Auditing: A Journal of Practice and Theory, Vol. 19, No. 1, Spring 2000, pp. 145-155.

Integrating Statistical and Non-Statistical Audit Evidence in Attribute Sampling Using Belief Functions

Peter R. Gillett
Associate Professor
Accounting and Information Systems Area
Faculty of Management
Rutgers University-Newark
Newark, NJ 07102-1895

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

Integrating Statistical and Non-Statistical Audit Evidence in Attribute Sampling Using Belief Functions

ABSTRACT

The main purpose of this article is to show how one can integrate statistical evidence from attribute sampling with non-statistical evidence within the Dempster-Shafer belief function framework. In particular, the article shows: (1) how to determine the sample size in attribute sampling to obtain a desired level of belief that the true attribute occurrence rate of the population lies in a given interval; (2) what level of belief is obtained for a specified interval given the sample result; and (3) how to integrate non-statistical evidence with the statistical evidence arising from the attribute sampling. These issues are important to the auditor and therefore we use auditing examples to illustrate the process. As intuitively expected, we find that the sample size increases as the desired level of belief in the interval increases. In evaluating the sample results, we again find results that are intuitively appealing. For example, provided the sample occurrence rate falls in the interval B for a given number of occurrences of the attribute, we find that the belief in B, Bel(B), increases as the sample size increases. However, if the sample occurrence rate falls outside of the interval then Bel(B) is zero. Note that, in general, both Bel(B) and Bel(notB) are zero when the sample occurrence rate falls at the end points of the interval. These results extend similar results already available for variables sampling. However, the auditor faces an additional problem for attribute sampling: how to convert belief in an interval for control exceptions into belief in an interval for material misstatements in the financial statements, so that it can be combined with evidence from other sources in implementations of the Audit Risk Model. We discuss this problem, and investigate conversion methods that are consistent with current auditing practice.

Integrating Statistical and Non-Statistical Audit Evidence in Attribute Sampling Using Belief Functions

1. INTRODUCTION

The main purpose of this article is to show how one can integrate statistical and non-statistical evidence in attribute sampling using belief functions. In particular, we show the following: (1) how to determine the sample size in attribute sampling to obtain a specified level of belief that the true attribute occurrence rate of the population lies in a given interval; (2) what level of belief is obtained for the interval given the sample result; and (3) how to integrate statistical and non-statistical evidence in attribute sampling.

The above issues are important to the auditor. However, before we give a specific example, we would like to give a brief introduction to the audit process for readers not familiar with what auditors do and how they do it, in order for them to have a better perspective for the paper. The accounting profession defines the audit process[1] as (see, e.g., Arens and Loebbecke 1994):

> Auditing is the accumulation and evaluation of evidence about quantifiable information of an economic entity to determine and report on the degree of correspondence between the information and established criteria (p. 1).

Thus, in essence, the audit process is the accumulation, evaluation and aggregation of evidence in order to express an opinion that the financial statements (FS) present fairly the financial position of the business. The quantifiable information in the above definition relates to the accounts on the FS such as cash, accounts receivable, inventory etc. and the established criteria relate to the generally accepted accounting principles (GAAP). GAAP provides guidelines on how economic transactions should be recorded and reported.

It is understood that the FS are the representations of management about the company's financial position. Through the FS, management is making certain assertions[2] about the assets, liabilities, sales, expenses, and other accounts in the FS. For example, for assets they are asserting that the reported assets do exist, all the assets that exist are recorded, the assets are properly valued, the company has the right to use them, and so on (see any auditing textbook for more details). The auditor's responsibility lies in accumulating enough evidence to make reasonably sure that the management assertions for each account are warranted, thus confirming the fair presentation of the FS.

It is generally accepted that a good accounting system with effective internal controls that prevent, detect, and correct errors in recording transactions will be more likely to generate reliable (i.e., error free) accounting information. In fact, through internal controls the company not only seeks to ensure that accounting information is reliable but also that the company's policies and procedures are being followed, the assets are protected, operational efficiency and effectiveness are being achieved, and the company is complying with the laws and regulations. However, even an excellent control system cannot prevent errors due to management fraud or employee collusion to steal company's assets (defalcation).

At present, on every audit engagement, the auditor is required by the American Institute of Certified Public Accountants (AICPA, 1988) to acquire sufficient understanding of the company's internal control structure. There are three main components of the internal control structure: control environment, accounting system, and control procedures. It is expected that the auditor will be able to assess the risk of misstatements of both kinds, intentional (irregularities due to management fraud or employees collusion) and unintentional (errors), after obtaining a good understanding of the control structure. To understand the control environment, the auditor must search for information such as management attitude towards controls, management integrity, their style of administration (authoritarian or consensus builder), employees' honesty and competence etc. In order to understand the accounting system, the auditor must gather information related to how transactions are processed and recorded in the system. For the third component, the auditor must find out what specific controls are in use by inquiry, observation and studying the company's policy and procedure manuals. Almost all of the above items of evidence related to acquiring sufficient understanding of the client's control structure are non-statistical in nature.

The next step in the audit process is that when the auditor feels that the control environment is sound, the accounting system appears to be processing information in accordance with GAAP, and the company seems to have good control procedures in place, the auditor may evaluate how effective the controls are. The more effective the controls, the less detailed tests of account balances the auditor may need to perform. Tests of details are more costly and therefore the auditor would like to depend heavily on internal controls if they are effective. How does an auditor test for the effectiveness of an internal control?

