Indian Journal of Accounting, Vol. XXVII, June 1996, pp. 4-18.

A BAYESIAN PERSPECTIVE ON THE STRENGTH OF EVIDENCE IN AUDITING

Saurav K. Dutta
Assistant Professor
Graduate School of Management
Rutgers University-Newark, Newark, N.J.

and

Rajendra P. Srivastava
Associate Professor and
Deloitte & Touche Faculty Fellow
School of Business
University of Kansas
Lawrence, Kansas 66045

(January 28, 1992)

The authors would like to thank Ed Blocker, Peter Gillet, Glenn Shafer, Tim Shaftel, and Prakash Shenoy for their insightful comments on an earlier version of this paper.

A BAYESIAN PERSPECTIVE ON THE STRENGTH OF EVIDENCE IN AUDITING

SYNOPSIS

This paper deals with a mathematical definition of the strength of evidence in the Bayesian framework for the (1) positive evidence, (2) negative evidence, and (3) confirming evidence. These concepts are important in auditing. The earlier definitions of these concepts by Toba and Kissinger depend on the prior probability and the posterior probability and thus lead to inconsistencies. The definitions discussed in this paper are based on the likelihood ratio and thus depend on the intrinsic properties of the evidence and therefore are free from the inconsistencies encountered in the Toba-Kissinger framework.

We have extended the concept of confirming evidence to a set of evidence which is more relevant in the auditing context. Further, we have used a simple example to demonstrate how aggregation of various items of evidence on an audit can be facilitated by using likelihood ratio as a measure of the strength of evidence.

Key Words: Strength of evidence; likelihood ratio; auditing; aggregation; evidence; confirming.

A BAYESIAN PERSPECTIVE ON THE STRENGTH OF EVIDENCE IN AUDITING

We have two objectives in this paper. First, we want to develop a mathematical definition for the strength of evidence in the Bayesian framework that depends on the intrinsic properties of the evidence. We extend this definition to describe (1) positive evidence, (2) negative evidence, and (3) confirming evidence. Second, we want to show how the above definition facilitates the evidence aggregation process for planning and evaluating an audit.

In auditing, it is important that an auditor determine the nature and strength of evidence, i.e., he must know whether an item of evidence supports or refutes an assertion. In addition, he must also assess how strongly the evidence supports or refutes the assertion. In fact, there have been several recent studies that deal with the notions of positive and negative, and strong and weak items of evidence (see, e.g., Ashton & Ashton 1988; Tubbs, Messier and Knechel 1990). These studies, however, do not mathematically define these notions. A mathematical definition of these notions is imperative to help auditors perform an effective and efficient audit. Also, such a definition should help in constructing decision support systems for auditing.

Toba (1975) in the original work on the theory of evidence dealt with some of the above notions. Later, Kissinger (1977) extended Toba's work by eliminating several oversights and conceptual difficulties. Unfortunately, as shown in Section I, even the modified theory by Kissinger (hereafter referred to as Toba-Kissinger (TK) framework) is found to be deficient in dealing with the above notions. For example, the TK framework is based on the prospective evidence, i.e., on the posterior probability. In this framework, the auditor has no knowledge of the strength of evidence at the time of planning because the evidence has not yet been observed and hence the posterior is not known.

Furthermore, as shown in Section I, the strength of evidence in TK framework depends on the prior probability of the assertion, a factor extrinsic to the evidence. This in turn leads to the following anomalies: (1) A positive item of evidence, however weak, is confirming when the prior is greater than 0.5. (2) An item of evidence may be confirming for a high prior but may not be confirming for a low prior. (3) The strength of evidence depends on the order in which the evidence is obtained. (4) An item of evidence which is confirming for the negation of an assertion may become confirming for the assertion when combined with another independent item of evidence. Thus, the TK framework which provided the initial conceptual foundation for the theory of evidence in auditing is found to be inappropriate for defining the strength of evidence. In principle, the strength of evidence should not depend on the prior probability of the assertion or on the order in which the evidence is collected, rather it should depend on the intrinsic properties of the evidence.

After reviewing the literature in philosophy, mathematics and law, we find that there are four types of measure of the strength of evidence (Friedman 1986; Lampert 1977; Jeffrey 1983; Cullison 1969; Hacking 1965; Schum & Martin 1982; Tribe 1971; Good 1950; Minsky & Selfridge 1961; Barnett 1982). These measures are based on (1) the posterior and prior; (2) the difference of likelihoods; (3) the likelihood ratio, and (4) the log-likelihood ratio. The first two measures suffer from the same deficiencies as the TK framework and will not be discussed in the paper[1]. The last two measures, unlike the TK framework, define various notions of positive, negative, strong, and weak items of evidence without any anomalies.

The remainder of the paper is divided into five sections and one appendix. In section I, we discuss the anomalies with the TK framework. In Section II, we discuss the measure of the strength of evidence based on the likelihood ratio and show that this measure is free from the anomalies encountered in the TK framework. Also, in Section II, we discuss the notions of positive, negative, weak, and strong items of evidence based on the likelihood ratio. In Section III, we discuss the notion of confirming evidence. In Section IV, we demonstrate how the strength of evidence based on the likelihood ratio can be used to aggregate evidence in evaluating as well as planning an audit. We use a numerical example to illustrate the process. In Section V, we present a brief summary and conclusions. Finally, in the Appendix, we discuss a general approach to aggregating evidence in auditing.

I. TOBA-KISSINGER FRAMEWORK

Toba (1975) defined positive evidence as an item of evidence, which when obtained increases the belief in the assertion. In other words, if E represents an item of evidence and a represents the assertion, then according to Toba, E is positive if the posterior is greater than the prior, i.e., $P(a|E) > P(a)$. Also, according to Toba an item of evidence E is confirming if the evidence renders the assertion more likely than its negation, i.e., $P(a|E) > P(\sim a|E)$. However, as pointed out by Kissinger (1977), the above definition of confirming evidence was deficient and led to internal inconsistencies. For example, an item of evidence while positive for an assertion may be, at the same time, confirming the negation of the assertion[2].

Kissinger modified the definition of confirming evidence to eliminate the above anomaly by defining an item of evidence E to be confirming for an assertion a if $P(a|E) > P(a)$ as well as $P(a|E) > P(\sim a|E)$. Thus, an item of evidence had to be positive in order to be confirming for the assertion. This modification removes the earlier mentioned anomaly, but is still fraught with many problems of its own, as discussed below.

First, the characterization of evidence depends on the prior probability, i.e., an item of evidence may only be supporting (positive) an assertion when the prior probability is low, but it may be confirming the assertion when the prior probability is high. For example, suppose the prior is 0.2 and the evidence obtained is positive and weak which yields 0.3 for the posterior. Then, such an item of evidence is just supporting the assertion but not confirming it[3]. However, if the prior were 0.7, then the posterior for the same evidence considered above would be 0.76[4]. In this case, the evidence is confirming the assertion, a strong item of evidence. In fact, any positive evidence, however weak, is a confirming item of evidence when the prior is 0.5 or greater. Thus, in the TK framework, the characteristic of an item of evidence that whether the evidence is confirming or not confirming and whether it is strong or weak, is based on the prior probability of the assertion. The prior is not an intrinsic property of the evidence, rather it is a subjective opinion of the observer.

