STATISTICAL MODELLING AND INFERENCE FOR FINANCIAL AUDITING

by

Yaw Bimpeh
BSc (Hon's) in Mathematics, University of Science and Technology, Ghana
MSc in Mathematical Sciences, Norwegian University of Science and Technology, Norway

A thesis presented for the degree of
Doctor of Philosophy
Modelling & Scientific Computing Group
School of Computing
Dublin City University
Ireland

Dublin, Ireland (2006)

Contents

List of Figures

List of Tables

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:______________________________

Candidate

ID No.:____________________________

Date:________________________________

Abstract

The fundamental problem addressed in this thesis is the problem of constructing confidence limits for mean or totals in finite populations, when the underlying distribution is highly skewed and contains a substantial proportion of zero values. This situation is often encountered in statistical applications such as statistical auditing, reliability, insurance, meteorology and biostatistics. The motivating example underlying this research is that of auditing (see the report published by the National Academy Press entitled "Statistical Models and Analysis in Auditing", Panel on Non-standard Mixtures of Distributions 1989), where interest is focused on computing the confidence bounds for the true total error amount. In such populations the use of the classical survey-sampling estimators such as the mean-per-unit, the difference, the ratio or regression, based on the normality assumption of the sampling distribution of the estimates, has been found unreliable, (e.g. Stringer 1963, Kaplan 1973, and Neter and Loebbecke 1975, 1977). Several alternative methods have been proposed, of which the Stringer bound (Stringer 1963), is the most widely used. This bound, while overcoming the unreliability problem of the classical estimators, has been found to be extremely conservative.

In this research, we develop new methods for constructing confidence intervals for the mean of a bounded random variable. Further, we apply these new methods to data that are heavily skewed and marked by many zero values. Our proposed confidence intervals have a good coverage probability and precision.

The first method is based on a novel use of the Edgeworth expansion for the studentised compound Poisson sum. In this work, we have reduced the problem of estimating the total error amount in auditing to the compound Poisson sum, and explored the asymptotic expansion for a compound Poisson distribution as a method of constructing confidence bounds on the total error amount. This method is less restrictive than the Stringer bound, and imposes no prior structure on the error distribution.

We obtain a bound on the cumulative distribution function of the prorated errors, which we then use to give an alternative form of the Stringer bound. With this form of the Stringer bound, we were able to use Bolshev's recursion to obtain a lower bound on its coverage probability, and showed that, for a sample size, $n \leq 2$, this lower bound is greater than or equal to the stated coverage probability. We illustrate numerically that the Stringer Bound is reliable when (n, α) falls into a number of ranges; specifically $n \leq 11$ and a significance level $\alpha \in (0, 0.05)$; $n \leq 10$ and $\alpha \in (0, 0.1)$; $n \leq 9$ and $\alpha \in (0, 0.20)$; $n \leq 8$ and $\alpha \in (0, 0.40)$; and $n \leq 7$ and $\alpha \in (0, .5)$. We also proposed an extension to the Stringer method based on Rom's adjusted significance levels, and illustrated

numerically the reliability of the extended Stringer bound for values of α in the range .05 to .5, and for $n = 1$ to $n = 20$.

For the new bounds, we provide explicit expressions which make their computations straightforward. Monte Carlo simulations are carried out to evaluate the performance of the methods developed in this thesis when applied to accounting data, we investigate the performance of each method and assess whether or not it is affected by varying the distribution of accounting data, the effects of 100-percent overstatement error and the effects of error rates, using real and simulated populations. The method based on compound Poisson sum seems to reliable for large samples. However, for small samples the compound Poisson bound has the poorest results (in the sense of coverage probability), in particular, for populations containing a lower concentration of small error amounts. Although the extended Stringer bound, has a good coverage probability for all sample sizes and significance levels, it shares the extreme conservativeness of the Stringer bound.

Acknowledgments

I would like to thank my supervisor Prof. Jane Horgan for her guidance. I am glad to have worked with her and benefited tremendously from her knowledge and experience in technical matters. I would like to add special thanks to Prof. Heather Ruskin. I appreciate the time she devoted in reading my work and her useful suggestions. I am also grateful to Dr. Emmanuel Aziz for his many helpful suggestions, to Dr. Benjiman Aziz and Dr. John Burns for being good friends. I want to express my gratitude to all people at the School of Computing, especially those in the Modelling and Scientific Computing group. Finally, I would like express my deep gratitude to my family for their love and support.

Dedication

This is dedicated to my wife Comfort Addae and beloved daughter Ama Agyemang Bimpeh.

List of Abbreviations

AICPA = American Institute of Certified Public Accountants

AUS = Auditing Standards Board of the Australian Accounting Research Foundation

ISA = The International Standards on Auditing

NAO =The National Audit Office

PNMD =Panel on Non-standard Mixtures of Distributions

PPS = Probability proportional to size sampling

SAS = AICPA Statement on Auditing Standard

SB = Stringer bound

SCP = Studentised Compound Poisson

SRS = Simple random sampling

STRS = Stratified random sampling

Chapter 1

Introduction

1.1 Background

Most financial databases are subject to annual audits, which often entail random sampling whenever the data set is too large for a complete audit. Very often, auditors use statistical sampling results and detailed testing of individual transactions to obtain reasonable assurance to support audit opinion on financial statements (The AICPA's *Audit Sampling* Guide 1992). The sample statistics are used to estimate the amount of monetary error in an account balance (or transaction) with reasonable accuracy. Their ultimate interest is to compute an upper confidence bound for the true total error amount in the recorded monetary value. This is used by the auditor as a threshold to accept or reject the account balance. That is, the auditor compares the upper bound with the tolerable error (the maximum error in the population that the auditor would be willing to accept while still concluding that the result from the sample has achieved the audit objective). If the upper bound exceeds the tolerable error, the auditor regards the statistical evidence as indicating the possibility of material error. This calls for a much more rigorous audit. When the computed upper confidence bound does not exceed the tolerable error amount, the auditor decides that there is no material error (Robert 1978).

Earlier attempts in analysing audit data using statistical sampling techniques adopted the classical survey-sampling estimators such as the Horvitz-Thompson (1952) estimator, difference, ratio and regression estimators. These methods, however, rely on the use of the central limit theorem, and have been found to be unreliable when making inferences about the mean or total monetary error (e.g. Stringer 1963, Kaplan 1973, Neter and Loebbecke 1975, 1977, Baker and Copeland 1979, and Beck 1980). Thus, the coverage probability attained by these estimators frequently falls substantially below the nominal level for the sample sizes commonly used by auditors. This unreliability is attributed to the high skewness, low incidence of error peculiar to auditing populations (e.g. Stringer 1963, Neter and Loebbecke 1975, 1977, Frost and Tamura 1986). The skewness effect of the underlying distribution of auditing data may be so severe that traditional large-sample techniques, based on the central limit theorem, are unreliable even if the sample size is moderately large or the population is stratified (Neter and Loebbecke 1975, 1977).

1.2 Motivation

The unreliability problem of the classical survey-sampling estimators has motivated a number of auditing practitioners and academics to explore other approaches. A variety of methods have been developed to overcome the limitations of the classical estimators; these include the Stringer Bound (Stringer 1963), Multinomial Bound (Fienberg, Neter, and Letich 1977 and Leitch et al. 1982), Cell Bound (Leslie, Teitlebaum, and Anderson 1979), Cox and Snell bound (Cox and Snell 1979), Moment Bound (Dworin and Grimlund 1984, 1986), Bayesian Normal (Menzefricke and Smieliauskas 1984), Multinomial-Dirichlet (Tsui, Matsumura, and Tsui 1985), Numerical Inversion of a Characteristic Function motivated by the Ferguson's Dirichlet process (Tamura 1988), Variance Augmentation (Rohrbach 1993), Likelihood Ratio Method (Kvanli, Shen, and Deng 1998), Empirical Cornish-Fisher Expansion and Bootstrap Calibration (Helmers 2000), and Monte-Carlo Simulation (Laws and O'Hagan 2000).

The differences in the various methods include both selection techniques and evaluation of the sample result. One problem they have in common is that the confidence level attained by the bounds constructed by these methods could differ substantially from the nominal coverage, depending on the distribution of the error. For example, the Moment and Bayesian Normal bounds have unpredictable coverage failures (Grimlund and Felix 1987), whereas the Cell bound achieves coverage probability larger than the nominal (Leitch et. al. 1982).

The most commonly used method is the Stringer bound. This bound, while overcoming the unreliability problem of the classical estimators, has been found to be extremely conservative, in the sense that the confidence bounds are substantially greater than the actual error amount (e.g. Leitch et. al. 1982; Bickel 1992). As pointed out by Leitch et al. (1982), this conservatism has unfortunate consequences in terms of power against immaterial monetary error. There is an interest in auditing practice to find a less conservative and still reliable alternative method (see PNMD[1] 1989). "The National Audit Office (NAO), which is responsible to the UK parliament for the audit of all Government departments and a wide range of other public sector bodies, has established a Statistics Advisory Panel of professional statisticians which advises on a range of statistical issues in audit" (Barnett and Howarth 1998). As evidenced by a recent survey in the United States "some auditors improperly rely on formal statistical methods to evaluate non-statistical samples" (Hall, Hunton, and Pierce 2002). An extensive simulation study comparing 14

[1] PNMD = Panel on Non-standard Mixtures of Distributions

bounds used in monetary-unit sampling performed by Swinamer et al. (2004) using both real and simulated data, suggested that no one method is superior in sense of giving smaller and reliable upper confidence bound. Motivated by these considerations, this work develops improved methods of constructing confidence bounds for the true total error amount in auditing populations.

1.3 Contributions of this Research

We develop alternative methods for constructing confidence bounds for population mean, when the underlying distribution is highly skewed and marked by many zero values. The first alternative method is based on the Edgeworth expansion for studentised compound Poisson processes. The basic idea underlying this approach is to construct a model using compound Poisson process incorporating two aspects of auditing populations: the error rate and the distribution of non-zero prorated errors. Then, based on this model, we estimate the upper confidence bound on the population mean prorated error. This method imposes no prior structure on the error distribution.

Second, we obtain a bound on the cumulative distribution function of the prorated errors, which we then use to give an alternative form of the Stringer bound. With this form of the Stringer bound, we were able to use Bolshev's recursion to obtain a lower bound on its coverage probability, and showed that, for a sample size, $n \leq 2$, this lower bound is greater than or equal to the stated coverage probability. We illustrate numerically that the Stringer bound is reliable when (n, α) falls into a number of ranges; specifically $n \leq 11$ and a significance level $\alpha \in (0, 0.05)$; $n \leq 10$ and $\alpha \in (0, 0.1)$; $n \leq 9$ and $\alpha \in (0, 0.20)$; $n \leq 8$ and $\alpha \in (0, 0.40)$; and $n \leq 7$ and $\alpha \in (0, .5)$. We also proposed an extension to the Stringer method based on Rom's adjusted significance levels, and illustrated numerically the reliability of the extended Stringer bound for values of α in the range .05 to .5, and for $n = 1$ to $n = 20$.