There are several different approaches and they all depend on the nature of the control. For certain controls simply observing the process (in conjunction with appropriate inquiries) is good enough. For example, if the auditor wants to check whether certain incompatible tasks such as recording of transactions and custody of the related assets are segregated then observing the activity on a surprise basis may be the best way to test this control. However, in making judgments about the effectiveness of certain other controls, such as "credit sales are properly approved by the credit manager before goods are shipped," the auditor, in general, accumulates three types of evidence: (1) 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 the economic environment etc; (2) test results on a sample basis as to whether the credit manager has really been approving the sales before the goods are shipped by looking for the presence or absence of the signature[3] of the credit manager on the processed sales orders; (3) test results on a sample basis to see whether the approval has been performed properly, i.e., the manager has really checked on the customer's credit before approving the sales[4]. The first type of evidence falls into the non-statistical category. The second and third types fall into a statistical category that relates to attribute sampling. In considering how internal controls are likely to affect the reliability of information in the financial statements, the auditor typically considers statistical evidence from attribute sampling in conjunction with non-statistical evidence from other sources. How can the auditor combine this evidence? We provide a belief-function approach to this problem[5]. We suggest that an objective approach to integrating statistical and non-statistical evidence will lead to an efficient audit.

Many researchers have argued that belief functions may provide a better framework for representing uncertainties in the audit evidence than probability theory (see, e.g., Akresh, Loebbecke, and Scott 1988, Gillett 1993, Shafer and Srivastava 1990, and Srivastava and Shafer 1992, and Srivastava

1993). In general, the structure of audit evidence forms a network of variables: the accounts in the financial statements, the audit objectives[6] for the accounts, and the financial statements as a whole. Aggregating all the evidence in an audit to determine whether the FS present fairly the financial position of the company becomes a problem of propagating beliefs in a network. Shafer, Shenoy, and Srivastava (1988, see also, Srivastava 1995, Srivastava, Dutta, and Johns 1994) have discussed this problem. However, none of these works deal with the issue of how to integrate statistical and nonstatistical evidence; they all assume that the auditor has the beliefs relevant to the appropriate variables from each item of evidence gathered irrespective of their nature.

Recently, Srivastava and Shafer (1994) have discussed the issue of integrating statistical and nonstatistical evidence related to mean per unit variable sampling. However, as discussed above, the integration of non-statistical evidence with the statistical evidence derived from attribute sampling is needed for determining the effectiveness of internal controls. This paper deals with this issue.

The remainder of the paper is divided into four sections. Section 2 deals with the standard statistical approach to attribute sampling. Section 3 discusses concepts related to consonant belief functions and their relationship with the statistical evidence. Section 4 presents the belief-function approach to attribute sampling and discusses how statistical and non-statistical evidence can be integrated using belief functions. Finally, Section 5 provides a summary and conclusion of the paper with potential research problems.

2. THE STANDARD STATISTICAL APPROACH

Let us consider the internal control example discussed earlier. The auditor wants to test with a certain level of confidence that the control 'credit sales are properly approved by the credit manager before goods are shipped' is effective. For this purpose, the auditor selects a sample of sales orders for which goods have been shipped, reperforms the procedures that are usually performed by the credit manager and makes a note of the sales orders that are not properly approved (i.e., approval is not noted on the document and the customer's file has not been checked for the credit limit and unpaid balance in the account). This type of sampling is called attribute sampling; the relevant attribute in this case is ‘failure of proper credit approval.’ From sampling, the auditor wants to predict the occurrence rate in the population.

Usually in attribute sampling, we test whether the occurrence rate, P, of an attribute in the population is equal to a certain value p_0 or different. This situation can be expressed in terms of the null hypothesis as :

$$P = p_0,$$

and the alternative hypothesis as:

$$P \neq p_0.$$

In effect, the null hypothesis is that p_0 lies in the interval B = [p_1, p_2], i.e., $p_1 = p_0 = p_2$, and the alternative hypothesis is that p_0 lies outside the interval where p_1 and p_2, respectively, represent the lower and upper precision limits of the occurrence rate judged by the decision maker.

2.1. Sample Size Determination

The sample size for the test is determined by solving the following equations simultaneously (see, e.g., Arens and Loebbecke 1981, and Bailey 1981):

$$\text{P(number of occurences } \leq \text{ cv}|np_2) = \prod_{r=0}^{cv} \frac{n!}{r!(n\text{-}r)!} p_2^r (1-p_2)^{n-r} = b\ , \tag{1}$$

$$\text{P(number of occurences } \leq \text{ cv}|np_1) = \prod_{r=0}^{cv} \frac{n!}{r!(n\text{-}r)!} p_1^r (1-p_1)^{n-r} = 1 - a\ , \tag{2}$$

where β, and α, respectively, represent the planned level of Type II and Type I errors, and cv represents the critical value of the number of occurrences in the sample for acceptance and rejection[7] of the null hypothesis.

For cv = 0, i.e., if we decide to reject the null hypothesis when the number of occurrences of the attribute is more than 0, the sample size from (1) is given by:

$$n = \frac{\text{Log}\beta}{\text{Log}(1 - p_2)}. \tag{3}$$

The level of Type I error (α-risk) is automatically zero with a choice of cv = 0. For all other values of cv, we need to use either a table for binomial distribution or a computer program to find a value for n that simultaneously satisfies both (1) and (2).

2.2. Evaluation of the Sample Result

Suppose during the test we obtained 'k' number of occurrences of the attribute in the sample. The lower and upper precision limits of the confidence interval are determined by solving the following equations for the desired level of α and β:

$$\prod_{r=0}^{k} \frac{n!}{r!(n-r)!} p_U^r (1 - p_U)^{n-r} = b \ ,$$

$$\prod_{r=0}^{k} \frac{n!}{r!(n-r)!} p_L^r (1 - p_L)^{n-r} = 1 - a \ .$$

Again, one can use a table of the binomial distribution or a computer program to obtain p_L and p_U from the above equations. If this interval, i.e., [p_L, p_U], falls in the original interval [p_1, p_2] established by the decision maker, then the null hypothesis is accepted otherwise rejected.