Second, in the TK framework, the order in which an item of evidence is collected determines whether the evidence is confirming or not. Consider the scenario in which the auditor receives two items of positive evidence. Suppose the prior probability on the assertion a that 'the accounts receivables are collectible' is 0.2, i.e., $P(a) = 0.2$. The auditor then learns that the major debtors of the client are in good financial position. Suppose this item of evidence, say E_1, increases the probability of collectibility to 0.4, i.e., $P(a/E_1) = 0.4$. Further, the auditor learns that for the last six months the collection has been good and there were very few bad debts. The second item of evidence, say E_2, further increases the probability of collectibility to 0.64[5], i.e., $P(a/E_1\&E_2) = 0.64$. In this scenario the evidence E_1 is positive because $P(a/E_1) > P(a)$, whereas the evidence E_2 is confirming since $P(a/E_1\&E_2) > P(a/E_1)$ and $P(a/E_1\&E_2) > P(\sim a/E_1\&E_2)$.

Now suppose the sequencing of the items of evidence is reversed. The auditor first learns that for the last six months the collection has been good and there were very few bad debts. Next, the auditor learns that the major debtors of the client are in good financial position. In other words, the auditor first obtains evidence E_2 and then receives evidence E_1. After receiving evidence E_2 the posterior becomes 0.4, i.e., $P(a/E_2) = 0.4$. Next, when he receives evidence E_1 then the new posterior becomes 0.64, i.e., $P(a/E_2\&E_1) = 0.64$. In this case, evidence E_2 is positive, whereas evidence E_1 is confirming. Thus, it appears from the above discussion that the order in which an item of evidence is collected and processed determines whether the evidence is confirming or not confirming.

The third anomaly with the TK framework is that an item of evidence could be confirming for $\sim a$, but when combined with another item of evidence could become confirming for a. For example, an auditor wishes to ascertain that the internal controls in the sales-collection cycle are good. He has a prior probability of, say, 0.6, i.e., $P(a) = 0.6$, and $P(\sim a) = 0.4$. The auditor learns that the company receives cash in the mail in addition to checks, and the receptionist opens all the mail. This is a sign of poor internal control and thus the auditor revises his belief to 0.45 on a, i.e., $P(a|E_1) = 0.45$, and $P(\sim a|E_1) = 0.55$. Thus, E_1 is confirming evidence for $\sim a$ in the TK framework, since $P(\sim a|E_1) > P(a|E_1)$ and $P(\sim a|E_1) > P(\sim a)$. Next, the auditor learns that the functions of receiving, depositing, and recording cash receipts are properly segregated. Also, he finds that monthly statements are sent to the customers (debtors) on a regular basis and any complaints on the account are promptly resolved. All this is a sign of good internal control. We will denote this set of evidence as E_2. Thus, the auditor revises his belief on a to, say, 0.61, i.e., $P(a|E_1\&E_2) = 0.61$. This situation makes the two items of evidence considered together confirming for a in the TK framework, since $P(a|E_1\&E_2) > P(\sim a|E_1\&E_2)$ and $P(a|E_1\&E_2) > P(a)$. Thus, evidence E_1 is confirming for $\sim a$, and the same evidence in conjunction with E_2 is confirming for a. This seems counter intuitive.

Further, the nature of evidence in auditing is retrospective and not prospective. That is, the evidence collected after the fact is not an indicator of the event, but a consequence of it. Consider the following internal control procedure, proper authorization of credit sales. The auditor wishes to ascertain whether the procedure was being followed. In order to do so, he examines a sample of approved sales orders and verifies the authorized signature and the credit limit. Based on his sample check of approved sales orders he will surmise whether the internal control procedure was being followed. The sample check is a retrospective evidence[6] for internal control procedure being followed. The TK framework is based on comparing the prior with the posterior, but the auditor has no knowledge of the posterior until the evidence is collected.

In summary, we have identified many anomalies in the TK framework. The characterization of evidence in the TK framework is based on extrinsic factors such as the prior and the order in which the evidence is gathered. Such a characterization of evidence leads to many inconsistencies as discussed earlier. Moreover, the framework compares the prior with the posterior, but the auditor has no knowledge of the posterior until the evidence is collected and assessed. The lack of a-priori knowledge of the posterior inhibits the use of this framework for planning purposes. Thus, the TK framework with its anomalies has limited application in auditing.

II. MEASURE OF THE STRENGTH OF EVIDENCE BASED ON THE LIKELIHOOD RATIO

Likelihood is defined as the probability of obtaining the evidence given the assertion. That is, $P(E|a)$ and $P(E|\sim a)$, are the likelihoods, where E is the evidence and a is the assertion. Fisher introduced the name likelihood (Fisher 1922). Since then, several writers have canvassed a 'likelihood principle', which states, that in assessing hypotheses in light of an item of evidence, only likelihoods count (see, e.g., Barnard 1947; Edwards 1982; Edwards 1984; Savage 1961; Birnbaum 1962). Commenting on the importance of likelihood, Savage (1961) states that "Given the likelihood function in which the experiment resulted, everything else about the experiment ...is irrelevant." Good (1983) also asserts that the likelihoods have "sharp uncontroversial values." Since the likelihoods are intrinsic to the evidence, a measure of the strength of evidence based on the likelihood ratio will also depend on the intrinsic properties of the evidence. Further, it is cognitively simpler to construct probability assessments in terms of likelihoods (Kuippers & Kassier 1984). Thus, a measure based on likelihoods will not only capture the intrinsic properties of the evidence, but also will be easier to elicit. Such a measure will be helpful in constructing decision support systems and expert systems in auditing.

According to the likelihood measure, the strength of evidence, λ, equals the likelihood ratio (see, e.g., Hacking 1965; Lampert 1977; Schum & Martin 1982; Tribe 1971). The likelihood ratio for evidence E (λ_E) is defined as:

$$\lambda_E = \frac{P(E|a)}{P(E|\sim a)}. \tag{1}$$

This definition has a simple multiplicative property. If E_1 and E_2 denote two independent items of evidence, then the combined strength of E_1&E_2 is the product of the individual strengths[7], that is, $\lambda_{E_1\&E_2} = \lambda_{E_1} x \lambda_{E_2}$. We will use this representation instead of the log-likelihood ratio (discussed later in this section) for the strength of evidence because it yields simpler forms when various items of evidence are aggregated in a complex situation (See the Appendix).