For these methods, we provide theoretical and numerical results to show the efficiency and reliability of the proposed methods. Large-scale simulation studies are also carried out to assess their performance using real and simulated populations. Our results indicate that these methods are reliable for all cases investigated, with some coverage failures for small samples sizes from populations containing a lower concentration of small error amounts in the first method. Comparisons with the Stringer are also discussed.

1.4 Outline of Scope of the Thesis

The remainder of the thesis is structured as follows:

In Chapter 2, the general background to financial auditing is described.

In Chapter 3, we review the characteristics of errors in accounting population, the various methods that have been proposed for computing confidence bounds on the total error amount, and we discuss their limitations.

In Chapter 4, we develop a new method for constructing confidence bounds for the total error amount in auditing population, when the underlying distribution is highly skewed and contains substantial proportion of zero errors. The problem of estimating total error amount in auditing is reduced to a compound Poisson sum. We explore the asymptotic expansion for a compound Poisson distribution as a method of constructing confidence bounds on the total error amount.

In Chapter 5, we obtain a bound on the cumulative distribution function of the prorated errors, which we then use to give an alternative form of the Stringer bound. With this form of the Stringer bound, we were able to use Bolshev's recursion to obtain a lower bound on its coverage probability, and showed that, for a sample size, $n \leq 2$, this lower bound is greater than or equal to the stated coverage probability. We illustrate numerically that the Stringer bound is reliable when (n, α) falls into a number of ranges; specifically $n \leq 11$ and a significance level $\alpha \in (0, 0.05)$; $n \leq 10$ and $\alpha \in (0, 0.1)$; $n \leq 9$ and $\alpha \in (0, 0.20)$; $n \leq 8$ and $\alpha \in (0, 0.40)$; and $n \leq 7$ and $\alpha \in (0, .5)$. We also proposed an extension to the Stringer's method based on Rom's adjusted significance levels, and illustrated numerically the reliability of the extended Stringer bound for values of α in the range .05 to .5, and for $n = 1$ to $n = 20$.

Creation of study populations for testing the bounds are discussed in Chapter 6.

In Chapter 7, we evaluate two methods developed in this thesis for estimating upper confidence bounds for error amount in accounting data: the studentised compound Poisson method and the extended Stringer method. We perform a Monte Carlo simulation to compare these methods. We investigate the relative performance of each method and assess whether or not it is affected by varying the distribution of accounting data, the effects of 100-percent overstatement error and the effects of error rates.

An overview and suggestion for future research are given in Chapter 8.

Chapter 2

Overview of Financial Auditing Procedure

This chapter provides a brief overview of the financial audit procedures. We explain what is involved in financial auditing, and also establish the relevance of statistical analysis in financial audits.

We start with a definition of auditing in Section 2.1. Section 2.2 presents financial auditing procedures. Types of errors in accounting system is given in Section 2.3. The methods of obtaining audit evidence are discussed in Section 2.4. Section 2.5 deals with the audit sampling. Section 2.6 gives a brief overview of statistical sampling strategies in auditing.

2.1 What is Auditing?

The American Accounting Association (AAA) define auditing as:

> A systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events, to ascertain the degree of correspondence between those assertions and established criteria, and communicating the results to interested users (Report of the Committee on Basic Audit Concepts 1972 p18).

This definition is broad but contains the basic idea that an audit is an investigation process. Most of the time, the term audit is used with a descriptive word to indicate the purpose of audit or the subject matter of the audit or both: for example, financial audit, tax audit or operational audit. This research deals with financial auditing, which is concerned with the collection and evaluation of evidence to support or refute assertions in the financial statements on which the auditor is required to form an opinion (Mautz and Sharaf 1961).

The accounting system that underlies the financial statements is the primary focus of the auditor. An integral component of the accounting system is a carefully documented set of procedures that prescribe how the accounting system should operate. The accounting system contains internal controls that are designed to prevent unintentional and intentional errors in the operation of the system. The auditors investigation usually begins with a general review of the client's accounting system, including its internal controls. This is for the purpose of concluding whether the system meets the needs for which it was designed.

The AUS defines *Internal control structure (internal controls)* as "management's philosophy and operating style, and all the policies and procedures adopted by management to assist in achieving the entity's objectives" (AUS 402). The internal control structure extends beyond those matters that relate directly to the financial report and consists of three elements:

(i) **Control environment,** which means "the overall attitude, awareness and actions of management regarding internal controls and its importance to the entity"(AUS 402).

(ii) **Information system,** which means " the methods and records established to identify, assemble, analyze, calculate, classify, record and report the transactions and other events that affect an entity, and to maintain accountability for assets, liabilities, revenues and expenditures"(AUS 402).

(iii) **Control procedures,** which means "those policies and procedures in addition to the control environment that management has established to ensure, as far as possible, that specific entity objectives will be achieved"(AUS 402).

2.2 The Financial Auditing Procedures

The ISA recognizes three main types of auditing procedures, risk assessment procedures, compliance testing and substantive procedures. A flow chart giving a summary of the auditing procedures is shown in Figure 2.1.

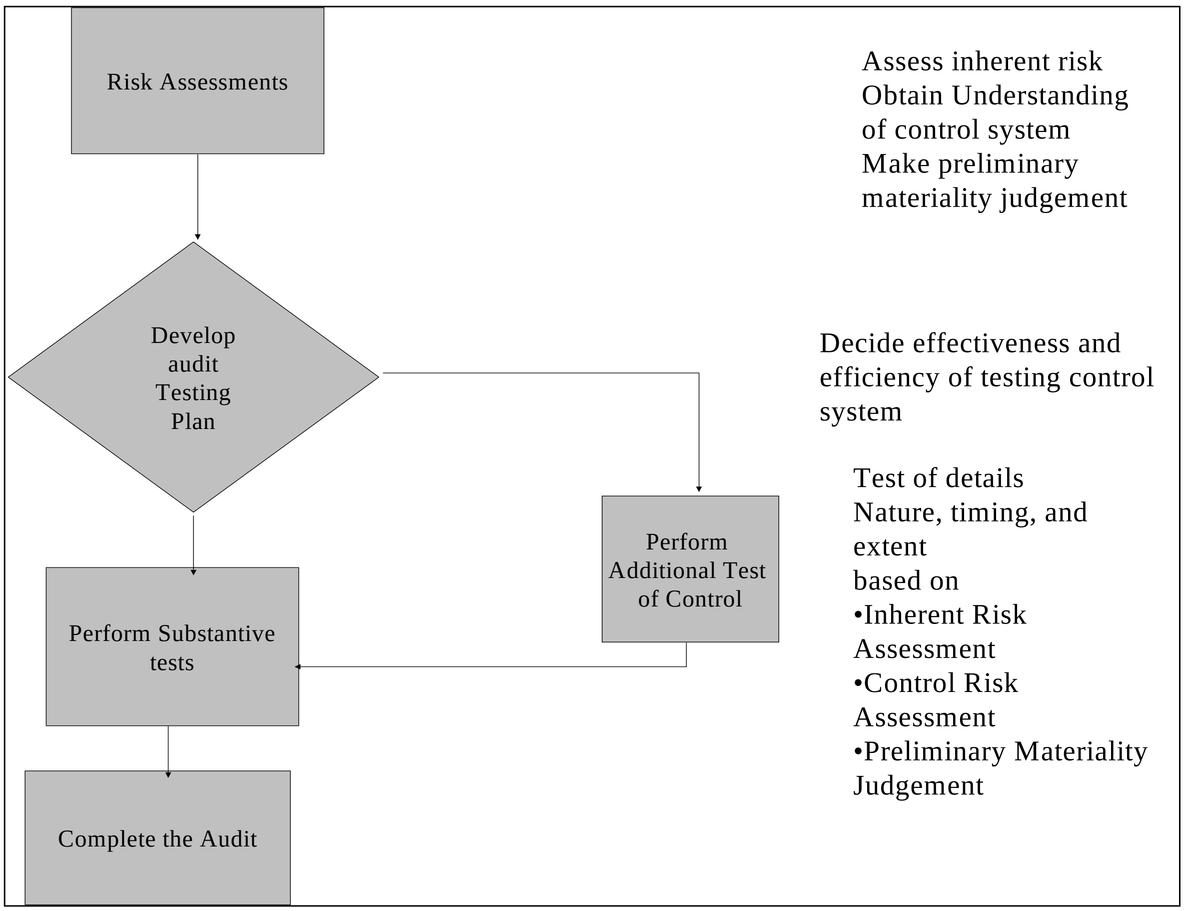

Figure 2.1: Summary of Auditing Procedures

2.2.1 Risk Assessment Procedures

Prior to the testing of the account balance and control procedures, the auditor generally performs risk assessment. By risk assessment we mean the assessment of the risk that the financial statement to be audited will contain an unacceptable error. The acceptable level of error is often referred to as *materiality* and may be expressed as a monetary value or as a percentage of the total account value. As AUS 402 says:

> The auditor should obtain an understanding of the internal control structure sufficient to plan the audit and develop an effective audit approach. The auditor should use professional judgement to assess audit risk and to design audit procedures to ensure it is reduced to an acceptable low level (AUS 402).

AUS 402 defines **Audit risk** as "the risk that the auditor gives an inappropriate audit opinion when the financial statement is materially misstated". Material misstatement in this context means an unacceptable misstatement. It is the auditor's responsibility to determine the margin of acceptable error or materiality. Audit risk has three components: inherent risk, control risk, and detection risk.

(i) Inherent risk means "the susceptibility of an account balance or class of transactions to misstatement that could be material, individually or when aggregated with misstatements in other balances or classes, assuming there were no related internal controls"(AUS 402).

(ii) Control risk means "the risk that misstatements that could occur in an account balance or class of transactions and that could be material, individually or when aggregated with misstatements in other balances or classes, will not be prevented or detected on a timely basis by the internal control structure"(AUS 402).

(iii) Detection risk means "the risk that an auditor's substantive procedures will not detect a misstatement that exists in an account balance or class of transactions that could be material, individually or when aggregated with misstatements in other balances or classes"(AUS 402).

According to AUS 402, the understanding of relevant aspects of the internal control structure, together with inherent and control risk assessments and other considerations, will enable the auditor to:

(a) identify the types of potential material misstatements that could occur in the financial statement;

(b) consider factors that affect the risk of material misstatements; and

(c) design appropriate audit procedures

The audit procedures performed for this purpose are referred to in auditing standards as “risk assessment procedures” (ISA 400). Such procedures by themselves do not provide sufficient appropriate audit evidence on which to base the audit opinion, but may be supplemented by further audit procedures in the form of compliance testing, and substantive procedures.

2.2.2 Compliance Testing

Gathering of evidence to verify that the accounting treatment of various transactions does not deviate from prescribed control procedures (e.g. an invoice is dated in the correct way, a cheque has all necessary signatures, and vouchers are filled out correctly) is referred to in the audit standards as “ test of controls” or “compliance testing”. This gives a quantitative measure of how effectively the internal control system works. As ISA 400 says “auditors test the operating effectiveness of the internal control system of their client, where necessary”. This is usually to ensure that the control relied on is operating as prescribed. Here it is the control which is being tested and not the transaction.