3. CONSONANT BELIEF FUNCTIONS AND STATISTICAL EVIDENCE

In this section, we briefly describe the basic concepts of *consonant belief functions* and how they are related to statistical evidence. Shafer (1976) first used consonant belief functions for determining the level of belief from the statistical evidence. Recently, Srivastava and Shafer (1994) have used the concepts to integrate non-statistical evidence with the statistical evidence arising from mean per unit variable sampling.

Shafer (1976) has shown that if f represents a real-valued function on the frame Θ with the property that

$$0 = f(\theta) = 1 \text{ for all } \theta \text{ in } \Theta, \qquad (4)$$

and

$$f(\theta) = 1 \text{ for at least one } \theta \text{ in } \Theta,$$

then the function Bel defined by

$$\mathrm{Bel}(B) = 1 - \max_{\theta \in \mathrm{not}B} f(\theta), \tag{5}$$

for each non-empty B is a belief function. He calls such a belief function a *consonant* belief function. The corresponding plausibility function is given by

$$\mathrm{PL}(B) = \max_{\theta \in B} f(\theta), \tag{6}$$

and the plausibility for a single point θ is

$$\mathrm{PL}(\{\theta\}) = f(\{\theta\}). \tag{7}$$

Equations (6) and (7) suggest that the plausibility of a singleton θ is f(θ) whereas the plausibility of a non-singleton set is the largest plausibility of any of its elements.

We can define a consonant belief function for a continuous function f in a similar manner as it is defined in a discrete case. The plausibility of a singleton θ in Θ is f(θ) whereas the plausibility of a subset B of Θ is given by

$$\mathrm{PL}(B) = \sup_{\theta \in B} f(\theta), \tag{8}$$

and the belief in B by

$$\mathrm{Bel}(B) = 1 - \mathrm{PL}(\mathrm{not}B) = 1 - \sup_{\theta \in \mathrm{not}B} f(\theta). \tag{9}$$

The term “sup” in (8) and (9) represents a *supremum*, the least upper bound of the function.

In the present article, our interest is to obtain beliefs from statistical evidence. We will use the renormalized[8] likelihood function associated with the statistical evidence to represent the function f defined in (4), as recommended by Shafer (1976). In attribute sampling, the objective is to determine the rate of presence or absence of an attribute (or the number of occurrences of such an attribute) in the population based on the sample result. We will limit our discussions to the case where we have only two possible outcomes because of our interest in the application to auditing. For such a case, we may consider three probability distributions that can be used to describe the random process: (1) the hypergeometric distribution when the sampling is done without replacement, (2) the binomial distribution when the sampling is done with replacement, and (3) the Poisson distribution.

Usually, the sample units, such as sales orders in our earlier example, are selected without replacement and thus the hypergeometric distribution is the appropriate distribution for the process.

However, if the sample size, n, is much smaller than the population size, N (as a rule of thumb n = 0.1N), the binomial distribution is a good approximation to the hypergeometric distribution. One can also use Poisson distribution to describe the above random process if the rate of occurrence, P, of the attribute in the population is small (P = 0.1), n = 0.1N, and n = 50 (Cochran 1977, Bailey 1981, Roberts 1978). We will use only the binomial distribution in our discussion because it is a good approximation of the theoretically correct distribution, the hypergeometric distribution, and much simpler to work with computationally. We will not provide any discussion on the Poisson distribution because the steps are identical and readers can obtain the results easily following the steps presented in the paper. However, we do provide the renormalized likelihood function for the Poisson distribution in footnote 9.

The renormalized likelihood function (see footnote 9) for the binomial distribution is

$$f(P) = \left(\frac{nP}{k}\right)^k \left(\frac{n - nP}{n - k}\right)^{n - k}, \tag{10}$$

where 'a' represents the number of occurrences of the attribute in the sample.

The belief that the true value of the population occurrence rate lies in an interval B = $[p_1, p_2]$ when the sample occurrence rate falls in the interval may be determined by (9) and (10) above:

$$\mathrm{Bel}(B) = 1 - \mathrm{PL}(\mathrm{not}B) = 1 - \max_{P \varepsilon \mathrm{not}B} f(P),$$

$$= 1 - \max_{P \varepsilon \mathrm{not}B} \left(\frac{nP}{k}\right)^k \left(\frac{n - nP}{n - k}\right)^{n - k}. \tag{11}$$

Since the renormalized binomial function in (10) above has its maximum value in each of the tails outside the interval B at the end points of the interval, we may write (11) as:

$$\mathrm{Bel}(B) = 1 - \max \left\{ \left(\frac{np_1}{k}\right)^k \left(\frac{n - np_1}{n - k}\right)^{n - k}, \left(\frac{np_2}{k}\right)^k \left(\frac{n - np_2}{n - k}\right)^{n - k} \right\}. \tag{12}$$

If the sample occurrence rate does not fall in B then (11) yields:

$$\mathrm{Bel}(B) = 0. \tag{13}$$

In many auditing applications, the lower limit of the interval is set to zero; i.e., B = $[0, p_2]$. When the sample occurrence rate falls in such an interval B, equation (12) reduces[10] to:

$$\text{Bel(B)} = 1 - \left(\frac{np_2}{k}\right)^k \left(\frac{n - np_2}{n - k}\right)^{n-k}. \quad (14)$$

If we further assume that there are no exceptions (i.e., k = 0), then the above expression further simplifies[11] to:

$$\text{Bel(B)} = 1 - (1 - p_2)^n. \quad (15)$$

4. BELIEF-FUNCTION APPROACH

In Section 3, we discussed how to assess the belief that the true population occurrence rate lies in an interval B = $[p_1, p_2]$ using the concepts of consonant belief functions. From (11), we know that if a falls in B then the belief in B is given by (12). However, if a does not fall in B then the belief in B is zero. Here we use (11) to (a) determine the sample size for a desired level of belief in an interval B, and (b) evaluate the sample results and determine the level of belief in the interval.