Positive, Neutral, and Negative Items of Evidence

A positive item of evidence is defined by a likelihood ratio of greater than one, i.e., $P(E|a)/P(E||\sim a) > 1$. This definition implies that the posterior is always greater than the prior[8] for a positive item of evidence. This was the basic definition used in the TK framework. But in the present case it is the consequence of the definition. The higher the likelihood ratio, the stronger the positive evidence. In the limit when the likelihood ratio goes to infinity, the evidence becomes confirming evidence. This topic will be further discussed in Section III.

A likelihood ratio of unity implies that the evidence is neutral, i.e., the evidence does not contribute any new knowledge about the assertion. For example, since $P(E|a)/P(E|\sim a) = 1$, Bayes' rule will yield $P(a|E)P(\sim a) / P(\sim a|E)P(a) = 1$. This expression simplifies to $P(a|E) = P(a)$ which implies that the evidence has no effect on the prior. Thus, likelihood ratio equal to one denotes irrelevant evidence, that is, attaining such an item of evidence will not affect our belief on the assertion.

Negative evidence is defined by a likelihood ratio of less than one. In such cases, the posterior decreases after the evidence has been observed, i.e., $P(a|E) < P(a)$. This property again is a consequence of the definition and not the definition as used in the TK framework. The following discussion shows how the definition of the negative evidence leads to the property $P(a|E) < P(a)$. Since $P(E|a)/P(E|\sim a) < 1$, from Bayes' rule one obtains $P(a|E)P(\sim a) / P(\sim a|E)P(a) < 1$. This expression yields $P(a|E) < P(a)$. The stronger the negative evidence the closer the likelihood ratio to zero.

Is Likelihood Ratio as a Measure of Strength of Evidence Free From Anomalies?

The measure of the strength of evidence based on the likelihood ratio is free from the anomalies encountered by the TK framework. As seen in Figure 1, the posterior probabilities are higher for stronger evidence. When the likelihood ratio is less than one, the evidence is negative. Thus, there is a decrease in probability. The stronger the negative evidence, the lower the posterior probability. The evidence with a likelihood ratio of 0.2 is stronger negative than the evidence with likelihood ratio of 0.4. The evidence with a likelihood ratio equal to one is neutral evidence. It does not render any proposition more likely than the other. For items of evidence with likelihood ratios greater than one, the posterior is greater than the prior. The evidence with likelihood ratio of 6 is stronger than the evidence with likelihood ratio of 2.5. The curve in Figure 1 denoting the posterior for a stronger evidence is higher than the curve denoting the moderate evidence. Unlike the TK framework, one can easily show that the strength of evidence based on the likelihood ratio does not change with the order in which the evidence is processed.

Figure 1 here

Measure Based on Log-Likelihood Ratio

Another very prevalent measure of the strength of evidence is the logarithm of the likelihood ratio (Good 1950; Minsky & Selfridge 1961; Barnett 1982).

$$\text{Str (Evidence)} = \ln \text{(Likelihood Ratio)} \quad (2)$$

With this definition, the strength is additive: Str (E_1&E_2) = Str (E_1) + Str (E_2). The advantage of the logarithmic measure, over the measure based on likelihood ratio, is that it measures the strength of negative evidence and positive evidence on equal scales. The range of log- likelihood ratios for positive evidence is from 0+ to +∞. The range of log- likelihood ratios for negative evidence is from 0- to - ∞. In contrast, the likelihood ratio measures positive evidence on a scale of 1+ to ∞, but it compresses the negative evidence to a range of (0,1).

Positive log- likelihoods denote positive evidence. The higher the number, the stronger the evidence. Negative log- likelihoods denote negative evidence. The higher the absolute value, the stronger the negative evidence. Log-likelihood ratios of zero denote irrelevant or neutral evidence.

The likelihood ratio or the log-likelihood as measures of the strength of evidence are more viable than the Toba-Kissinger measure or the difference of likelihoods measure. The likelihood ratios do not refer to the prior odds. They capture the "intrinsic" strength of the evidence. The strength of the evidence is the same whether the evidence is introduced in support of an assertion that has been shown to be highly likely or in support of an initially implausible assertion. The strength of evidence, thus measured, is simply a function of evidence itself, and not of extrinsic factors such as the order and the prior. In addition, the probability models based on likelihoods require assessment of fewer parameters and are 'portable' (Shachter & Heckerman 1987). Thus, in addition to being theoretically sound, likelihood ratio is also a more viable measure of the strength of evidence from a decision support/expert system perspective.

III. CONFIRMING EVIDENCE

Prior to mathematically defining confirming evidence, it is important to distinguish between the two statements: (1) an item of evidence E is confirming for the assertion a, and (2) the assertion a is confirmed after observing the evidence E. The two statements are not equivalent, but they have been erroneously interpreted as being equivalent. Instead, we perceive we should interpret the two statements as follows: The first relates to an item of evidence or a set of items of evidence which, when observed, will provide confirming support that the assertion is met, irrespective of the prior probability of the assertion. Such an item of evidence (or a set of items of evidence) we will define to be confirming evidence, whereas the second statement implies that, given the prior, observing an item of evidence (or a set of items of evidence) provides enough support that the assertion is confirmed. The first statement deals with the strength of evidence in isolation, whereas the second one deals with the overall support, after observing the evidence, that the assertion is met. Both of these concepts are important for the auditor; the first one is needed for planning and the second one is needed for evaluation and to determine the extent of additional work that has to be performed in order to establish the assertion.

The above concepts have been studied by many researchers in various disciplines including philosophy, mathematics, and law. In general, the empiricists have addressed the problem of how a hypothesis can be confirmed by an item of evidence by devising various strategies that relate evidence to theory: Elimination of theory, Deductive method, Bootstrap method, and Probabilistic strategies (Glymour 1980). For the purpose of this paper, we will employ probabilistic strategy for defining confirming evidence[9].

We will first discuss the mathematical, i.e., the ideal definition of a confirming item of evidence and then later discuss an acceptable definition of when an assertion is established to be confirmed.

A confirming item of evidence is defined such that it is confirming for all priors, no matter how small. The notion of confirming evidence is regarded as an intrinsic property of the evidence and does not depend on extraneous factors as used in the TK framework. Figure 2 (a) shows a Venn diagram for a confirming item of evidence. In this case, E is a confirming item of evidence for the proposition a. If we observe E, then we are sure that a is true. Suppose a is the assertion that, 'there is cloud in the sky'. The evidence E is that, 'it is raining'. We know that it can rain only if there are clouds. Thus, evidence E is a subset of the event a. Further, if we observe E then we are sure that a is true ($P(a|E) = 1$), regardless of the prior in a. An item of evidence is confirming for an assertion, if and only if, the likelihood of the evidence given the negation of the assertion is zero, i.e., $P(E|{\sim}a) = 0$. In the above example, the conditional probability of rain given no clouds is zero. Thus, rain is a confirming evidence for clouds. Figure 2 (b) shows an illustration of disconfirming evidence.