The auditor draws a conclusion as to whether the control was operating as intended for the period covered by the test. This conclusion is used by the auditor to decide how much audit assurance regarding the completeness and accuracy of the data can be placed on the internal control system, and much of this assurance needs to be obtained from substantive testing. Due to inherent limitations of the internal control structure, for example the risk of management override, the potential for human error due to carelessness, distraction, mistakes of judgement, misunderstanding of instructions, and the effect of system changes, auditors are not permitted to rely completely on the internal control systems, (ISA 400). They are required, in all cases, to make some substantive test.

2.2.3 Substantive Procedures

These are designed to obtain evidence concerning the validity and accuracy of transactions, balances, and the various elements of financial statements. The nature, timing, and the extent of substantive procedures are based on the auditors’ (i) inherent risk assessment, (ii) control risk assessment and (iii) preliminary materiality judgements. There are two main types of substantive testing:

(a) Analytical Review

Analytical review involves comparing current financial information with those for previous period (e.g. year) to see if there are any significant trends or variances. Various financial ratios are calculated, which sometimes involve regression analysis. According to ISA 520, analytical procedures compare an entity's financial information with, for example:

- Comparable information for prior periods;
- Anticipated results of the entity, such as budget or forecasts, or expectations of the auditor such as depreciation;
- Similar industry information, such as a comparison of the entity's ratio of sales to the accounts receivable with industry averages, or with the same ratio in comparable entities in the same industry.

(b) Substantive Tests of details

These are concerned with the examination of the correctness of recorded monetary values in a financial statement. These tests provide direct evidence about the accuracy of total recorded monetary values. The auditor either applies substantive tests of detail extensively, or applies compliance tests to see if reliance on those controls are efficient and effective in reducing the tendency of material error in accounts; if so, then although substantive procedures are required, they would not be as extensive as for a full verification approach. The choice of the approach depends on the efficiency and effectiveness of the internal control system, and also the cost of compliance testing compared with that of substantive testing (ISA 530).

2.2.4 Dual-Purpose Tests

According to AICPA *Audit Sampling* (1992), in some circumstances an auditor may design a test that will have a dual purpose: testing for compliance with prescribed control procedures and with regard to the value of the recorded balance or class of transactions in which they occur.

2.3 Types of Errors in Accounting System

Generally, there are two different types of errors in accounting system:

Errors of Commission: Occur when errors are committed on inputs to the accounting system. An error, in accordance with ISA 530, means either control deviation, when performing compliance tests, or misstatements, when performing substantive test of details. Misstatements could be caused by human error or fraud. *Fraud* is defined as intentional acts designed to deceive another person causing him financial loss. An error that arises from an isolated event that has not recurred, other than on specifically identifiable occasions and therefore not reflective of similar errors in the accounting population, is referred to as an *anomalous error*. An example of this would be an error caused by a computer breakdown that is known to have occurred on only one day during the period.

Errors of Omission: Occur when amounts are not entered into the accounting system.

2.4 Audit Evidence

Audit evidence is obtained from an appropriate combination of compliance tests and substantive procedures, and is defined in ISA 500 as:

> Audit evidence is all of the information used by the auditor in arriving at the conclusions on which the audit opinion is based, and includes the accounting records underlying the financial statement and other information (para 30).

Methods of obtaining audit evidence include inspection, observation, inquiry and confirmation, computation and analytical procedures (ISA 500). The choice of appropriate methods is a matter of professional judgement in the circumstances. The application of these methods will often involve the selection of items for testing from an accounting population, which is composed of a finite set of accounts, called line items.

As mandated by the ISA 500, when designing audit procedures, the auditor should determine appropriate means of selecting items for testing. In the manual (AUS 514, p9), the following selection methods have been suggested:

(a) *Selecting all items* (100% examination): The auditor may decide it will be most appropriate to examine the entire population that make up an account balance or class transactions (or a stratum within that population). 100% examination is unlikely in the case of test of control; however it is more common for substantive procedures. For example 100% examination may be appropriate when the data constitutes a small number of large value items, when both inherent and control risks are high and other means do not provide sufficient appropriate audit evidence, or when the repetitive nature of a calculation or other process performed by a computer information system makes a 100% examination cost effective.

(b) *Selecting specific items:* The auditor may decide to select specific items from an a population based on such factors as knowledge of client's business, preliminary assessments of inherent and control risks, and the characteristics of the population being tested. The judgemental selection of specific items is subject to non-sampling risk. Specific items selected may include:

- *High value or key items*: The auditor may decide to select specific items within a population because they are high value, or exhibit some other characteristic, for example items that are suspicious, unusual, particularly risk-prone or that have history of error.

- *All items over a certain amount*: The auditor may decide to examine items whose values exceed a certain amount, so as to verify a large proportion of total amount of an account balance or class transactions.

- *Items to obtain information*: The auditor may examine items to obtain information about matters such as the client's business, the nature of transactions, accounting and internal control systems.

- *Items to test procedures*: The auditor may use judgement to select and examine specific items to determine whether or not a particular procedure is being performed.

  While selective examination of specific items from an account balance or class of transactions will often be an efficient means of gathering audit evidence, it does not constitute audit sampling. The results of procedures applied to items selected in this way cannot be projected to the entire population. The auditor considers the need to obtain appropriate evidence regarding the remainder of the population when that remainder is material.

(c) *Audit sampling:* The auditor may decide to apply audit sampling to an account balance or class of transactions. Audit sampling can be applied using statistical or non-statistical sampling methods. Audit sampling is discussed in detail in the next section.

Remark: The selection procedures in (b) above is purposive (non-random) where as the selection procedure in (c) is random.

2.5 Audit Sampling

Audit sampling deals with the application of auditing procedures to *less than* 100 percent of the items within an account balance or class of transactions such that all sampling units have a chance of selection, in accordance with ISA 530. The purpose of the sample is to evaluate some characteristics of the balance or class. Auditors may use a statistical or non-statistical sampling (judgmental sampling) approach (ISA 530). Sampling is used in both compliance and substantive testing, and is treated in various standard textbooks in auditing (see e.g. Arkin 1984, Guy, Carmichael, and Whittington 1994).

According to AICPA Audit Sampling (2001), cost-benefit is the determining factor of using either a statistical or non-statistical sampling approach. The AICPA statement on auditing standard No. 39 (1981) says that either approach is valid, and requires that the auditor use professional judgment in "planning, performing, and evaluating a sample and in relating the evidential matter produced by the sample to other evidential matter when forming a conclusion about the related account balance or class of transactions". Any choice of approach depends on the circumstances.

The theory and practice of audit sampling is detailed in Bailey (1981), Arkin (1984), the AICPA *(*1992*).*

2.5.1 Statistical Sampling

A sampling approach is considered *statistical* if the selection of sampling items are random, each item having a calculated probability of being selected. Inferences about the population parameters may be made from the sample statistics. Random sampling enables the auditor to project sample results mathematically and to state, with measurable precision and confidence, the estimated rate of deviation in the population under audit (compliance audit sampling), or the estimated monetary misstatement in the population (substantive audit sampling) (Carmichael and Benis 1993). The most important benefit which statistical sampling offers is reduction of the risk of overauditing or underauditing.

2.5.2 Non-statistical Sampling

A sampling strategy is said to be *non-statistical* if either the selection method is non-probabilistic, or the result from a probabilistic sample are evaluated judgementally.

We next discuss the use of statistical sampling and inference procedures in financial auditing.

2.5.3 Role of Statistical Techniques in Auditing

The auditor's ultimate desire is to plan audits in a way that minimizes the total expected cost of performing the audit procedures while also giving a fair opinion on the financial statement. Sampling is therefore important in meeting these requirements. There are two audit procedures for which statistical sampling has been utilized. These are compliance and substantive tests. Statistical sampling in auditing seeks to assist auditors to use random selection methods and statistical evaluation techniques in testing, whether for compliance or substantive purposes. The objective is to reduce the risk of biased selection and quantify the sampling confidence level achieved. We discuss various statistical methods used in compliance and substantive tests in the next section.

We conclude this section with the general guidance on the use of sampling in an audit of financial statements provided by the ISA 530, which states that there are three main considerations in audit sampling, namely (a) the selection of the sample, (b) its size and (c) how the results are evaluated. These considerations are described as follows:

(a) *Selection of the Sample:* As ISA 530 says: "the auditor should select a sample that is representative of the population".

(b) *Determining the sample size:* According ISA 530, when determining the sample size, the auditor should consider sampling risk, tolerable error, and expected error.

- *Sampling risk* is "the risk that the auditor's sample will yield a conclusion different from the conclusion that would be reached if the entire population were tested. The lower the sampling risk that the auditor is willing to accept, the larger the sample will need to be".
- *Tolerable error* is "the maximum error in the population that the auditor would be willing to accept and still conclude that the result from the sample has achieved the audit objective. The tolerable error should be related to the auditor's judgement about materiality levels. The smaller the tolerable error, the larger the sample size will need to be".
- *Expected error* is "the error the auditor expects to be present in a population. If the auditor expects an error, a larger sample should be taken to ensure that the actual error is not larger than the planned tolerable error".

(c) *Evaluation of the results:* According to ISA 530, after performing tests of control or substantive procedures on one sample, the auditor should:

- Analyze any error detected in the sample;
- Project the errors to the population;
- Reassess sampling risk.

It is an auditor's responsibility to choose testing methods which is in his/her professional judgement are sufficient to satisfy generally accepted auditing standards. In the next section we review some of the statistical sampling methods used in substantive testing.

2.6 Statistical Sampling Methods in Auditing

Items to be reviewed during an audit are usually selected through one the following probability sampling methods: simple random, stratified random, systematic and probability proportional to size sampling, and therefore it is useful to provide a quick overview of these methods. The set of accounts to be audited is assumed to consist of N recorded (book) amounts, denoted by $y_1, y_2, ..., y_N$. The total recorded amount and the progressive sum of the recorded amount shall be denoted by $T_y = \sum_{i=1}^{N} y_i$ and $Y_i = \sum_{j=1}^{i} y_j$ respectively.

2.6.1 Simple random sampling (SRS)

In this case a sample of line items of the fixed size n is drawn one by one with the same probability but without replacement. That is each draw is carried out among items that have not already been chosen. There are thus $N!/(N-n)!$ samples each consisting of a combination of n of the N line items, and each such sample item has the probability $(N-n)!/N!$ of being selected. This is detailed by Cochran (1977) and Särndal, Swensson, and Wretman (1992). Robert (1978) gives a detailed account of the simple random sampling in auditing.

2.6.2 Stratified random sampling (STRS)

Stratified random sampling in auditing consists of dividing the auditing population into strata according to the sizes of the recorded amounts, and then selecting sampling items from each stratum independently by simple random sampling without replacement. Cyert, Hinckley, and Monteverde (1960) introduced the idea of achieving greater sampling efficiency through stratified sampling in auditing. A number of methods of stratifying audit populations effectively have been suggested, for example by Arkin (1974) and Robert (1978).