4.1. Sample Size Determination

Similar to the statistical approach described earlier, we need first to make a judgment about the interval in which we want a desired belief of say x before we determine the sample size for the test. Let us assume that the interval is defined by the lower precision limit p_1 and the upper precision limit p_2, as earlier. Next, we need to estimate k, the number of occurrences in the sample. This judgment is similar to the judgment for the critical value cv in the statistical approach; however, the interpretation is different. In the statistical approach, if the sample occurrence number exceeds the critical value, we reject the null hypothesis. However, in the belief-function approach, if the sample occurrence number is higher than the estimated value for k then we do not reject the hypothesis, rather the belief in B is accordingly adjusted. Rejection of the hypothesis will occur once the statistical evidence is integrated with the non-statistical evidence and if the total belief is still below the overall desired belief then, of course, the hypothesis is rejected.

Thus, for a desired level of belief x in B, the sample size is determined by solving the following equation from (12) above, knowing the interval B = $[p_1, p_2]$ and k, the sample occurrence number:

$$x = 1 - \max\left\{ \left(\frac{np_1}{k}\right)^k \left(\frac{n - np_1}{n - k}\right)^{n-k},\ \left(\frac{np_2}{k}\right)^k \left(\frac{n - np_2}{n - k}\right)^{n-k} \right\},$$

i.e.,

$$x - 1 + \max\left\{ \left(\frac{np_1}{k}\right)^k \left(\frac{n - np_1}{n - k}\right)^{n-k},\ \left(\frac{np_2}{k}\right)^k \left(\frac{n - np_2}{n - k}\right)^{n-k} \right\} = 0. \qquad (16)$$

For B = $[0, p_2]$, as shown in (14), the above equation reduces to:

$$x - 1 + \left(\frac{np_2}{k}\right)^k \left(\frac{n - np_2}{n - k}\right)^{n-k} = 0, \qquad (17)$$

and for k = 0 (along with $p_1 = 0$), we get a simple solution for n (see equation 15):

$$n = \frac{\text{Log}(1 - x)}{\text{Log}(1 - p_2)}. \qquad (18)$$

This solution is similar to the solution (3) obtained using the statistical approach. The only difference is that (1 – x) is replaced by β. This is an interesting result. For k = 0 and B = $[0, p_2]$, the belief-function approach is identical to the statistical approach, and belief in B becomes the power of the test, i.e., x = 1 – β (compare columns 2 and 4 of Tables 1 and 2 for k = 0).

Example

We consider an auditing example here to illustrate the process of determining the sample size for an attribute sampling using belief functions. Usually in auditing, the interval is chosen with the lower precision limit to be zero ($p_1 = 0$) and the upper limit to be some value that the auditor considers acceptable for the particular attribute being tested. We will call this value TER for tolerable exception rate. Let us consider the example discussed earlier that the auditor is testing the effectiveness of an internal control through attribute sampling and the attribute being tested is the "the sales order is not properly approved by the credit manager." The auditor feels that a belief of, say 0.8, that the population occurrence rate is in the interval B = [0, 0.1] is needed in order to achieve the desired overall belief, say 0.95, when combined with the other items of evidence. Also, suppose that 0.8 level of belief in B is desired for k = 1 (i.e., with no more than one exception in the sample). Using (17), we obtain a sample size of 39 (i.e., n = 39, see Table 1). In practice, the auditor would probably select a sample of 40.

--- Tables 1 and 2 here ---

Tables 1 and 2 give sample sizes for different values of k, x and TER using the belief-function approach. We also list the corresponding values of the planned power of the test. There are some interesting observations about the sample results we would like to comment on.

First, as one can see, the sample size decreases as the desired belief in the interval decreases (compare columns 2 and 3 in Tables 1 and 2). Second, as the interval size increases the sample size decreases for the same level of belief (compare Table 1 and 2). These are intuitive results. You do expect that you have to have a bigger sample size to achieve a higher belief in a given interval. Also, as the tolerable exception rate increases, you do expect to do relatively less work (a smaller sample size for a given belief).

Third, as seen from columns 1 and 3 of Tables 1 and 2, the sample size increases as the value of k increases. The sample size is the smallest for a given belief in the interval for k = 0. This result is similar to the result one obtains in the statistical approach with cv = 0, as discussed earlier. This finding is again intuitive; you do expect to sample less if you expect fewer occurrences in the sample.

Finally, let us compare the planned power of the test and the desired belief in the interval (columns 2 and 4 in Tables 1 and 2). For k = 0, as discussed earlier, the desired belief and the planned power are the same. However, for k > 0, the planned power decreases with a slower rate than the desired belief. This is again expected. In general, in belief functions, a zero belief on all the subsets of a frame and a value of one on the entire frame represents ignorance whereas in probability theory equal probabilities of occurrence of all singletons in the frame represents ignorance. In the present context with only two outcomes, a zero belief on both B and notB, i.e., Bel(B) = Bel(not B) = 0, implies ignorance but in probability theory 0.5 probability on each outcome represents ignorance. In principle, we expect to obtain 0.5 level of power of the test for a zero belief in the interval. However, this does not happen in the present case, as the results in Tables 1 and 2 show, because the distribution is not symmetric. Srivastava and Shafer (1994) obtain such a result in the case of a normal distribution which is a symmetric distribution (see their equation 15).