Figure 2 here

The above definition of confirming evidence yields a likelihood ratio of infinity, i.e., $\lambda = P(E|a)/P(E|{\sim}a) = 8$ for any value of the likelihood $P(E|a)$, since $P(E|{\sim}a) = 0$. This in turn yields $P(a|E) = 1$ for any value of the prior $P(a)$ and likelihood $P(E|a)$. Thus, in terms of likelihood ratio, an item of evidence is confirming if λ is infinity. This is consistent with the positive evidence. A positive item of evidence becomes a confirming item of evidence when its strength becomes infinitely large.

In auditing, very rarely will any single item of evidence be confirming. In general, a set of evidence is jointly confirming for a proposition. Thus the notion of confirming evidence should be broadened from a single item of evidence to a set of evidence. The set may consist of items of evidence which, in isolation, are merely suggestive, but when taken in conjunction with other evidence, is confirming. Such a scenario is presented in Figure 3.

Figure 3 here

Consider the following scenario: an auditor discovers that the controller had signed two checks of $5,000 each to Beaux Painters, when the bills were only for $500 each. This could be an error of oversight or it could be intentional, in which case it is fraud. The evidence by itself renders only little support to fraud, because it is possible that the error was unintentional. Suppose the auditor finds out that Beaux Painters is owned by the wife of the controller. The wife's owning a painting shop, in isolation, is irrelevant to fraud in the company. The knowledge that the painting shop is owned by the controller's wife and the company does business with the painting shop may provide little support for fraud. However, the two items of evidence together, two significantly overvalued checks to the painting shop which is owned by the controller's wife, makes it conclusive that the controller is involved in fraud. This scenario is diagrammatically presented in Figure 3, where a is the assertion that there is management fraud, E_1 is the evidence that the wife of the controller owns a painting shop and the company does business with the painting shop, and E_2 is the evidence that the controller had written two overvalued checks. The intersection of E_1 and E_2 is a subset of a, and thus confirming for a. In practice, auditors do consider an item of evidence in conjunction with the other evidence when making decisions.

So far we have discussed the definition of a confirming item of evidence, E (or a set of items of evidence, E) for a. This definition requires that the conditional probability for the evidence (or the set of items of evidence) given the negation of the assertion be zero, i.e., $P(E/{\sim}a) = 0$ or the likelihood ratio be infinite, i.e., $P(E/a)/P(E/{\sim}a) = 8$. This definition yields a posterior probability of one when the evidence or the set of items of evidence are obtained, i.e., $P(a/E) = 1$. But in practice, is it possible to obtain such an item of evidence (or a set of items of evidence) that would yield $P(a/E) = 1$? Maybe, but at an enormous cost. The auditor may not deem it necessary to be absolutely sure that the assertion is met. He may be satisfied with only 0.95 posterior that the assertion is met given the evidence. In fact, one does consider 95% confidence interval quite acceptable in statistical tests of hypotheses. Thus, we can assume similar criterion that irrespective of prior probability of the assertion when the posterior probability is 0.95 or more then we should accept the assertion to be true. This statement means that all the evidence together provide strong enough posterior that the assertion is acceptable. However, we should not confuse this with the definition of confirming evidence; still the above evidence (or the set of evidence) may not be confirming (P(E/~a) ? 0).

IV. EVIDENCE AGGREGATION AND AUDIT PLANING

In this section, we show how various items of evidence are aggregated in an audit for planning and evaluation. We will limit our discussion to a simple example given in Figure 4. However, one can derive a general rule for combining evidence in complex situations. Such a case will be a topic for future research.

Figure 4 here

Evidence Aggregation for Evaluation

Assume that the auditor is interested in determining whether the accounts receivable balance (AR) is fairly stated. For simplicity of exposition, we assume that AR has only two audit objectives, existence (E) and valuation (V). Let us call AR, E, and V as variables and assume that these variables are binary variables. In our case this means that the accounts receivable balance is either fairly stated (ar) or not fairly stated (~ar); the existence objective is either met (e) or not met (~e); and the valuation objective is either met (v) or not met (~v). For the purpose of aggregating evidence, we will consider the relationship between the account balance, AR, and the two audit objectives, E & V to be an 'and' relationship. This relationship implies that 'ar' is true if and only if 'e' and 'v' are true.

Assume that the auditor accumulates four items of evidence in this case as described in Table 1 (Arens and Loebbecke, 1991). For simplicity of exposition, we consider a set of procedures as one item of evidence. Such an assumption is not necessary in general. One can consider each procedure as one item of evidence but the combination rule will be changed accordingly.

Table 1 here

The three variables AR, E, and V in Figure 4 are connected through an 'and' relationship as described earlier. For simplicity, we consider only four items of evidence. Three of them bear individually on the three variables and the fourth one bears on E and V both.

Let us consider that the auditor has performed all the procedures and assessed the following values for the strength of evidence: $\lambda_{AR} = 2.5$, $\lambda_E = 4.0$, $\lambda_V = 4.0$, and $\lambda_{EV} = 5.0$. As discussed in Appendix, the posterior odds (that the account balance is fairly stated and all the objectives have been met given all the evidence) is given by (A-12):

$$\text{Posterior Odds} = \frac{\lambda_{AR}\lambda_E\lambda_V\lambda_{EV}\pi_E\pi_V}{1+\lambda_E\pi_E+\lambda_V\pi_V}. \tag{3}$$

Assuming that all prior odds are one, i.e., we have no prior knowledge about whether the account balance (AR) or the objectives (E and V) are in error then we obtain:

$$\text{Posterior Odds} = \frac{\lambda_E\lambda_V\lambda_{EV}\lambda_{AR}}{1+\lambda_E+\lambda_V} \tag{4}$$

or

$$\text{Posterior Odds} = \frac{4.0\text{x}4.0\text{x}5.0\text{x}2.5}{1+4.0+4.0} = 22.22$$

This yields the following posterior probability that the account balance is fairly stated:

$$P(ar|E_E\&E_V\&E_{E\&V}\&E_{AR}) = \frac{\text{Posterior Odds}}{1+\text{Posterior Odds}}$$

$$= \frac{22.22}{1+22.22} = 0.9569.$$

Audit Planning

Aggregation of evidence for planning of an audit is similar to the above case with one difference. In the case of planning, the auditor has to assess, a-priori, the likelihood ratios, i.e., the strength of evidence that will be obtained after conducting the procedures. The strength of evidence, in general, depends on the extent, nature and timing of the procedures. Thus, the auditor has to make judgments about the extent, nature and timing of a procedure based on the desired level of the strength.

