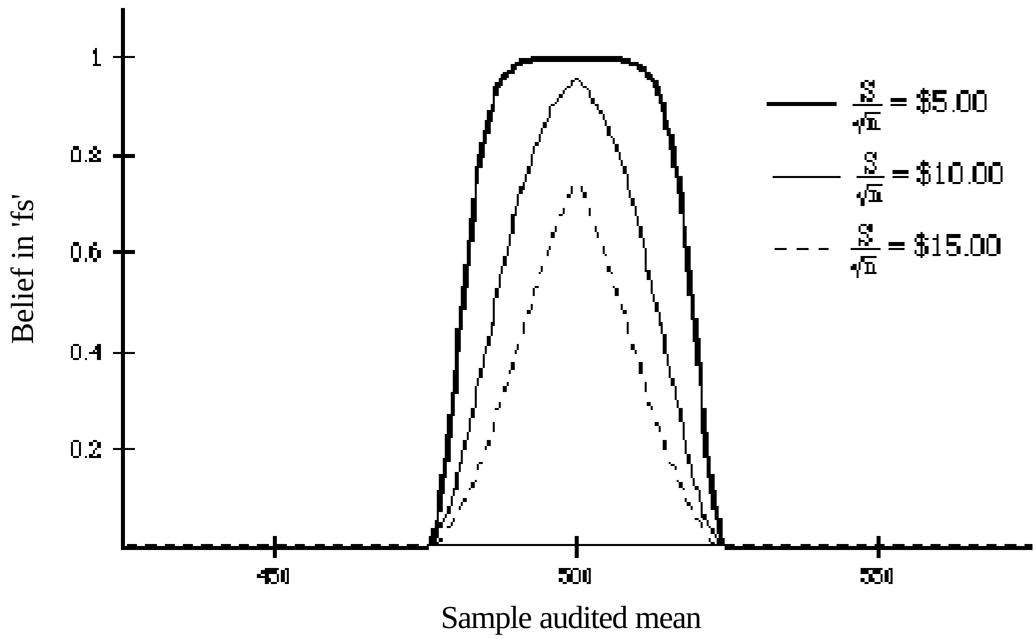
Srivastava, R. P. and G. R. Shafer. 1992. Belief-Functions Formulas for Audit Risk. *The Accounting Review* (April): 249-283.

Srivastava, R. P. and P. R. Gillett. 1995. Integrating Statistical and Non-Statistical Audit Evidence in Attribute Sampling Using Belief Functions. Working Paper, School of Business, The University of Kansas.

Figure 1

Belief in 'fs' and 'not fs' (Equations 22-26) as a function of the sample audited mean $\overline{y}$ for different values of the observed standard error of the mean ($S/\sqrt{n}$). B = [\$475, \$525], μ_r = \$500, n = 100, and TE = \$25 (taken from Srivastava and Shafer 1994).

Panel A: Belief in 'fs' as a function of the sample audited mean.



Panel B: Belief in 'not fs' as a function of the sample audited mean.