4.2. Evaluation of the Sample Result

Continuing with our auditing example discussed earlier, suppose the auditor has performed the test as planned in the previous subsection with a sample size of 40 and a tolerable exception rate of 0.1 (TER = 0.1, i.e., B = [0, 0.1]). Now, suppose that the auditor has found just one occurrence of the attribute in the sample, i.e., k = 1. This means that the auditor has achieved the level of belief as planned. This can be seen by calculating the belief in the interval B = [0, 0.1] using (14) which yields 0.8236 degree of belief.

Let us consider another situation. Suppose the auditor finds two occurrences of the attribute in the sample (k = 2) in the above case. Using (14), the belief in B = [0, 0.1] is only 0.4874. This is significantly below what the auditor had planned to achieve. There are several options available to the auditor including: (1) Increase the sample size based on the information that a = 2 and reperform the test procedures for the remaining sample units. (2) Think of the consequences of the weakness and see if there are any other controls that would compensate for this weakness and take that into consideration. (3) Do not depend as much on the control system as originally planned and increase the tests of details of the account balance. In the first alternative, with a new estimate of k =2, the sample size obtained by solving (17) is 56[12]. The auditor can select additional sample units to perform the test procedures and if he or she does not find any further occurrence of the attribute then the belief in the interval is as desired.

--- Figure 1 ---

Figure 1 shows graphs of Bel(B) and Bel(notB) as a function of the number of occurrences in the sample where B is the interval defined by [0, TER] for TER = 0.1, and n = 40, 60, 80, and 100. Here are some interesting observations about the assessed beliefs in different situations that make intuitive sense: (1) Bel(B) is maximum at k = 0 for all n's. This maximum increases with the increase in n. At large n with k = 0, the Bel(B) is almost unity. This simply means that the control is definitely effective. (2) Bel(B) decreases as k increases and becomes zero at the upper end of the interval and beyond (i.e., Bel(B) = 0 for TER = k/n). (3) Bel(notB) is zero within the interval (i.e., Bel(notB) = 0 for k/n = TER), however, it is non-zero outside the interval when k/n > TER and increases with the increase in k, reaching unity for a large value of k. This simply means that the control is definitely not effective at all when the sample occurrence rate is much higher than the tolerable exception rate (i.e., TER << k/n). (4) At the upper end of the interval (i.e., TER = k/n), both Bel(B) and Bel(notB) are zero. As discussed earlier, this simply means that the auditor is ignorant about the state of the internal control; the auditor lacks evidence in support of or against the control being effective.

Auditors commonly evaluate attribute samples by establishing an upper precision limit for control exceptions at a fixed level of reliability such as 90 percent (Arens and Loebbecke 1994). Table 3 shows the upper precision limits, p_2, for intervals $[0, p_2]$ with 90% belief for various combinations of sample sizes and exceptions found, together with comparable upper precision limit calculated using the equivalent probability framework. Note that the intervals are identical when there are no exceptions found. However, when control exceptions are found the intervals for 90 percent belief are wider than the equivalent 90 percent reliability intervals in the probability framework. In some sense, then, a 90 percent belief is a stronger condition than 90 percent reliability. Nevertheless, the upper precision limits generated by the two different frameworks are comparable in magnitude.

--- Table 3 here ---

4.3. Integrating Statistical and Non-Statistical Audit Evidence

Once we are able to assess beliefs from the statistical evidence, its integration with the non-statistical evidence becomes in general straightforward. The standard way will be to combine the beliefs obtained from two types of evidence using Dempster's rule (see, e.g., Shafer 1976, Shafer and Srivastava 1990, and Srivastava 1993). There are two situations where the auditor will integrate such items of evidence. One when he or she is planning the audit and the other when evaluating the audit results. In each case, the auditor may wish to combine environmental evidence regarding effectiveness of controls with the results of attribute sampling. Moreover, the auditor will wish to combine evidence relating to controls with substantive evidence regarding the financial statement assertions. The problem remaining to be addressed is that although the methods outlined above will enable beliefs to be developed based on the results of attribute sampling, the different pieces of evidence the auditor wishes to combine do not have the same frame of discernment. The attribute sampling gives a belief x on an interval $[p_1, p_2]$ for the failure rate of controls. The substantive evidence, for example, will give a belief for an interval of tolerable misstatement in a FS amount (see, e.g., Srivastava and Shafer 1994). Since the main focus of a FS audit is just such misstatement, we need to convert beliefs in other intervals to the same frame before they can be combined.

This problem exists regardless of whether belief functions or probability theory are used for attribute sampling. There is little or no empirical evidence as to how auditors do this in practice even within a probability framework. The balance between subjective assessment by individual auditors and the mandates of firm policy in this area are also unresearched. For example, a sample of 10 in which no exceptions are found gives rise to an interval of [0, 20.6%] within both frameworks, as shown in Table 3. This might be interpreted as providing only limited reliance on controls (i.e., control risk might be assessed as slightly less than maximum). In one firm's approach this corresponds to a control risk of 56.2 percent (Grant Thornton 1990). Similarly, a sample of 50 in which two exceptions are found gives 90 percent belief in an interval of [0,12.8%], corresponding to a 90 percent reliability interval of [0, 10.3%] in a probability framework, and warrants maximum reliance on controls (i.e., a control risk of 13.3 percent, see Grant Thornton 1990).

Analytically, given a belief x in the interval $[p_1, p_2]$ for the control exception rate, we require a function F such that the belief z in the interval for tolerable misstatement is given by:

$$z = F(x, p1, p2, \underline{\theta}) \qquad (19)$$

where $\underline{\theta}$ is a vector of other relevant parameters.