2.6.3 Systematic Sampling

Systematic sampling of line items is carried as follows: For a desired sample size n, a fixed sampling interval $I = N/n$ is calculated, where I is an integer. A random number r is chosen between 1 and I, and the subsequent sample consists of items r, $r+I, ...,$ $r+(n-1)I$ of the population with corresponding recorded amount $y_r, y_{r+I}, ..., y_{r+(n-1)I}$ respectively.

2.6.4 Probability proportional to size sampling (PPS)

PPS selection is sampling with unequal probabilities of selecting items. If items with larger values are relatively more important, then sampling with probability proportional to size will be useful. With PPS selection a sample of line items are taken in such a way that the inclusion probability, π_i is proportional to y_i, that is $\pi_i = ny_i/T_y$, provided that $y_i < T_y/n$. This design implicitly stratifies the sample by recorded amount.

PPS was originally developed in survey sampling theory by Hansen and Hurwitz (1943) for selection of clusters of unequal size. In the auditing context, this method and its variations are referred to as *Monetary Unit Sampling* (MUS) (see for example The AICPA's *Audit Sampling Guide* 1992), also known as *Dollar Unit Sampling*.

The idea of using individual monetary values as the sampling units was suggested by Deming (1960). The basic concept of monetary unit sampling in auditing was developed independently, first by van Heerden (1961) and later by Stringer (1963) and Stephen. van Heerden suggested that an account balance or the line item could be regarded as a cluster of monetary units being either correct or in error. A monograph by Meikle (1972) discussed an early version of MUS. Monetary unit sampling was made popular by the work of Anderson and Teitlebaum (1973). Their article expatiated upon monetary unit sampling in a way that is understandable to audit practitioners. Leslie, Teitlebaum, and Anderson (1979) gave a detailed account of Monetary Unit Sampling.

There are a number of ways of achieving the inclusion probability, $\pi_i = ny_i/T_y$. We will discuss the main ones below:

Systematic PPS:

The N items in the population are listed in a random order, their y_i are cumulated and a systematic selection of n elements from a "random start" is then made as follows: For a desired sample size n, a fixed sampling interval $I = T_y/n$ is calculated. A random value r between 0 and I is chosen, r is called the random start. The item i is selected if $Y_{i-1} < r \leq Y_i$. The subsequent items j are selected if $Y_{j-1} < c \leq Y_j$, where

$$c = \{h : h = r + kI;\ k = 1, 2, ..., n-1\}.$$

Systematic selection is a well known procedure and the mathematical theory associated with this procedure is provided in Hartley and Rao (1962). Anderson and Teitlebaum (1973) suggested this selection method should be used in auditing, and it is widely used.

Unrestricted random selection of line items with PPS:

This is simple random sampling without replacement of the monetary values. The unrestricted random selection is performed by selecting n random monetary values, $\mathbf{r} =$

$(r_1, ..., r_n)$, such that $1 \leq r_j \leq T_y$ $(j = 1, ..., n)$. The item i is selected if $Y_{i-1} < r_j \leq Y_i$. This method was initially described in the auditing context by Meikle (1972) with the title; Cumulative-Monetary-Amounts (CMA) sampling. This method is also known as unrestricted Dollar-Unit sampling (Anderson and Teitlebaum 1973). It is an accepted audit sampling procedure.

Cell selection:

The method of cell selection was developed by Leslie, Teitlebaum, and Anderson (1979). It is implemented in the same fashion as systematic sampling except that an independent selection is made in each interval. For a desired sample size n, cell sampling is performed by dividing the population of monetary units into n cells of size T_y/n each. One monetary unit is selected from each cell. A random selection is made for each cell independently to identify the sample monetary values.

Sieve selection:

Sieve selection was developed by Rietveld (1978, 1979) to avoid the need to sort random numbers or accumulate the recorded value subtotals. Sieve sampling is performed by selecting a random number between 0 and T_y/n independently for each line item in population. If the random number for line item i is y_i or less $(i = 1, 2, .., N)$, then the item i is selected. If the random number is greater than y_i, then the line item is not selected. This method has the advantage of selecting distinct items. However, the achieved sample size could be greater than or less than the required sample size.

2.6.5 Statistical Sampling Strategies in Compliance Testing

In compliance tests the variable of interest is an error rate (proportion of transactions for which the internal control operates wrongly). Samples of transactions are used to make inferences about the error rate. Many of the statistical methods adopted for quality control have been utilised in compliance testing. These methods are often referred to in the auditing context as attribute sampling (Robert 1978). Based on the auditor's understanding of the accounting and internal control system, the attributes that indicate performance of a control, as well as possible conditions of deviation, are identified, e.g. failure to obtain suitable authorization for a purchase order, which does not necessarily lead to a monetary loss. The auditor generally makes a preliminary assessment of the rate of error he/she expects to find in the population to be tested and the level of control risk. This assessment is based on the auditor's prior knowledge or the examination of a small number of items from the population. The preliminary assessment is used by the auditor to design the audit sample and to determine the sample size.

The random selection methods typically used in compliance tests of internal control procedures include simple random sampling, systematic sampling, and some non-statistical sampling methods such as haphazard selection and block selection (see The AICPA *Audit Sampling Guide* 1992*).* Sampling is not applicable to compliance testing if the internal control procedures provide no documentary evidence of performance.

The main attribute sampling methods used in compliance testing are:

1. *Estimation sampling*: A simple random sample is selected without replacement. The sample is examined to find the number of deviations from control procedures. This is then used to estimate the upper limit on the rate of deviation from control procedures in the population at some desired confidence level.

   The major drawback of this method is that, unless the expected error rate (an anticipation of the deviation rate in the entire population) is estimated in advance, it is not possible to determine, in advance, actual sample size to meet the auditor objectives.

2. *Acceptance sampling:* This a specialized sampling scheme developed for use in quality control. Acceptance sampling involves selecting a fixed sample of items from a population and, based on criteria established in advance, the population is either accepted or rejected. This sampling scheme is used in tests of control and provides the auditor with grounds to decide whether or not a population with a given error rate is acceptable. Here, the auditor needs to specify the tolerable error rate (maximum population rate of deviation from prescribed control procedures that the auditor will tolerate without modifying the planned reliance on the internal controls) and the expected error rate in the population before carrying out the selection and examination. The difference between the tolerable error rate and the expected error rate is used to determine the sample size. If the evaluation of the sample results show that it is unlikely for the tolerable error rate to be exceeded, the population is accepted and the relevant internal controls relied on. Acceptance sampling plans were developed by Dodge and Roming (1959). There are many variations of this sampling scheme. One of the earliest is a double sampling plan. A further extension of double sampling is sequential sampling.

3. *Discovery sampling*: This is a statistical sampling method which provides an indication of the probability of finding at least one attribute in question. It is not designed to estimate the error rate in the population. This method has been used in tests of control when the objective is to discover one example of an error if errors are occurring at or above a given rate. It is a special case of acceptance sampling (for which the expected error rate is set at zero). This sampling approach is very useful when no errors are found in the sample, the conclusion being that the auditor can rely on the relevant internal control. On the other hand when one or more errors

are found, the auditor reduces his/her planned reliance on internal control. This method gives the smallest sample size possible.

The theoretical properties of these methods, which usually use the Binomial or Poisson approximations to the appropriate Hypergeometric distribution are well known. Their application in auditing was first suggested by authors such as Vance (1947), Vance and Neter (1956), and Arkin (1961).

The use of each of the above three attribute models depends on the auditor's purpose for the compliance tests. For example, if the auditor wants to estimate the rate of occurrence of compliance error in the population, estimation sampling is employed. On the other hand if the auditor's purpose is to determine whether the occurrence rate of compliance error is above or below a certain level, acceptance sampling is used. Discovery sampling is also used if the auditor wants to preset the error rate at zero.

The use of statistical sampling and inference procedures for compliance testing has not posed any special problem (see Robert 1978), and will not be dealt with any further in this work.

2.6.6 Statistical Sampling Strategies in Substantive Testing

The statistical sampling methods often used in substantive auditing include simple random, stratified random, systematic and probability proportional to size sampling. The problem in auditing is the evaluation of substantive tests of detail, which entails using a valid evaluation technique to determine whether the recorded amount could be materially misstated. According to SAS No. 39 the auditor should

> "project the error results of the sample to the population from which the sample was selected and should add that amount to the errors discovered in any item examined 100 percent. Regardless of whether the sample results support the assertion that the recorded amount is not misstated by an amount greater than tolerable error, the client may adjust the recorded amount of the account because of the errors identified in the population. The total projected error after the client has adjusted the recorded amount should be compared with the tolerable error (materiality) for the account balance or class of transactions" (AICPA statement on auditing standard No. 39, 1981).

The methods used for evaluating account balances can be grouped as classical or non-classical methods. The classical methods are the survey-sampling estimators such as the Horvitz-Thompson (1952) estimator, difference, ratio and regression estimators (see appendix A for the formulation of these estimators). These methods, which rely on the use of the central limit theorem, were first employed in analysing audit data. Many

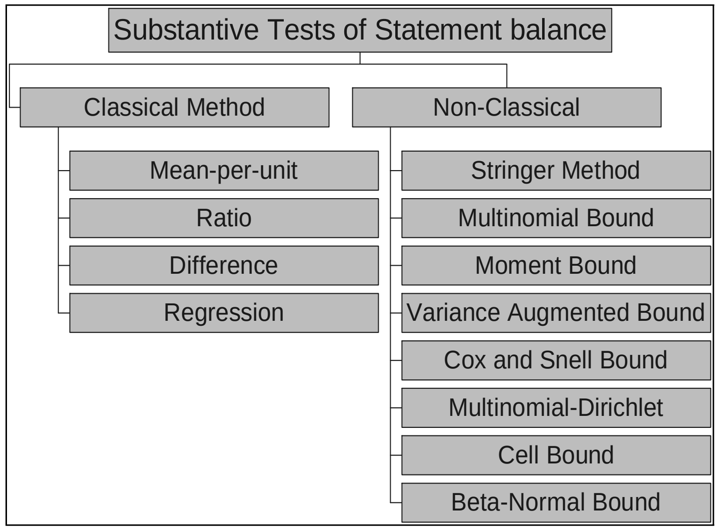

Figure 2.2: Statistical Sampling Techniques in Substantive Auditing

researchers including Stringer (1963), Kaplan (1973), and Neter and Loebbecke (1975, 1977) have acknowledged that the classical survey-sampling estimators are unreliable when making inferences about the total monetary error amount. An important feature of this problem is that a substantial proportion of the items in the population are usually error free, while the non-zero errors are highly skewed to the right (Johnson, Leitch, and Neter 1981 and Neter, Johnson, and Leitch 1985).