Continuing with the previous example, suppose the auditor is planning the audit at 0.95 level of posterior probability, i.e., a posterior odds of 19, and has obtained all the evidence except E_{EV} (See Figure 4 and Table 1). Also, assume that the auditor has obtained the following values for the strengths: $\lambda_{AR} = 2.5$, $\lambda_E = 4.0$, and $\lambda_V = 4.0$. The desired level of the strength to be obtained from E_{EV} would be determined by solving (4):

$$\text{Posterior Odds} = \frac{4.0 \text{x} 4.0 \text{x} 2.5 \text{x } \lambda_{EV}}{1+4.0+4.0} = 19, \tag{5}$$

which yields

$$\lambda_{EV} = 4.275. \tag{6}$$

V. SUMMARY AND CONCLUSION

In this paper we have defined the strength of evidence and discused the notions of positive, negative, and confirming evidence. The earlier work in the area, the Toba-Kissinger framework, was critically examined and was found to be inadequate in dealing with auditing situations. The Toba-Kissinger framework is based on comparing the prior with the posterior which leads to many inconsistencies.

We compared various measures of the strength of evidence and demonstrated that the likelihood ratio provides an appropriate measure of the strength of evidence. Also, we used the likelihood ratio to define positive, negative and confirming evidence. Further, the likelihoods provide an easier representation of uncertainty for audit planning and evaluation as shown in Section IV.

We feel that the concepts developed in this article should help the auditor aggregate various items of evidence in the Bayesian framework. Also, this research should aid the researchers building expert and/or decision support systems for planning and evaluation of an audit. We should point out that we did not address how the likelihood ratios can be obtained from qualitative and statistical evidence. We feel that this is an important issue and future research should be directed toward this issue.

FOOTNOTE

1. *Measure Based on Posterior and Prior* - The measure of the strength of evidence based on the difference between the posterior and the prior (Friedman 1986) is similar to Toba-Kissinger framework and suffers from the same limitations as the Toba-Kissinger framework (see Section I for details).

   Measure Based on Difference of Likelihoods - This approach uses the difference between the likelihoods as the strength of evidence, i.e., $P(E|a) - P(E|\sim a)$ determines the strength where $P(E|a)$ and $P(E|\sim a)$ represent the likelihoods with E being the evidence and a being the assertion (Cullison 1969; Jeffrey 1983). The evidence is supportive of a if the difference is positive, and vice versa. According to this definition, the higher the difference, the greater the strength of evidence. However, this definition is also fraught with problems. One would expect a higher increase in support from an item of evidence of higher strength. That is, if evidence E_1 has higher strength than evidence E_2, then one would expect the posterior probability after obtaining E_1 to be greater than the posterior probability after obtaining E_2. Unfortunately, that is not the case as shown here. Suppose, $P(E_1|a) = 0.6$ and $P(E_1|\sim a) = 0.4$, and $P(E_2|a) = 0.2$ and $P(E_2|\sim a) = 0.1$. Thus the strength of evidence E_1 is 0.2, and the strength of evidence E_2 is 0.1. Evidence E_1 is stronger than evidence E_2, and we will expect a higher increase in probability after obtaining E_1. Let the prior on a be 0.4 then $P(a|E_1) = 0.5$, and $P(a|E_2) = 0.57$ (The choice of 0.4 is arbitrary, the result holds for all priors.). The increase in belief is 0.1 for E_1 and 0.17 for E_2. The 'weaker' evidence has more impact than the stronger evidence.

2. Assume $P(a) = 0.2$, and $P(a|E) = 0.4$ then $P(a|E) > P(a)$ which means E is positive and supports a. But, we also find that $P(\sim a|E) = 0.6 > P(a|E) = 0.4$ which implies that E is confirming $\sim a$. The statements are contradictory to each other.

3. Since $P(a|E) = 0.3 > P(a) = 0.2$, the evidence is supporting the assertion. However, since $P(a|E) = 0.3 < P(\sim a|E) = 0.7$, the evidence is not confirming the assertion.

4. The item of evidence which would increase the probability in the assertion from 0.2 to 0.3, can be characterized as $P(E|a) = 0.4$, and $P(E|\sim a) = 0.27$. The same evidence when combined with a prior of 0.7 would result in a posterior of 0.76 by applying Bayes' rule.

5. Both items of evidence can be represented by likelihoods of $P(E|a) = 0.533$ and $P(E|\sim a) = 0.2$.

6. Time precedence and causation are two of the factors that distinguish prospective evidence from retrospective evidence.

7. The Likelihood ratios for E_1, E_2, and $E_1\&E_2$, are given by $P(E_1|a) / P(E_1|\sim a)$, $P(E_2|a)/P(E_2|\sim a)$ and $P(E_1\&E_2|a) / P(E_1\&E_2|\sim a)$, respectively. Since $P(E_1\&E_2|a) = P(E_1|a)P(E_2|a)$ and

$P(E_1 \& E_2|\sim a) = P(E_1|\sim a)P(E_2|\sim a)$, as the two items of evidence are assumed to be independent, the likelihood ratio for $E_1 \& E_2$ becomes the product of the likelihood ratio for E_1 and the likelihood ratio for E_2 (Edwards 1984).

8. For a positive item of evidence, $P(E|a)/P(E|\sim a) > 1$. From Bayes' rule we know,

$$P(a|E) / P(\sim a|E) = P(E|a)P(a) / P(E|\sim a)P(\sim a)$$

Thus, for a positive evidence we can write:

$$P(E|a) / P(E|\sim a) = P(a|E)P(\sim a) / P(\sim a|E)P(a) > 1,$$

which implies that

$$P(a|E)P(\sim a) > P(\sim a|E)P(a),$$

or

$$P(a|E) [1 - P(a)] > [1 - P(a|E)] P(a),$$

or

$$P(a|E) > P(a).$$

9. "Almost everyone interested in confirmation theory believes that confirmation relations ought to be analyzed in terms of probability relations" (Glymour 1980, p. 64).

REFERENCES

Arens, A.A. & Loebbecke, J.K. 1991. *Auditing: An Integrated Approach*, Prentice Hall, New Jersey.

Ashton, A. H. and Ashton,R.H. 1988. Sequential Belief Revision in Auditing. *The Accounting Review,* (October): 623-41.

Barnard, G.A. 1947. Review of “Sequential Analysis. *Journal of the American Statistical Association*, XLII: 659.

Barnett, V. 1982. *Comparative Statistical Inference.* J.Wiley Press, New York, 2nd Edition.

Birnbaum, A. 1962. On the Foundations of Statistical Inference. *Journal of the American Statistical Association,* LVII: 269-306.

Cullison, A. D. 1969. Probability Analysis of Judicial Fact Finding: A Preliminary Outline of the Subjective Approach. *University of Tol. Law Review,* Vol. 1:(1).

Edwards, A. W. F., *Likelihood: An account of the statistical concept of likelihood and its application to scientific inferences*, Cambridge University Press, Cambridge (1984).

Edwards, W., D. A. Schum, and R. L. Winkler, "Murder and (of?) the Likelihood Principle: A Trialogue," *Journal of Behavioral Decision Making*, Vol. 3 (November 1982), pp. 75-89.