A number of other parameters are likely to be relevant to this function; for example: sample size (n), tolerable error (TE), and book value of the population (BV). As an illustration, a function is given below which depends also on a failure rate (r) describing the rate at which the auditor believes misstatements are likely to have occurred in the presence of exceptions. In other words, not every control exception will necessarily relate to an error in the FS.

We use as an illustration the function:

$$z = x.(1- p_2).\left[1 - \exp\left\{ - \frac{n.(TE - \min(TE, BV.\ p_2.r))^2}{2.BV^2.p_2.r.(1 - p_2.r)}\right\}\right] \qquad (20)$$

for which certain values are shown in Table 4.

--- Table 4 ---

Suppose in the example we were considering earlier that based on one exception in a sample of 40 items the belief in the interval [0, 0.1] for control exceptions was 0.8236, and the auditor's tolerable error for a book value of $1,000,000 was $50,000. Suppose also that the auditor believed that there would be only one error for every ten control exceptions (i.e., r = 0.1). Then from Table 4, the belief in the financial statement interval would be 0.6916. This belief can now be combined with the belief in the same interval derived from other audit evidence such as substantive procedures.

5. SUMMARY AND CONCLUSION

We have demonstrated how beliefs can be assessed from the statistical evidence in attribute sampling. We have illustrated using an auditing example how to determine the sample size for a desired level of belief in an interval B = [0, TER], and what level of belief is obtained when the sample results are analyzed, with TER representing the auditor's judgment about the highest occurrence rate tolerable. As expected, the sample size increases as the desired level of belief in the interval increases. Also, the sample size increases as the number of occurrences expected in the sample increases. Furthermore, when TER increases the sample size decreases because the auditor is willing to take more risk by increasing the interval for the same level of belief. The analyses of the sample findings also show results that make intuitive sense. For example, we find that Bel(B) increases as the sample size increases for the same number of occurrence of the attribute. Also, Bel(B) is maximum when the number of occurrence, k, in the sample is zero. It is interesting to note that both Bel(B) and Bel(notB) are zero at the upper end of the interval (i.e., for k = n*TER). This implies that the auditor has no evidence in support of or against the control being effective.

We have discussed the problem of converting belief in an interval for control exceptions into belief in a financial statement interval so that it can be combined with evidence from other sources.

There several issues related to integrating statistical and non-statistical evidence in attribute sampling that we have not covered here in the article. These include: (1) How do we deal with dependencies among such items of evidence? (3) How to evaluate what level of belief is obtained from the nonstatistical evidence, i.e., what are factors that determine the level of belief? (4) What relationship does exist between an effective control and its impact on the fair presentation of the related account balance? Further research is needed in these areas which are important for real decisions.

FOOTNOTES

1. In general, there are four types of audit: audit of the financial statements, tax audit, compliance audit, and operational audit. We will use examples from financial audits. However, in principle, all audits involve the same basic process of accumulation, evaluation, and aggregation of evidence to express an opinion.

2. These assertions are known as management assertions. The accounting profession classifies them into five categories: Existence or occurrence, Completeness, Valuation or allocation, Rights and obligations, and Presentation and disclosure (AICPA 1980(SAS 31)).

3. Such a tests is called 'test of control (TC)'. Usually, a TC simply determines whether the control is in operation or not and rarely tests for the effectiveness of the control. For example, performing an attribute test that each sales order has the credit manger's signature does not tell how effective the credit approval is; it simply tells that the manager is signing all the sales orders. Whereas, in the case of testing for segregation of incompatible functions, an observation of the process on a surprise basis to see who does what will give the auditor certain level of confidence that the control is in operation and effective. Auditors usually perform tests of transactions to determine the effectiveness of controls (see footnote 4).

4. Such a test is known as a test of transactions (TT). It determines the effectiveness of the control by reperforming the control procedures to see how effective it has been.

5. It should be pointed out that a test of transaction (TT) relevant to a control process provides direct beliefs to the control that it is effective or not effective. However, a test of control (TC) may not necessarily provide beliefs directly to the control being effective or not. The auditor will have to make a judgment based on the nature of control and the level of its operation (see the example in footnote 2).

6. In order to facilitate accumulation of evidence in meeting the overall objective of the audit that the FS present fairly the financial position of the company, the auditing profession has developed a set of audit objectives. There are, in general, nine audit objectives: Existence, Completeness, Accuracy, Classification, Cutoff, Detail tie-in, Realizable value, Rights and Obligations, Presentation and Disclosure, related to each account on the balance sheet (see, e.g., Arens and Loebbecke 1994). These audit objectives are directly related to the management assertions listed in footnote 2. It is assumed that (1) the FS are fairly presented if all the accounts are fairly stated, (2) each account is fairly stated if all its management assertions are met, and (3) each assertion is met if the related audit objectives are met.

7. When the number of occurrences in the sample is less than or equal to the critical value, we accept the null hypothesis to be true and when it greater, we reject the null hypothesis.

8. Renormalization is done by dividing the likelihood function by its maximum value, so that the new maximum is one.

9. The likelihood function for the binomial distribution is given by

$$L(P|k,n) = \frac{n!}{k!(n-k)!}.P^k(1-P)^{n-k},$$

which has a maximum at P = k/n, i.e.,

$$L_{max}(P = \frac{k}{n}|k,n) = \frac{n!}{k!(n-k)!}.(\frac{k}{n})^k(1-\frac{k}{n})^{n-k},$$

The renormalized likelihood function is obtained by dividing the above maximum value into the likelihood function. The result is:

$$f(P) = \left(\frac{nP}{k}\right)^k\left(\frac{n-nP}{n-k}\right)^{n-k}.$$

For Poisson distribution, the likelihood function is given by:

$$L(P|kn) = Exp(-nP).\frac{(nP)^k}{k!}.$$

This function has a maximum at P = k/n which yields the following expression for f:

$$f(P) = Exp(-nP+k).\left(\frac{nP}{k}\right)^k.$$

10. For $p_1 = 0$, the set 'notB' consists of only those values of P that are greater than p_2. This condition simplifies (12) to (14).