Much effort has been spent on developing alternative strategies (sampling designs and estimators) that might provide more satisfactory upper confidence bound for the total error amount (PNMD 1989). The alternative methods which are mostly heuristics and sometimes *ad hoc* are referred to as non-classical methods. Some of these methods which are in use among auditing practitioners are summarized in Figure 2.2 and their definitions will be detailed in Chapter 3. However, the confidence level attained by the non-classical methods, which are discussed in Chapter 3, could differ substantially from the nominal coverage, depending on the distribution of the error amount. So the development of new methods and approaches for setting appropriate confidence bounds for the total error amount in auditing data is a challenge for statisticians. This problem forms the core of much of the subsequent research.

2.7 Summary

We recall the basic procedures of financial auditing and also establish the role of statistical analysis in financial audits. Two main types of audit tests frequently rely on statistical sampling. First, when an auditor needs evidence to verify that the accounting treatments of various individual transactions comply with prescribed procedures. Second, when an audit requires evidence to verify that account balances are not materially misstated. To evaluate the results of a sample the auditor is required to project errors found in the sample to the entire population. For compliance testing, error projection does not involve any statistical difficulties. For substantive testing, the auditor projects the total error for the population to obtain an estimated total error amount. This poses a challenging problem when the auditing data is highly skewed and contains few errors. This problem is the subject of this research.

Chapter 3

Related Work

The aims of this chapter are to formulate the statistical auditing problem, to review the current state of the art of statistical sampling as applied to auditing, and to establish the need for the research. Research on the use of statistical sampling in financial auditing includes investigations of the behaviour of estimators when applied to audit populations. The range of statistical techniques employed in substantive auditing is quite diverse.

The plan of the chapter is as follows. Section 3.1 presents the notations and formulation of the statistical auditing problem. Section 3.2 reviews the empirical evidence on the characteristics of errors in accounting populations. Section 3.3 reviews the classical sampling estimation methods used for projecting the total error amount in substantive auditing and the problems associated with them. Section 3.4 through 3.7 review respectively the combined attribute and variable methods, likelihood ratio methods, methods based on asymptotic expansions, and Bayesian approaches to analysis of audit data. Comparative studies are reviewed in Section 3.8. A summary is given in Section 3.9.

3.1 Substantive Testing: Statement of the Problem

The account to be audited is assumed to consist of a set of N recorded (book) amounts, denoted by $y_1, y_2, ..., y_N$, and the total recorded amount by T_y :

$$T_y = \sum_{i=1}^{N} y_i. \tag{3.1}$$

The audited (true) amount of the N line items in the population will be denoted by $x_1, x_2, ..., x_N$, and the total audited amount by

$$T_x = \sum_{i=1}^{N} x_i \tag{3.2}$$

The error in item i, is defined by $z_i = y_i - x_i$, $1 \leq i \leq N$. When $z_i > 0$, the ith item is said to be overstated and when $z_i < 0$, it is understated. When $z_i = 0$, the account is said to be error free. The total error amount is denoted by

$$T_z = \sum_{i=1}^{N} z_i. \tag{3.3}$$

For $y_i \neq 0$, $t_i = z_i/y_i$ is called the fractional or prorated error (taint).

The values $(x_1, x_2, ..., x_N)$ are unknown and inaccessible before sampling, whereas $(y_1, y_2, ..., y_N)$ are known. We assume that the amount of any overstatement does not exceed the stated recorded value. The purpose of the audit is to estimate the total error amount T_z, on the basis of an examination of a sample of size n items in the account. The auditor's interest usually focuses on obtaining as accurate an upper bound on T_z possible, at a specified confidence level. If this upper bound exceeds the tolerable error amount allocated to the statistical tests of details, the auditor regards the statistical evidence as indicating the possibility of material error. When the upper confidence bound computed does not exceed the tolerable error amount, the auditor decides that the recorded value is a fair reflection of the accounts.

An important feature of this problem is that a substantial proportion of the items in the population are usually error free, while the non-zero errors are highly skewed to the right (Johnson, Leitch, and Neter 1981; Neter, Johnson, and Leitch 1985). The remainder of this chapter will review various estimators developed to provide more satisfactory upper confidence bound for the total error amount.

3.2 Characteristics of Accounting Populations

In testing the effectiveness of various sampling and estimation methods in substantive auditing, assumptions are required regarding the distribution of errors and rate of occurrence of errors in auditing populations. Here we review empirical articles on the characteristics of accounting populations, which will be used later in Chapter 8.

3.2.1 Distribution of Auditing Populations

Error characteristics are of great importance in determining the most appropriate and efficient sampling plan. In what follows, we provide an overview of the empirical evidence on characteristics of errors (mis-statements) in auditing populations. In particular, we discuss error types, (i.e. whether errors are overstated -the recorded value exceeds the audited value or understated -the audited value exceeds the recorded value), magnitude of fractional or prorated error and any other pertinent details.

Work by Stringer (1963), Kaplan (1973), and Neter and Loebbecke (1975, 1977) has indicated that auditing populations are highly positively skewed, that is the small-value items have a high frequency compared with the large-value line items.

Ramage, Krieger, and Spero (1979), Johnson, Leitch, and Neter (1981), and Neter, Johnson, and Leitch (1985) analysed a data set from Peat, Marwick, Mitchell, & Co. Approximately two-thirds of the data were accounts receivable audits and one-third inventory audits. The data set contains the sample error rate, the recorded and audited values for all sample items with error, as well as various characteristics of the recorded amounts in each of the accounting populations. The errors considered in their studies are those errors observed by the audit firm in audit samples for substantive tests. In each case the sampling method employed by the auditor is stratified random sampling.

Ramage, Krieger, and Spero (1979) found that this auditing population was highly positively skewed, and the distribution of error amounts for both inventory and accounts receivable is not normal. Also the proportion of overstatement and understatement errors differ for different types of population. About 80% of accounts receivable were reported to have mainly overstated errors, whereas errors in inventory include both overstatement and understatement with about equal frequency.

Empirical studies carried out by Johnson, Leitch, and Neter (1981) on the analysis of error distributions in 55 accounts receivable and 26 inventories gave the following findings:

i. The distribution of error amounts for both inventory and accounts receivable is not normal.

ii. Most errors in receivables audits are overstatements, whereas in inventory audits overstatement and understatement errors are about equal frequency.

iii. The distribution of overstatement and understatement error amounts are unimodal.

iv. Prorated error distributions are often degenerate at 1.0 and usually more than 10% of the line item have this degeneracy.

v. Large negative prorated errors are present in many inventories.

vi. Mean prorated error for accounts receivable tends to be 0.2 or higher.

vii. The shape of the prorated error distributions tends to be unimodal and negatively skewed for inventory and either reverse J-shaped or unimodal and positively skewed, with overstatement errors predominating, for accounts receivable.

Neter, Johnson, and Leitch (1985) also analyzed these data sets on the basis of monetary units. Their studies indicate that

i. The monetary unit prorated error distributions are much more concentrated near zero than the line-item prorated error distributions.

ii. The shape of the monetary-unit prorated error distributions are similar to those of the corresponding line-item prorated error distributions.

iii. The median monetary-unit prorated error rate for accounts receivable is 0.040 whereas the median monetary-unit prorated error rate for inventory is 0.186.

iv. The monetary-unit prorated error rates for both accounts receivable and inventory accounts tend to be higher than the line-item prorated error rates.

v. The size of the prorated error tends to vary inversely with recorded amount. The distributions of the proportion of overstatement errors in audits measured on a monetary-unit basis are similar to the corresponding line-item basis.

vi They inferred that the shapes of the prorated error distributions vary largely, that no one standard distribution (such as exponential distribution, gamma, beta and so on) can be used satisfactorily, in all the cases, to describe prorated error distribution.

Ham, Losell, and Smieliauskas (1985) investigated the error characteristics of five accounting categories using a data set from PriceWaterhouse, which consist of audit files of five annual audits for each of 20 companies. The following error characteristics were observed :

i. Error values for accounts receivable and sales are likely to be overstated, while accounts payable and purchase errors tend to be understatement.

ii. Mean error amount for accounts receivable tends to be greater than the mean error amount of the other four accounts.

iii. Error amounts are highly variable and are not normally distributed.

iv. The distributions of inventory, accounts receivable, and accounts payable errors are significantly different. Accounts receivable, and accounts payable errors are affected by individual firm characteristics which do not simply represent industry differences.

v. Accounts receivable prorated errors are larger than the prorated errors of the accounts payable, purchase, sales, and inventory. Inventory prorated errors are the smallest.

All the above studies on the error characteristics in audit population have been based on manufacturing and merchandising companies. Bell and Knechel (1994) extended the research by examining the characteristics of errors in audit of property and casualty insurers. They reported that the error amounts are not normally distributed, with many of the errors recurring.

3.2.2 The Rate of Error Incidence in Auditing Population

Ramage, Krieger, and Spero (1979) examined the rate of error incidence of 97 different audit populations. These studies used archived data of detected errors obtained from Peat, Marwick, Mitchell, & Co described in subsection 3.2.1. They reported that auditing populations have a low error occurrence rate and that the error rates vary substantially by both accounts receivable and inventory type of population.

Johnson, Leitch, and Neter (1981) also found great variation in the error rate for both accounts receivable and inventory using the same database. They reported that accounts receivable populations that they examined have first quartile, median and third quartile error rates of 0.4%, 2.4% and 8.9% respectively. Thus a random sample of 100 items from this distribution will contain, on the average, 2 non-zero errors. They also reported that inventory audit populations have first quartile, median and third quartile error rates of 7.73%, 15.4% and 39. 9% respectively. That is a random sample of 100 items from this distribution will contain, on the average, 15 non-zero errors.

Ham, Losell, and Smieliauskas (1985) examined the rate of error incidence, using a data set from PriceWaterhouse, mentioned in subsection 3.2.1. Their study expands upon previous work in two important directions. Firstly, they examined errors in accounts payable, purchases, sales, accounts receivable and inventory, whereas the previous studies (by Ramage, Krieger, and Spero 1979 and Johnson, Leitch, and Neter 1981) examined only the error characteristics in accounts receivable and inventory. Secondly, four error rates to describe error characteristics were defined and calculated. The previous studies concentrated on error frequency, and prorated error, but these studies, in addition to these two, included the proportion of monetary units in error and total discovered error amount as a proportion of the recorded value of the sample. The authors reported that the rate of error incidence is higher for receivable accounts than for inventory and the error rates differ significantly between accounting categories except for the proportion of the net monetary error. Ham, Losell, and Smieliauskas (1985) found that accounts payable have the highest number of errors in relation to the number of items tested.

Neter, Johnson, and Leitch (1985) reported that, among the accounts receivable populations that they examined, 40% had error rates below 2.5% and up to 73% of them had error rates below 12%.