Fisher, R. A. 1922. On The Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society,* Vol. CCXXII: 310.

Friedman, R. D. 1986. A Close Look at Probative Value. *Boston University Law Review*, Vol. 66:(3&4), (May): 733- 759.

Good, I.J. 1950. *Probability and Weighing of Evidence*, London, Charles Griffin; New York, Hafners.

Good, I.J. 1983. *Good Thinking: The Foundation of Probability and its Applications*, University of Minnesota Press, Minneapolis.

Glymour, C. 1980. *Theory and Evidence*, Princeton University Press, Princeton, N.J.

Hacking, I. 1965. *Logic of Statistical Inference,* Cambridge University Press, London.

Jeffrey, R.C. 1983. *The Logic of Decisions* (2nd ed.) McGraw Hill, New York.

Kaye, I. 1986. Comment: Quantifying Probative Value. *Boston University Law Review*, Vol. 66:(3&4), (May): 761- 766.

Kissinger, J.N. 1977. A General Theory of Evidence as the Conceptual Foundation in Auditing Theory: Some Comments and Extensions. *The Accounting Review* (April): 322-39.

Kuipers, B. and J.Kassier. 1984. Causal Reasoning in Medicine: Analysis of a Protocol. *Cognitive Science,* 8: 363-385.

Lampert, M. 1977. Modelling Relevance. *Michigan Law Review*, (April) Vol. 75: (5&6): 1021-1057.

Minsky, M., and Selfridge, O.G. 1961. Learning in Random Nets. *Information Theory,* Colin Cherry, ed., London: Butterworths: 335-347.

Savage, L.J. 1961. The Foundations of Statistics Reconsidered. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* I: 583.

Schum, David A. and Anne W. Martin. 1982. Formal and Empirical Research on Cascaded Inference in Jurisprudence. *Law and Society Law Review,* Vol. 17.

Shachter, R.D. and D. E. Heckerman. 1987. Thinking Backward for Knowledge Acquisition. *AI Magazine,* Vol.8:(3), (Fall): 55-61.

Toba, Y. 1975. A General Theory of Evidence as the Conceptual Foundation in Auditing Theory. *The Accounting Review* (January): 7-24.

Tribe, Lawrence H. 1971. Trial by Mathematics: Precision and Ritual in the Legal Process. *Harvard Law Review,* Vol. 84: 1329- 1393.

Tubbs, Richard M., William F. Messier and Robert Knechel. 1990. Recency Effects in the Auditor's Belief Revision Process. *The Accounting Review* (April): 452-460.

Appendix

Aggregation of Evidence

We have already shown that if there were more than one item of evidence bearing on a single objective then the combined evidence is simply the product of the strengths of the individual items of evidence. That is, if the strength of evidence is represented by the Likelihood Ratio, λ, then the combined strength, Λ, of all the items of evidence bearing on a single variable is: $\Lambda = \sum_{i} l^{i}$ where λ^{i} represents the strength of ith item of evidence on the variable. However, when we want to combine evidence bearing on a cluster of variables then the combination rule is not so straight forward. We will use the example in Figure 4 to illustrate the process of combining evidence in such a case.

Assume that the auditor is auditing accounts receivable balance (AR). We further assumed that the account has only two audit objectives, Existence (E) and Valuation (V). The accounts receivable balance is fairly stated (*ar*) only when the two objectives have been met, i.e., '*ar*' is true if and only if '*e*' and '*v*' are true. As shown in Figure 4 and in Table 1, we have four items of evidence: E_{E}, E_{V}, E_{EV}, and E_{AR}. The auditor would like to aggregate all these items of evidence to determine whether the account is fairly stated and the objectives are met. In other words, the auditor needs to determine the posterior probability $\mathrm{P}(ar\&e\&v|E_{\mathrm{AR}}\&E_{\mathrm{E}}\&E_{\mathrm{V}}\&E_{\mathrm{EV}})$.

Our objective is to express the above posterior in terms of various likelihood ratios, i.e., the the strength of evidence. In order to achieve this objective, we complete the following steps. Let us assume

$$A = ar\&e\&v \quad \text{(A-1)}$$

and

$$\sim A = (\sim ar\&\sim e\&v) \cup (\sim ar\&e\&\sim v) \cup (\sim ar\&\sim e\&\sim v) \quad \text{(A-2)}$$

Applying Bayes' rule we obtain:

$$\mathrm{P}(A|E_{\mathrm{AR}}\&E_{\mathrm{E}}\&E_{\mathrm{V}}\&E_{\mathrm{EV}}) = \mathrm{P}(ar\&e\&v|E_{\mathrm{AR}}\&E_{\mathrm{E}}\&E_{\mathrm{V}}\&E_{\mathrm{EV}})$$
$$= \frac{\mathrm{P}(E_{\mathrm{AR}}\&E_{\mathrm{E}}\&E_{\mathrm{V}}\&E_{\mathrm{EV}}|ar\&e\&v)\mathrm{P}(ar\&e\&v)}{\mathrm{P}(E_{\mathrm{AR}}\&E_{\mathrm{E}}\&E_{\mathrm{V}}\&E_{\mathrm{EV}})} \quad \text{(A-3)}$$

We can assume that each item of evidence is conditionally independent of every thing else given the variable(s) it bears on [Edwards 1984]. These conditions yield the following expression for (A-3):

$$P(A|E_{AR}\&E_E\&E_V\&E_{EV}) = \frac{P(E_{AR}|ar)P(E_E|e)P(E_V|v)P(E_{EV}|e\&v)P(ar\&e\&v)}{P(E_{AR}\&E_E\&E_V\&E_{EV})}. \quad \text{(A-4)}$$

Also, we know that

$$\begin{aligned} P(\sim A|E_{AR}\&E_E\&E_V\&E_{EV}) = {} & P(\sim ar\&\sim e\&v|E_{AR}\&E_E\&E_V\&E_{EV}) \\ & + P(\sim ar\&e\&\sim v|E_{AR}\&E_E\&E_V\&E_{EV}) \\ & + P(\sim ar\&\sim e\&\sim v|E_{AR}\&E_E\&E_V\&E_{EV}) \end{aligned} \quad \text{(A-5)}$$

With a similar argument that we have used in obtaining (A-4) from (A-3), we can write each term on the right hand side of (A-5) as:

$$P(\sim ar\&\sim e\&v|E_{AR}\&E_E\&E_V\&E_{EV}) = \frac{P(E_{AR}|\sim ar)P(E_E|\sim e)P(E_V|v)P(E_{EV}|\sim e\&v)P(\sim ar\&\sim e\&v)}{P(E_{AR}\&E_E\&E_V\&E_{EV})}, \quad \text{(A-6)}$$

$$P(\sim ar\&e\&\sim v|E_{AR}\&E_E\&E_V\&E_{EV}) = \frac{P(E_{AR}|\sim ar)P(E_E|e)P(E_V|\sim v)P(E_{EV}|e\&\sim v)P(\sim ar\&e\&\sim v)}{P(E_{AR}\&E_E\&E_V\&E_{EV})}, \quad \text{(A-7)}$$

$$P(\sim ar\&\sim e\&\sim v|E_{AR}\&E_E\&E_V\&E_{EV}) = \frac{P(E_{AR}|\sim ar)P(E_E|\sim e)P(E_V|\sim v)P(E_{EV}|\sim e\&\sim v)P(\sim ar\&\sim e\&\sim v)}{P(E_{AR}\&E_E\&E_V\&E_{EV})}. \quad \text{(A-8)}$$