11. For k = 0, the likelihood function for the binomial distribution becomes $L(P) = (1 - P)^n$ which is also the function f.

12. A more rigorous calculation using sequential sampling techniques would be used in practice, and would give a larger value.

REFERENCES

Akresh, A. D., J. K. Loebbecke, and W. R. Scott. 1988. Audit Approaches and Techniques. *Research Opportunities in Auditing: The Second Decade*. edited by A. R. Abdel-khalik and Ira Solomon. Sarasota, FL:AAA.13-55.

American Institute of Certified Public Accountants. 1980. *SAS 31: Evidential Matter.* 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.* Hracourt Brace Jovanovich.

Cochran, W. 1977. *Sampling Techniques. Third Edition.* John Wiley & Sons, New York.

Gillett, P. R. 1993. Automatic Dynamic Audit Programme Tailoring: An Expert Systems Approach. *Auditing: A Journal of Practice and Theory* (Supplement): 173-89.

Grant Thornton. 1990. *Audit Manual.* Longman Group UK, London.

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. P.P. Shenoy, and R.P. Srivastava. 1988. AUDITOR'S ASSISTANT: An Knowledge Engineering Tool for Audit Decisions. *Proceedings of the 1988 Touche Ross/University of Kansas Symposium on Auditing Problems.* Lawrence, KS: School of Business, University of Kansas, 61-84.

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. 1995. A General Scheme for Aggregating Evidence in Auditing: Propagation of Beliefs in Networks. *Artificial Intelligence in Accounting and Auditing,* Vol. 2, Miklos A. Vasarhelyi, editor, Markus Wiener Publishers.

Srivastava, R.P. S. Dutta and R. Johns. 1994. An Expert System Approach to Audit Planning and Evaluation in the Belief-Function Framework. Working Paper, School of Business, The University of Kansas. An earlier version of the paper by Srivastava and Johns was presented at the *1992 American Accounting Association Workshop on AI/ES in Accounting, Auditing and Tax*, Washington, DC., August 1992, and also presented at the *Conference on Expert Systems in Auditing, Accounting and Telecommunications*, Rutgers University, Newark, November 20-21, 1992.

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

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

Table 1

Sample Size for Attribute Sampling Using Binomial Distribution for a Desired Belief in the Interval B = [0, TER] with TER = 0.10

Expected Number of Occurrence in the Sample (k)	Desired Level of Belief in the Interval [0, 0.10]	Sample Size n From (17) and (18)	Corresponding Planned Power of the Test (1-β), From (1)
0	0.95	28	0.95
	0.90	22	0.90
	0.80	15	0.80
	0.70	11	0.70
	0.60	9	0.60
	0.50	7	0.50
	0.40	5	0.40
	0.30	3	0.30
1	0.95	56	0.9802
	0.90	47	0.9560
	0.80	39	0.9124
	0.70	34	0.8671
	0.60	30	0.8163
	0.50	26	0.7487
	0.40	24	0.7075
	0.30	21	0.6353
2	0.95	74	0.9825
	0.90	65	0.9640
	0.80	56	0.9281
	0.70	50	0.8883
	0.60	45	0.8410
	0.50	41	0.7914
	0.40	37	0.7297
	0.30	34	0.6745
3	0.95	92	0.9855
	0.90	82	0.9695
	0.80	71	0.9335
	0.70	64	0.8937
	0.60	59	0.8537
	0.50	54	0.8015
	0.40	50	0.7497
	0.30	47	0.7044

Table 2

Sample Size for Attribute Sampling Using Binomial Distribution for a Desired Belief in the Interval B = [0, TER] with TER = 0.05

Expected Number of Occurrences in the Sample (k)	Desired Level of Belief in the Interval [0, 0.05]	Sample Size n From (17) and (18)	Corresponding Planned Power of the Test (1-β), From (1)
0	0.95	58	0.95
	0.90	45	0.90
	0.80	31	0.80
	0.70	24	0.70
	0.60	18	0.64
	0.50	14	0.50
	0.40	10	0.40
	0.30	7	0.30
1	0.95	113	0.9789
	0.90	96	0.9560
	0.80	79	0.9103
	0.70	68	0.8601
	0.60	60	0.8085
	0.50	53	0.7500
	0.40	47	0.6883
	0.30	42	0.6276
2	0.95	151	0.9826
	0.90	133	0.9648
	0.80	113	0.9256
	0.70	100	0.8817
	0.60	91	0.8391
	0.50	82	0.7837
	0.40	75	0.7303
	0.30	68	0.6672
3	0.95	186	0.9847
	0.90	166	0.9686
	0.80	144	0.9330
	0.70	130	0.8942
	0.60	119	0.8511
	0.50	110	0.8055
	0.40	102	0.7560
	0.30	94	0.6972

Table 3

Upper Precision Limit, p_2, for the Interval [0, p_2] having 90% Belief compared with Upper Precision Limit for 90% Reliability (10% Risk of Overreliance) in probability framework.