3.2.3 Summary of the Characteristics of Accounting Populations

The foregoing studies provide some insight into the evidence on the characteristics of errors in auditing populations. The accounting populations are highly positively skewed (e.g. Stringer 1963, Kaplan 1973), and there is considerable diversity in the characteristics of error amounts in accounting populations across the accounting subsystem. The studies suggest that there is considerable variation in each account type, error amount

Table 3.1: Summary of the Principal Findings

Author(s)	**Account type**	**Error Rate**	**Error Distribution**	**Prorated Error Distribution**
Ramage et al. 1979	Inventory Receivable	low low	non-normal	N/A
Johnson et al. 1981	Inventory Receivable	.154 .024	non-normal	unimodal reverse J
Neter et al. 1985	Inventory Receivable	0.186 0.040	non-normal	unimodal reverse J
Ham et al. 1985	Inventory Receivable Payable Sales Purchases	low low moderately high low low	non-normal	N/A

distribution as well as in error rates and balance of errors between overstatement and understatement. The lack of normality of the error amounts distribution was noted by many researchers (e.g. Ramage, Krieger, and Spero 1979; Johnson, Leitch, and Neter 1981; Neter, Johnson, and Leitch 1985 and Ham, Losell, and Smieliauskas 1985).

The results of Johnson, Leitch, and Neter (1981) supports the results obtained by Ramage, Krieger, and Spero (1979). The finding of Ham, Losell, and Smieliauskas (1985) are, in general, consistent with Ramage et al. (1979), Johnson, Leitch, and Neter (1981), and Neter, Johnson, and Leitch (1985). One notable exception was that Ham, Losell, and Smieliauskas (1985) found that the rate of error incidence is higher for receivable accounts than for inventory. In Table 3.1, we list principal findings with authors.

These studies may be biased since the audit data sets used in the study is restricted to certain industries and clients and it is uncertain how representative it is of the mass of audit populations. Furthermore, the data is mostly confined to two accounting categories, accounts receivable and inventory. However, the studies have provided some insight into the characteristics of errors in auditing populations.

In the remainder of the chapter we consider the various estimation methods that have been proposed for computing the upper confidence bound on T_z and which are in use among the auditing practitioners, together with the problems associated with them, commencing with the classical estimators.

3.3 Estimating the Total Error Amount: Classical Approach

The estimators commonly used in the survey sampling such the Horvitz-Thompson (1952) estimator, the difference, ratio or regression estimator were first used in the estimation of the total error amount, T_z (see appendix A). These methods are described in standard textbooks such as Cochran (1977) and Särndal, Swensson, and Wretman (1992) in the general statistical context, and their application in auditing is discussed in textbooks such as Robert (1978), Bailey (1981), and Arkin (1984). The Horvitz-Thompson estimator with simple random sampling without replacement is often referred to in auditing context as *mean-per-unit*. The mean-per unit, difference, ratio and regression estimators are referred to in auditing literature as *variable sampling* (e.g. The AICPA's *Audit Sampling Guide 2001*). These estimators are used in auditing with SRS, STRS and PPS sampling.

The consideration of the above mentioned estimators in auditing started with Prythercy (1942). He raised numerous issues, including the need to stratify populations and examine the high-value stratum on 100 percent basis and perform some sampling in the lower strata. Hill, Roth, and Arkin (1962) discussed the application of stratified sampling and cluster sampling, as well as difference estimators in auditing. Elliot and Rogers (1972) integrated quantitative evidence about the internal control system, along with other audit procedures, in the determination of the required sample size for tests of details using statistical sampling methods. The classical survey-sampling estimators were implemented widely in auditing in 1970s and 1980s, their use has declined in the 1990s, (AICPA Auditing Practice Research and Education 1995, p94). The application of the Horvitz-Thompson (1952) estimator, the difference, ratio and regression estimators to estimate the total error amount, however, lead to certain problems which we now discuss.

Kaplan (1973) performed a simulation study to examine the behaviour of the mean-per-unit, difference, ratio, and regression estimators with sample sizes ranging from 25 to 200, and at 95 percent confidence level. He found that the sampling distribution of these estimators often deviate from the Normal or Student's "t" distribution. He also found that the estimates of the mean were highly correlated with estimates of the standard error. Kaplan (1973) hypothesized that high correlation between the mean of mean-per-unit, difference, ratio, and regression estimators and standard error estimates would inflate (reduce) the probability of a type II (type I) error.

The unsatisfactory performance of the mean-per-unit, the difference, ratio and regression estimators in highly skewed accounting populations with low error rates was also noticed by Neter and Loebbecke (1975, 1977). They extended Kaplan's work by empirically assessing the reliability of ratio, mean-per-unit and difference estimators using four real audit populations. Twenty study populations from the four real audit populations were simulated with error rates 0.5%, 1%, 5%, 10% and 30%. Sample sizes of 100 and 200 with both unstratified and stratified samples of line items were utilised. Neter and

Loebbecke (1975, 1977) observed that the confidence interval for difference and ratio estimators, whether with unstratified or stratified samples of audit units do not provide the nominal level when error rates are low or errors refer to overstatement. The stratified mean-per-unit estimator achieved a coverage near the nominal level when the error rates were low, but fell substantially below the nominal level by about six percent points for the ten percent error-rate study population for population 4 in Neter and Loebbecke (1977).

Neter and Loebbecke (1975, 1977) acknowledged that stratification improved the coverage probability of the mean-per-unit estimator. For the ratio and difference estimators they observed that stratification did not improve this probability when the error rate was low. It was also reported that increasing the sample size from 100 to 200 improves the coverage probability of the estimators but the substantial fall below the nominal 95 percent level as observed in the sample of size 100 remained.

The combined mean-per-unit and difference estimators and combined mean-per-unit and ratio estimators suggested by Jones (1972), were also investigated by Neter and Loebbecke (1975, 1977) using a weight of 0.1. They reported that the combined mean-per-unit and difference estimators and combined mean-per-unit and ratio estimators attained a coverage probability near the nominal level when the error rate was low or moderate for populations with low skewness. For highly skewed populations the coverage probabilities were found to be substantially below the nominal level.

As reported by Neter and Loebbecke (1975, 1977), the ratio and difference estimators had the smallest relative standard error for all the study populations they considered. The standard error for stratified mean-per-unit estimator was greater than that of the difference and ratio estimators. The stratified difference estimator tends to be more precise than the other estimators, although it frequently fails to attain the nominal coverage. The mean-per-unit estimator was inefficient compared to the difference estimator.

Further research on the performance of the classical estimators was carried out by Burdick and Reneau (1978) on the same audit population used by Neter and Loebbecke (1975, 1977) with systematic PPS. These authors observed that the coverage probabilities of these estimators were less than the nominal confidence level.

Baker and Copeland (1979) studied the performance of the stratified regression estimator, using a real audit population. They observed that the coverages for regression estimators were lower than the nominal level.

Beck (1980) used a method similar to Neter and Loebbecke (1975) to assess the performance of the regression estimator in auditing populations. Beck observed that the coverages for the regression estimator were lower than the nominal level, based on unstratified and stratified sampling of line items with sample sizes of 200 and 600.

Garstka and Ohlson (1979) suggested a variance augmentation technique for the Horvitz-Thompson (1952) estimator. The Garstka-Ohlson procedure uses a PPS sample of n items. The upper confidence bound for T_z is constructed as:

$$\widehat{T}_z + C_{m,n} s\left(\widehat{T}_z\right), \tag{3.4}$$

where

$$C_{m,n} = [\frac{np_n(m; 1-\alpha)}{m} - 1][\frac{nm}{n-m}]^{1/2}, \tag{3.5}$$

$\widehat{T}_z$ is the Horvitz-Thompson estimator, $s\left(\widehat{T}_z\right)$ is the standard error of $\widehat{T}_z$ and m is the number of errors observed in the sample ($m \geq 1$), and $p_n(m; 1-\alpha)$ is the $1-\alpha$ upper confidence limit for the binomial parameter p (error rate) when m errors are observed in the sample. Simulation studies by Garstka and Ohlson (1979) and Duke, Neter, and Leitch (1982) have shown, however, that the Garstka-Ohlson bound may not necessarily attain the nominal confidence level. A further analysis of this bound has been performed by Tamura (1985) and his study indicated that the applicability of the bound is severely restricted; specifically, lower bounds tend to be too conservative and upper bounds unreliable when there is a concentration of low prorated error.

Several theories have been put forward to explain the coverage problems of the ratio, mean-per-unit, difference and regression estimators. For example, Jones (1975) argued, that the coverage problems of the ratio, mean-per-unit and difference estimators were attributable not to correlation between the mean and standard error, as suggested by Kaplan (1973), but rather to the biased estimates of the standard error. Neter and Loebbeck (1977) noticed that the bias and correlation problems appear to be interrelated, and called for more research. Bias was also mentioned by Baker and Copeland (1979) as the cause of the coverage problems reported in their studies. Beck (1980) attributed the coverage problem of the regression estimator to heteroscedasticity which has been induced by the distributions of accounting errors and recorded values. Frost and Tamura (1986) also extended Kaplan's investigation of the performance of classical estimators, and attributed the coverage problem to the high skewness inherent in most auditing populations.

The concern of these coverage problems stimulated interest in academic research to further refine the classical methods in audit sampling, using computer intensive methods. We discuss these in the next section.

3.3.1 Computer Intensive Methods in Auditing

The use of computer intensive methods such as the bootstrap and the jackknife (in the auditing context) to improve the performance of classical estimators has been investigated by many researchers including Frost and Tamura (1982), Biddle, Bruton, and Siegel (1990) and Rohrbach (1993).

In particular, Frost and Tamura (1982) used the jackknife to improve the performance of the ratio estimator. The jackknife was introduced in statistical theory by Quenouille (1949, 1956) as a method of reducing bias. Tukey (1958) used it to estimate variance and calculate confidence intervals. Frost and Tamura utilised the *delete-1 jackknife, w*here the ratio estimator is consistently applied to all samples obtained from the original sample with just one sample value deleted.

These authors performed simulations using the Neter and Loebbecke (1975) data set to compare a jackknifed ratio estimator, with the conventional ratio estimator (with sample sizes of 50, 100 and 200). Their results suggested that the ratio estimator gives a reliable confidence interval when the error rate is high and the errors are composed of overstatement and understatement. The jackknifed ratio estimator was superior to the conventional ratio estimator but both estimators failed to attain nominal coverage for lower error rates. They acknowledged the need for new procedures to be developed for populations with low error rates.

Biddle, Bruton, and Siegel (1990) investigated the performance of computer-intensive methods in the auditing context, by applying the bootstrap to difference and ratio estimators. The bootstrap method is a computer based technique for estimating the distribution of an estimator by resampling and simulation. It was developed by Efron (1979, 1982) and described in detail by Efron and Tibshirani (1993). It provides a way to substitute computation for mathematical analysis if calculating the asymptotic distribution of an estimator or statistic is difficult.

Simulation studies were again carried out by Biddle, Bruton, and Siegel (1990) using the Neter and Loebbecke (1975) data set. They observed that the bootstrap difference and ratio estimators have the potential to increase reliability when compared with the conventional counterparts. It was also clear from their findings that these bootstrap estimators will not always attain the nominal confidence level.