The posterior odds can be written as (A-4) divided by (A-5):

$$\text{Posterior Odds} = \frac{P(A|E_{AR}\&E_E\&E_V\&E_{EV})}{P(\sim A|E_{AR}\&E_E\&E_V\&E_{EV})}. \quad \text{(A-9)}$$

We assume here that two objectives 'E' and 'V' are independent, that is, $P(e\&v) = P(e)P(v)$.

Also, we know that 'ar' is true if and only if 'e' and 'v' are true implying that $P(ar|e\&v) = P(\sim ar|\sim e\&v) = P(\sim ar|e\&\sim v) = P(\sim ar|\sim e\&\sim v) = 1$. These conditions imply that:

$$\begin{aligned} P(ar\&e\&v) &= P(ar|e\&v)P(e)P(v) = P(e)P(v), \\ P(\sim ar\&\sim e\&v) &= P(\sim ar|\sim e\&v)P(\sim e)P(v) = P(\sim e)P(v), \\ P(\sim ar\&e\&\sim v) &= P(\sim ar|e\&\sim v)P(e)P(\sim v) = P(e)P(\sim v), \\ P(\sim ar\&\sim e\&\sim v) &= P(\sim ar|\sim e\&\sim v)P(\sim e)P(\sim v) = P(\sim e)P(\sim v). \end{aligned} \quad \text{(A-10)}$$

Using (A-9) and substituting (A-10) in (A-4) and (A-6) - (A-8), dividing the numerator and denominator of the resulting expression by $P(\sim e)$, $P(\sim v)$, $P(E_{AR}|\sim ar)$, $P(E_E|\sim e)$, $P(E_V|\sim v)$, and $P(E_{EV}|\sim e\&\sim v)$, and remembering that

$$P(E_{EV}|\sim e\&\sim v) = P(E_{EV}|\sim e\&v) = P(E_{EV}|e\&\sim v), \quad \text{(A-11)}$$

we obtain the desired expression for the posterior odds:

$$\text{Posterior Odds} = \frac{\lambda_{AR}\lambda_E\lambda_V\lambda_{EV}\pi_E\pi_V}{1+\lambda_E\pi_E+\lambda_V\pi_V}. \tag{A-12}$$

The symbol π stands for the prior odds and λ for the strength of evidence or the likelihood ratios:

$$\pi_{AR} = \frac{P(ar)}{P(\sim ar)},\ \pi_E = \frac{P(e)}{P(\sim e)},\ \text{and } \pi_V = \frac{P(v)}{P(\sim v)},$$

and

$$\lambda_{AR} = \frac{P(E_{AR}|ar)}{P(E_{AR}|\sim ar)},\ \lambda_E = \frac{P(E_E|e)}{P(E_E|\sim e)},\ \lambda_V = \frac{P(E_V|v)}{P(E_V|\sim v)},\ \lambda_{EV} = \frac{P(E_{EV}|e\&v)}{P(E_{EV}|\sim e\&\sim v)}.$$

The posterior probability can be obtained by:

$$P(A|E_{AR}\&E_E\&E_V\&E_{EV}) = \frac{\text{Posterior Odds}}{1 + \text{Posterior Odds}}. \tag{A-13}$$

where (A-12) defines the posterior odds.

Equation (A-13) is the desired result. It provides the posterior probability in terms of the strength of evidence and prior odds that the account is fairly stated and all the audit objectives have been met. One can derive a general expression for more complex situation but we postpone this problem for future research.

It should be noted that Equations (A-12) and (A-13) yield intuitive results. For example, let us assume that: (1) we have no prior knowledge that the objectives have been met, i.e., the prior odds are one ($\pi_E = \pi_V = 1$), (2) we have no evidence bearing directly on the account and on the cluster of objectives {E,V}, i.e., $\lambda_{AR} = \lambda_{EV} = 1$, and (3) we have direct evidence bearing individually on the two objectives E and V of strengths $\lambda_E = 4$, and $\lambda_V = 3$, respectively. The evidence bearing on objective E implies that the posterior probability that the objective is met is 0.8 (by using a relationship similar to (A-13)). Similarly, the evidence bearing on V implies that the posterior probability that the objective is met is 0.75. The posterior probability that both the objectives are met is simply the product of the two posteriors which is 0.6 (0.8x0.75). Since we have assumed that the account is fairly stated if and only if the two objectives have been met, the posterior probability that the account is fairly stated and the two objectives have been met is simply 0.6. This is what one would obtain if Equations (A-12) and (A-13) were used directly. For example, for the above values of π's and λ's, the posterior odds from Equation (A-12) is 1.5 which yields a posterior probability of 0.6 from Equation (A-13).

Table 1

List of Items of Evidence (i.e., Audit Procedures) Gathered by the Auditor for Accounts Receivable Audit (Arens and Loebbecke, 1991)

Evidence bearing on AR (E_{AR}) - Analytical Procedures: (i) Review accounts receivable trial balance for large and unusual receivables. (ii) Calculate ratios indicated in carry-forward working papers and follow up any significant changes from prior years. The test results are favorable.

Evidence Bearing on E (E_E) - Test of Transactions for sales being valid and cash receipts being complete: (i) Trace recorded sales from the sales journal to the file of supporting documents, which includes a duplicate sales invoice, bill of lading, sales order, and customer order. (ii) Obtain the prelisting of cash receipts, compare prelisting with the duplicate deposit slip and also trace amounts to the cash receipts journal, testing for names, amounts, and dates. The test results are favorable.

Evidence bearing on V (E_V) - Test of Transactions for sales and cash receipts being properly valued: (i) Trace selected duplicate sales invoices numbers from the sales journal to: a) Duplicate sales invoices, and check for the total amount recorded in journal, date, customer name and account classification. Check the pricing, extensions and footings. b) Bill of lading, duplicate sales order, and customer order and test for customer name, product description, quantity, and date. (ii) Perform a proof of cash receipts. The test results are favorable.

Evidence bearing on both E & V (E_{EV}) - Test of Details of Balance: Confirm accounts receivable using positive confirmations above a given amount and perform alternative procedures for all confirmations not returned on the first and second request. The test results are favorable.