Sample Size (n)	Number of Errors in the Sample	Upper Precision Limit, p_2, for a belief of 0.90 in the interval [0, p_2]	Upper Precision Limit for 10% Risk of Overreliance
10	0	20.6%	20.6%
20	0	10.9%	10.9%
	1	22.2%	18.1%
30	0	7.4%	7.4%
	1	15.3%	12.4%
	2	20.7%	16.8%
40	0	5.6%	5.6%
	1	11.6%	9.4%
	2	15.9%	12.8%
	3	19.6%	16.0%
	4	23.1%	19.0%
50	0	4.5%	4.6%
	1	9.4%	7.6%
	2	12.8%	10.3%
	3	15.9%	12.9%
	4	18.7%	15.4%
	5	21.4%	17.8%
100	0	2.3%	2.3%
	1	4.8%	3.9%
	2	6.6%	5.3%
	3	8.2%	6.6%
	4	9.7%	7.9%
	5	11.1%	9.1%
	6	12.5%	10.3%
	7	13.8%	11.5%
	8	15.1%	12.7%
	9	16.4%	13.9%
	10	17.6%	15.0%

Table 4

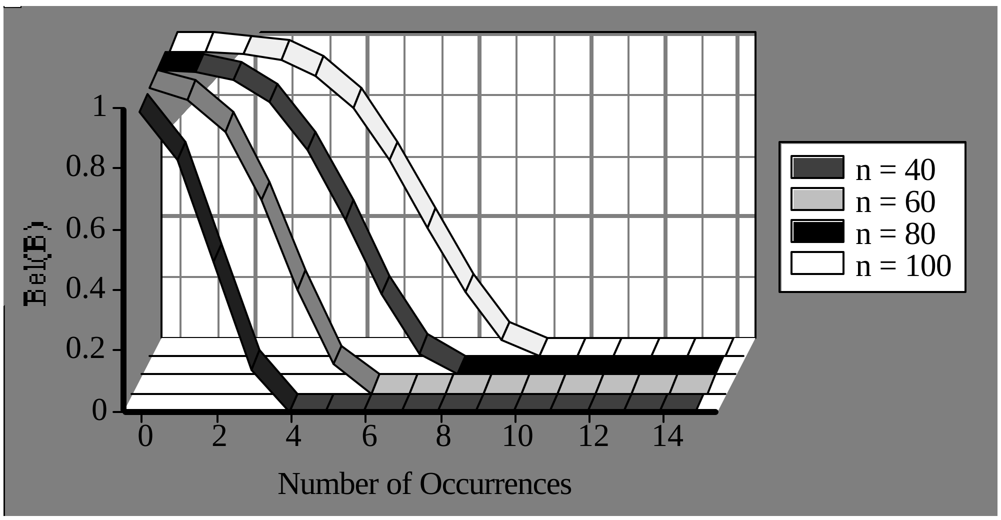
Belief in the Interval for Tolerable Error in the Financial Statement Amount $1,000,000 Based on the Interval [p_1, p_2] for Control Exceptions with p_1 = 0, p_2 = 0.1, and Sample Size of 40.

TE	x	r	$x^*(1-p_2)$	Bel(fs)
50,000.00	0.8	1	0.72	0
100,000.00			0.72	0
200,000.00			0.72	0.6420
500,000.00			0.72	0.7200
1,000,000.00			0.72	0.72
50,000.00	0.8	0.5	0.72	0
100,000.00			0.72	0.4687
200,000.00			0.72	0.7199
500,000.00			0.72	0.7200
1,000,000.00			0.72	0.7200
50,000.00	0.8	0.1	0.72	0.6916
100,000.00			0.72	0.7200
200,000.00			0.72	0.72
500,000.00			0.72	0.72
1,000,000.00			0.72	0.72
50,000.00	0.8	0.05	0.72	0.7199
100,000.00			0.72	0.7200
200,000.00			0.72	0.72
500,000.00			0.72	0.72
1,000,000.00			0.72	0.72
50,000.00	0.8	0.01	0.72	0.72
100,000.00			0.72	0.72
200,000.00			0.72	0.72
500,000.00			0.72	0.72
1,000,000.00			0.72	0.72

Figure 1

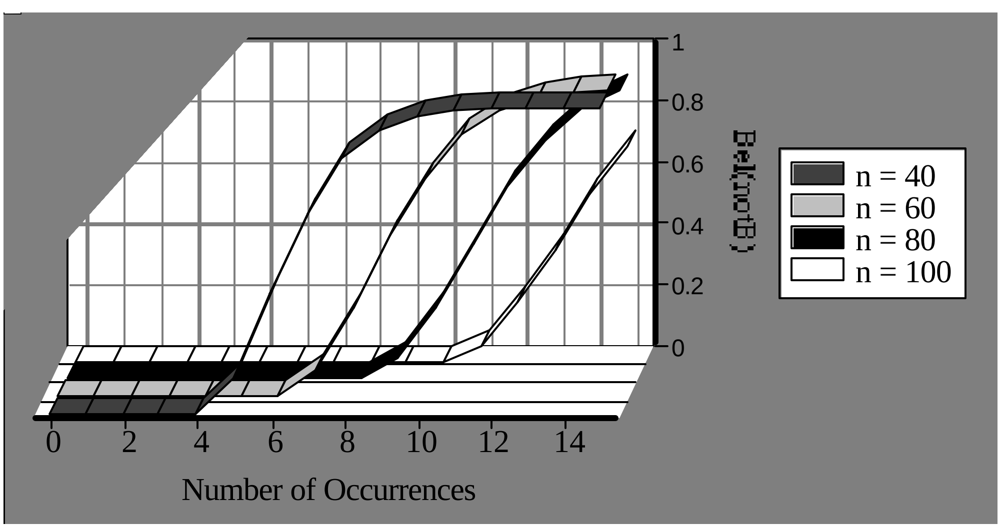
Bel(B) and Bel(notB) for B = [0, 0.10] as a Function of the Number of Occurrences* in the Sample for Different Values of the Sample Size n.

Panel A: Belief in B



- -

Panel B: Belief in notB



*This number takes only integer values and thus the distributions shown in the figure are all discrete. However, the lines are drawn just to separate the data points for each n.