Rohrbach (1993) argued that the failure of classical estimators to provide nominal coverage is not the results of incomplete sample information utilization. He applied a simple adjustment to the jackknife variance of the Horvitz-Thompson (1952) estimator using PPS without replacement. Rohrbach (1993) defines the bound for the population mean prorated error as

$$U_{1-\alpha} = M + z_{1-\alpha}\sqrt{(1-f)}\frac{s}{\sqrt{n}}$$

where $M = 1 - \frac{1}{n}\sum_i w_i,\;\; f = n/N,\; w_i = x_i/y_i,\; z_{1-\alpha}$ is the $100(1-\alpha)$ quantile of the standard normal distribution, and s^2 is adjusted jackknife variance of the mean-of ratio estimator given by

$$s^2 = \left(\frac{1}{n}\sum_i w_i^2\right) - (2 - \frac{2.7}{n})\left(\frac{1}{n(n-1)}\sum_i \sum_{i<j} w_i w_j\right).$$

This approach gives a non-zero bound when no error is detected in the sample, but will not always attain the nominal coverage under all conceivable error distributions, (as acknowledged by Rohrbach 1993, p96).

The failure of the classical survey-sampling estimators to give reliable confidence bounds, when used in sampling from highly skewed auditing populations with low error rates, has motivated a number of auditing practitioners and academics to explore other approaches. Various evaluation methods have evolved; these are grouped under the headings (i) combined attribute and variable method, (ii) likelihood ratio approach, (iii) asymptotic expansion approach and (iv) Bayesian approach. We discuss these methods in order.

3.4 Combined Attribute and Variable (CAV) Methods

Several statistical methods have been purposely developed for use in the auditing environment, mostly heuristic and sometimes *ad hoc*, to overcome the limitations of the classical methods based on the central limit theorem, particularly for populations with low error rates. These methods are based on sampling with probability proportional to recorded monetary value. However, they do not rely on a large-sample theory. These estimation methods are commonly referred to as *combined attribute and variable estimation* (Goodfellow, Loebbecke and Neter 1974). We discuss the most commonly used methods below.

3.4.1 Stringer Bound

The Stringer bound is one of the first CAV methods developed and is still the most widely used. It was introduced by Stringer (1963) and is a heuristic evaluation method. The Stringer-based method for obtaining an upper bound for the total overstatement error is given by

$$\widehat{T}_z(1-\alpha) = T_y\overline{\mu}_{st} \tag{3.6}$$

where

$$\overline{\mu}_{st} = p_n(0; 1-\alpha) + \sum_{i=1}^{n} \left[p_n(i; 1-\alpha) - p_n(i-1; 1-\alpha)\right] t_{n-i+1:n}, \tag{3.7}$$

for a sample of size n; and $0 \leq t_{1:n} \leq t_{2:n} \leq ... \leq t_{n:n} \leq 1$ are the order statistic of sample prorated errors arranged in decreasing order of magnitude. $p_n(i; 1-\alpha)$ is the $1-\alpha$ upper confidence limit for the binomial parameter when i errors are observed in a sample of size

n. That is to say, for a given α, n and number of errors i, we find the value $p_n(i; 1-\alpha)$ which satisfies

$$\sum_{j=0}^{i} \binom{n}{j} [p_n(i; 1-\alpha)]^j [1 - p_n(i; 1-\alpha))]^{n-j} = \alpha. \tag{3.8}$$

The Stringer bound is sometimes calculated using the Poisson approximation for obtaining the upper confidence limits $p_n(i; 1-\alpha)$.

The bound can be used with unrestricted, cell or systematic selection. The theoretical justification of the Stringer bound remains an important and interesting open problem (PNMD, 1989). There have been many studies on the Stringer bound. Much empirical evidence supports that its coverage is at least the nominal confidence level. It has been found to be overly conservative (e.g. Leitch et. al. 1982, Reneau 1978, Anderson and Teitlebaum 1973). As pointed out by Leitch et al. (1982), this conservatism often leads to the rejection of acceptable accounting populations.

Attempts to improve on the Stringer bound include the pioneering work of Bickel (1992) who studied the asymptotic behaviour of the bound. He showed that for a large sample, the confidence level achieved by the Stringer bound is most often greater than the nominal level. He proposed a new bound, which is a compromise between the Stringer and Gaussian bounds, behaving like Stringer when few errors are observed, and like the Gaussian when more errors are observed.

Pap and Van Zuijlen (1996) extended Bickel's work on demonstrating the asymptotic conservatism of the Stringer bound. They showed that the bound is asymptotically conservative for confidence level $1-\alpha$, when $\alpha \in (0, 1/2]$ and asymptotically not conservative when $\alpha \in (1/2, 1)$. On the basis of the asymptotic analysis, Pap and Van Zuijlen (1996) proposed a modified Stringer bound which is correct for very large samples. De Jager, Pap, and Van Zuijlen (1997) acknowledged that this modified method is not suitable when both the sample size and the error rate are small. They acknowledged the need for further enhancements of the Stringer bound in finite sample situations.

3.4.2 Multinomial Bound

Fienberg, Neter, and Letich (1977) developed a multinomial distribution-based method for evaluating a monetary unit sample. Their method was intended to be a less conservative approach to the Stringer bound. The authors treated the monetary unit sampling in a discretized form as multinomial sampling, where the multinomial classes represent the different prorated errors expressed to the nearest cent. The computation of the bound requires the use of nonlinear optimization methods. For this, let p_i be the probability that an item falls into the ith class, where $0 \leq p_i < 1$ and $\sum p_i = 1$. Then

$$t = \frac{i}{100} \text{ with probability } p_i,$$

and

$$T_z = \frac{T_y}{100} \sum_{i=1}^{100} ip_i.$$

Let d_i be the number of observations in a sample of n that falls into ith class, so that $\sum d_i = n$. Then $\mathbf{d} = (d_0, d_1, ..., d_{100})$ follows a multinomial distribution with parameters $(n, \mathbf{p})$, where $\mathbf{p} = (p_0, p_1, ..., p_{100})$ (assuming sampling is done with replacement). The multinomial distribution was therefore used to develop the confidence region for p_i. For a given confidence level $(1 - \alpha)$, the multinomial bound for total error amount, T_z, is obtained by solving the following maximization problem

$$\text{Maximize } T_z = \frac{T_y}{100} \sum_{i=1}^{100} ip_i \tag{3.9}$$

$$\text{subject to } \sum_{s} \frac{n!}{v_0!v_1!...v_{100}!} \Pi_{i=0}^{100} p_i^{v_i} \geq \alpha, \sum v_i = n \tag{3.10}$$

where S is the set of outcomes $\mathbf{v} = (v_0, v_1, ...v_{100})$ which are as extreme or less extreme than, the observed counts $\mathbf{d}$.

A difficulty with this method is the complexity of finding the joint confidence region and maximizing over it. This method has been used as a backup method by some auditing firms, but the difficulty even large mainframe computers had in computing the error bound has made the method unpopular, (AICPA *Auditing Practice Research and Education* 1995, p98).

The bound was modified in Leitch et. al. (1982) to make it computationally feasible. The authors grouped observed prorated errors into clusters of similar sizes, then all errors in the same cluster were assigned the maximum error value of the cluster. Extensive simulation studies by Neter, Leitch and Fienberg (1978) and Leitch et. al. (1982) indicate that the multinomial bound (unmodified or modified) is substantially tighter than the Stringer, although the former may fail to achieve the nominal coverage.

3.4.3 Cell Bound

The cell bound was proposed as a less conservative approach to the Stringer bound by Leslie, Teitlebaum, and Anderson (1979). This bound was developed for PPS with cell selection. However, can also be used with unrestricted or systematic selection.

Suppose that $t_{1:n} \geq t_{2:n} \geq ... \geq t_{m:n}$ are ordered sample taints; the upper error limit of the kth ordered sample taint, Λ_k, is determined as follows:

$$\Lambda_k = \max[\Lambda_k(1), \Lambda_k(2)]$$

where

$$\begin{aligned}\Lambda_0 &= \lambda_u(0; 1-\alpha) \\ \Lambda_k\,(1) &= \Lambda_{k-1} + t_{k:n} \\ \Lambda_k\,(2) &= \lambda_u(k; 1-\alpha)\frac{\sum_{i=1}^{k} t_{i:n}}{k}\end{aligned}$$

and $\lambda_u(k; 1-\alpha)$ denotes the upper $1-\alpha$ confidence limit for the Poisson distribution parameter λ when k errors are observed in the sample and m is the number of errors in the sample. The upper cell bound is computed as:

$$\widehat{T}_z(1-\alpha) = \frac{T_y}{n}\Lambda_m.$$

For auditing populations containing low error amount, the bound may be substantially greater than the actual error amount (Neter, Leitch, and Plante 1985).

3.4.4 Moment Bound

Another method of evaluating PPS sampling is the moment bound suggested by Dworin and Grimlund (1984, 1986). They constructed an upper bound for the total error amount by approximating the sampling distribution of the mean prorated error by a three-parameter gamma distribution. Using the sample moments, together with heuristic approximations, Dworin and Grimlund (1984) obtained estimates of the parameters of gamma distribution. The $1-\alpha$ upper moment bound for T_z is given by :

$$\widehat{T}_z(1-\alpha) = T_y\widehat{w}(1-\alpha), \tag{3.11}$$

where $\widehat{w}(1-\alpha)$ is defined as

$$\widehat{w}(1-\alpha) = G + A * B[1 + z_\alpha/\sqrt{9A} - 1/(9A)]^3, \tag{3.12}$$

where A, B and G are gamma distribution parameters given by $A = 4UC_2^3/UC_3^2$, $B = 0.5UC_3/UC_2$, $G = UN_1 - 2UC_2^2/UC_3$ and

$$\begin{aligned}UC_2 &= UN_2 - UN_1^2, \\ UC_3 &= UN_3 - 3UN_1 * UN_2 + 2UN_1\end{aligned}$$

where

$$\begin{aligned}UN_1 &= RN_1 * TN_1, \\ UN_2 &= [RN_1 * TN_2 + (n-1)RN_2 * TN_1^2]/n,\end{aligned}$$

$$UN_3 = [RN_1 * TN_3 + 3(n-1)RN_2 * TN_1 * TN_2 + (n-1)(n-2)RN_3 * TN_1^3]/n^2,$$

and

$$RN_1 = \frac{m+1}{n+2},\ RN_2 = \frac{m+2}{n+3}RN_1,\ RN_3 = \frac{m+3}{n+4}RN_2,$$

$$TN_j = [(t^*)^j + \sum_{i=1}^{m} t_i^j]/(m+1)\ ,\ \ j = 1,2,3,$$

$$t^* = 0.81[1 - 0.667\tanh(10\bar{t})] * [1 + 0.667\tanh(m/10)],$$

where $\bar{t} = \sum_{i=1}^{m} t_i/m$, n is the sample size and m number of non-zero errors. Their method is able to handle both overstatement and understatement errors.