Figure 1

The posterior probability on the assertion is plotted against the prior probability for different likelihood ratios, λ. λ = 0.2 and 0.4 represent strong and moderate negative evidence, respectively. λ = 1 represents neutral evidence; and λ = 2.5 and 6 represent moderate and strong positive evidence, respectively.

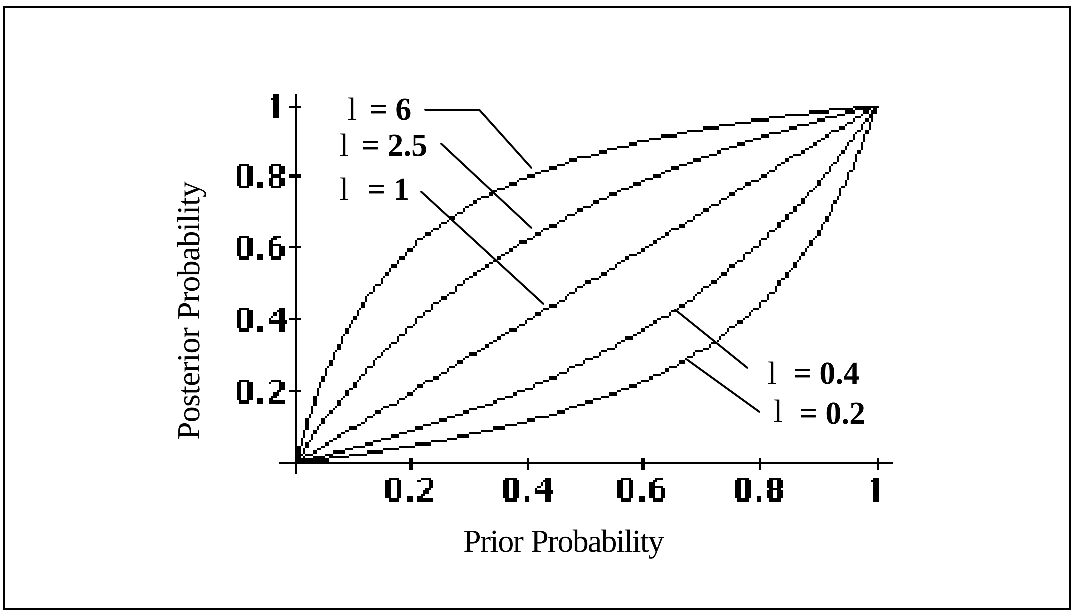

Figure 2

Diagrammatic Representation of Confirming and Dis-confirming Evidence. (a) *E* is Confirming Evidence for *a*, (b) *E* is Dis-confirming Evidence for *a*.

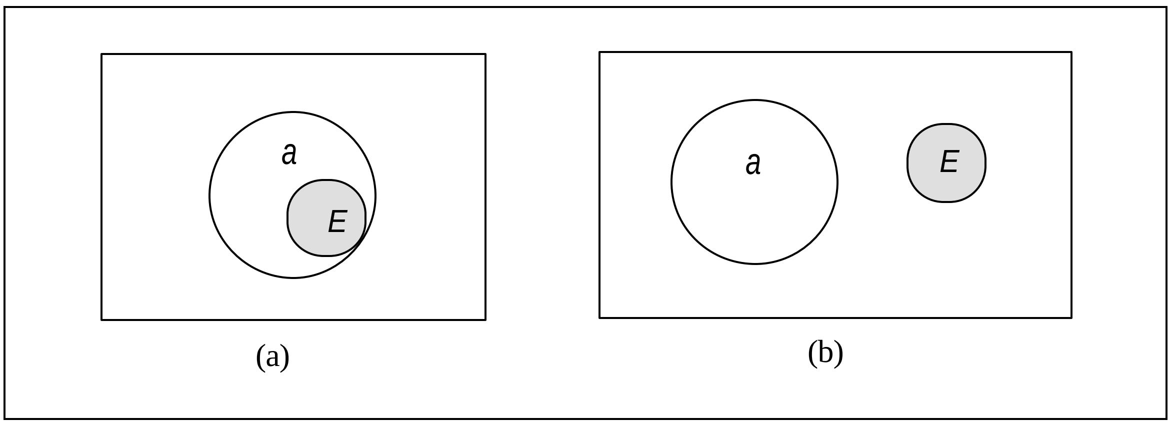

Figure 3

Diagrammatic representation of a set of evidence that is confirming. a is the assertion. E_1 and E_2 are two independent items of evidence. Individually they are not confirming for a. But taken together they are confirming for a.

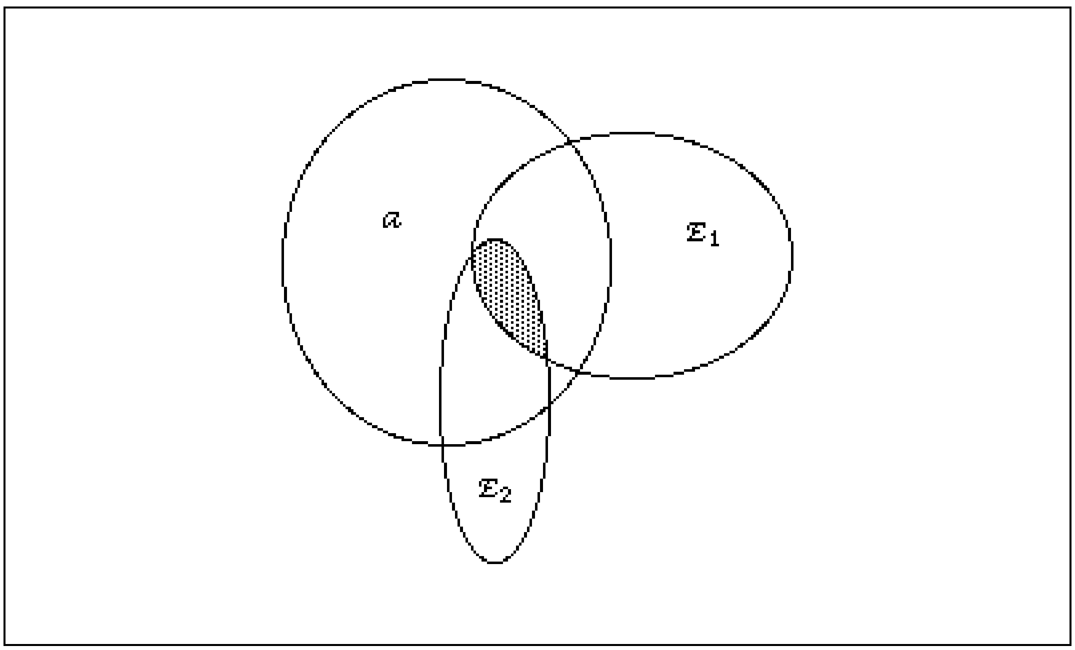

Figure 4

Evidential Network for Accounts Receivable (The rectangular boxes represent items of evidence bearing on various variables. These items of evidence are described in Table 1).