Extensive simulation studies by Dworin and Grimlund (1984, 1986) suggested that the nominal coverage is close to the required. They compared their bound with the Stringer and multinomial bounds for accounts receivable. Their results indicated that the moment bound is less conservative than the Stringer bound, and would correctly accept more accounts than the multinomial bound when the materiality limit is less than 4% of total recorded value. However, for materiality limits in excess of 5%, the multinomial bound correctly accepted more accounts than the moment bound. The moment bound has sporadic coverage failures (Grimlund and Felix 1987).

3.4.5 Bounds based on Hoeffding's inequalities

Hoeffding's inequality can be used to obtain confidence bounds for the mean prorated error. Fishman (1991) described a confidence interval-estimation procedure based on this idea. In particular, when observations are drawn from a population with cumulative distribution function F with mean $\mu(F)$, he obtained a finite sample $100(1-\alpha)$ confidence interval for the population mean, $\mu(F)$, when F has bounded support, based on Hoeffding's inequality (1963) as follows: Suppose that $\Pr(0 \le T_i \le 1) = 1$, define

$$H(\varepsilon,\mu) = (\varepsilon+\mu)\ln(\frac{\mu}{\varepsilon+\mu}) + (1-\varepsilon-\mu)\ln(\frac{1-\mu}{1-\varepsilon-\mu}),\ \text{for } 0 < \varepsilon < 1-\mu$$

and let

$$\mu_L(\overline{T}_n) = \begin{cases} \{\mu : 0 < \mu \le \overline{T}_n \le 1 \text{ and } \exp(nH(\overline{T}_n - \mu, \mu)) = \alpha/2\}\,; & \text{for } \overline{T}_n > 0 \\ 0;\ \text{for } \overline{T}_n = 0 & \end{cases}$$

and

$$\mu_U(\overline{T}_n) = \begin{cases} \left\{\mu : 0 \leq \overline{T}_n \leq \mu \leq 1 \text{ and } \exp(nH(\mu - \overline{T}_n, 1-\mu)) = \alpha/2\right\}; & \text{for } \overline{T}_n < 1 \\ 1; \text{ for } \overline{T}_n = 1 \end{cases}$$

Fishman (1991) showed that

$$\Pr[\mu_L(\overline{T}_n) \leq \mu(F) \leq \mu_U(\overline{T}_n)] \geq 1 - \alpha.$$

This bound is more conservative than the Stringer bound.

Hoeffding's inequalities based confidence bound has recently rediscovered by Bentkus and Van Zuijlen (2003). They described various forms of upper confidence bounds based on the inequality in Theorem 3 of Hoeffding (1963). The inequality in Theorem 3 of Hoeffding (1963) is best possible inequality that can be obtained from using the exponential moment method as noticed by Hoeffding (1963). A modification of this inequality allowing the use of a prior information and incorporating the estimation of the variance are discussed in Bentkus and Van Zuijlen (2003).

3.4.6 Combined bound

The combined bound proposed by Howard (1994) is based on a combination of a bootstrap approximation of the bound generated by Hoeffding's inequality and a modified nonparametric bootstrap-t method. Using Hoeffding inequality

$$\Pr\left(\overline{R} - \rho \geq c\right) \leq \left\{ \left(\frac{\rho}{\rho + c}\right)^{\rho + c} \left(\frac{1-\rho}{1-\rho-c}\right)^{1-\rho-c} \right\}^n, \tag{3.13}$$

where ρ is the population mean proportion of monetary unit that is correct,

$$\overline{R} = n^{-1} \sum_{i=1}^{n} (1 - t_i), \tag{3.14}$$

is the sample mean proportion of monetary unit that is correct, and c is some constant $(0 < c < 1 - \rho)$, Howard derived a $1 - \alpha$ lower confidence bound, L_ρ, for ρ, by replacing $\rho + c$ in the right side of relation (3.13) by $\overline{R}$, then setting the resulting expression to α, and solving for ρ. The $1 - \alpha$ upper confidence bound for the population mean prorated error was then obtained as $1 - L_\rho$.

Using the assumption that each sample of monetary values is resampled with replacement, many times with the bootstrap samples having the same size, n, Howard obtained a new $1 - \alpha$ upper confidence bound for the population mean prorated as

$$\widehat{w}(1 - \alpha) = 1 - \overline{R} + c,$$

where c is obtained by solving the Hoeffding inequality (3.13), replacing $\overline{R}$ by $\overline{R}^*$,and ρ by $\overline{R}$, then setting the resulting expression to α. Here, $\overline{R}^*$is the mean proportion of monetary units that are correct in each bootstrap sample. Simulation studies were carried out by Howard (1994) to examine the behaviour of the combined bound. The results showed that the combined bound is not uniformly better than the Stringer bound. However the combined bound seemed to offer a moderate advantage over the Stringer for accounts receivable with low error rates, low proportion of 100% overstatement and high proportion of understatement.

3.5 A Likelihood Ratio Approach

The confidence bound approach discussed in the previous section is based on heuristic approximations and *ad hoc* methods. An alternative methods based on the likelihood ratio has been suggested by Kvanli, Shen and Deng (1998), Chen, Chen, and Rao (2002) and Chen and Qin (2003). These methods are discussed below.

3.5.1 Parametric Likelihood Method

Kvanli, Shen and Deng (1998) suggested that if the non-zero values in given auditing data can be assumed to come from some appropriate parametric model, then the likelihood ratio test can be used to construct a two-sided approximate confidence interval for the mean error amount. The theory behind their procedure is the results of Wilks (1938) on the chi-square limiting distribution of the likelihood ratio statistics. They provided simulation results for which the distribution of error amount is assumed to be normal or based on an exponential density function. Their results indicated that the coverage attained by this method is very close to nominal when the auditing data follows the assumed parametric model.

The limitation of this procedure is that it depends on the choice of the underlying model. The bound may not perform up to expectations if the data departs from the model assumed. As evidenced by empirical studies, the error distributions in practice varies widely so that it is unlikely to be modelled by one standard distribution (e.g. see Johnson, Leitch, and Neter 1981; Neter, Johnson, and Leitch 1985).

3.5.2 Empirical Likelihood Method

A competitive method for constructing tests and confidence intervals is the empirical likelihood method introduced by Owen (1988). The empirical likelihood method has many advantages over normal approximation-based methods and the bootstrap for constructing confidence intervals (e.g. Hall and La Scala 1990). One of the interesting features of the empirical likelihood is that it is transformation invariant and respects range restrictions on parameters. A comprehensive account and update of developments in empirical likelihood method can be found in Owen (2001). Let $T_1, ..., T_n$ be independent and identically distributed real-valued random observations drawn from a population with cumulative distribution function F with mean $\mu(F)$ and F_n be the empirical distribution function of this sample. For a distribution supported on the $T_i's$, that is $F(\{T_1, ..., T_n\}) = 1$, let

$$\mathcal{F}_c = \{F \mid R(F) \geq c, F(\{T_1, ..., T_n\}) = 1\}$$

for some $0 \leq c \leq 1$, where $R(F)$ is the empirical likelihood ratio function define by $R(F) = L_n(F(t))/L_n(F_n(t))$ and $L_n(F(t)) = \prod_{i=1}^n (F(t_i) - F(t_i-))$. Let $T_{L,n} = \inf_{F\in\mathcal{F}_c} \int tdF$ and $T_{U,n} = \sup_{F\in\mathcal{F}_c} \int tdF$. Under a weak moment condition Owen (1988) showed that

$$\lim_{n\to\infty} \Pr(T_{L,n} \leq \mu(F) \leq T_{U,n}) = \Pr(\chi_1^2 \leq -2\log c),$$

where χ_1^2 denotes a random variable with chi-squared distribution of one degree of freedom. The interval $[T_{L,n}, T_{U,n}]$ is the empirical likelihood confidence interval for $\mu(F)$. This interval has a coverage error of $O(n^{-1})$ (see e.g. Hall and La Scala 1990). DiCiccio, Hall and Ramano (1991) showed that the coverage error of empirical likelihood confidence interval could be reduced to $O(n^{-2})$ by Bartlett correction.

Recently Chen, Chen, and Rao (2002) have suggested a confidence interval for the mean of a highly skewed data containing many zero based on a profile empirical likelihood ratio function. Their approach is based on results of Owen (1988,1990), which show that the profile empirical likelihood ratio function tends to the chi-square with one degree of freedom as the sample size tends to infinity. The authors acknowledged that this approach works well when the proportion of non-zero values is not too small, which is not the case in audit populations.

Chen and Qin (2003) proposed empirical likelihood-based confidence intervals for mean of a data with many zeros observations. Both direct application of Owen's empirical likelihood and Bartlett correction to Owen's empirical likelihood confidence intervals based on bootstrap were investigated. In their empirical likelihood approach, zero and non-zero observations were separated. Their simulation results indicated that empirical likelihood confidence intervals perform better than normal approximation based counterparts. However, there were still under-coverage problems with empirical likelihood confidence intervals. The under-coverage was improved to certain degree by their proposed empirical Bartlett correction.

3.6 An Asymptotic Expansion Approach

Suppose $x_1, x_2, ..., x_n$ are n independent and identically distributed observations with underlying distribution F, and let $S_n(x_1, x_2, ..., x_n)$ be some statistic with distribution

F_n. If, after suitable standardization, S_n is asymptotically normally distributed then one can obtain a higher order approximation to F_n using the first few terms of the Edgeworth expansion (see Appendix B). This method is based on the Fourier inversion of an approximate characteristic function of the statistic. The inversion of Edgeworth expansion yields the so called Cornish-Fisher expansion, which provides an asymptotic expansion for an $\alpha-$level quantile of a statistic in terms of the $\alpha-$level quantile of the standardized normal distribution.

Helmers (2000) derived an upper confidence bound for the total error amount based on the empirical Cornish-Fisher expansion and bootstrap calibration. Using a simple random sampling of the line items, he obtained an empirical Cornish-Fisher bound on the total error amount. However, Helmers observed that the coverage probability of Cornish-Fisher bound may not achieve the nominal confidence level, and suggested direct use of bootstrap calibration of the nominal coverage probability. That is, resample the coverage probability, with α replaced by δ, for a grid of values of δ in $(0, 1)$, and select the largest value $\widehat{\delta}$ for which the bootstrap estimate is at least $1 - \alpha$ (see Hall and Martin 1988). Helmers (2000) also acknowledged that this method is not suitable for a sample size as small as 100 and also for cases where no error or one error is observed. The sampling selection method used is not the preferred sample design in financial auditing, which is PPS.

3.7 A Bayesian Approach to Analysis of Audit Data

Up to now we have been largely concerned with methods which are based solely on the sample data. We have paid no heed to the fact that the auditor may have prior information about the data which can be explicitly incorporated into the estimation method. This section reviews various methods that have been proposed, which allow the auditor to explicitly incorporate his/her prior information about the data into the estimation method.

3.7.1 Parametric Bayesian Approach

Cox and Snell (1979) studied the statistical auditing problem, and proposed a parametric Bayesian approach based on PPS sampling. In their approach, a finite population of the recorded values was treated as a realization from a superpopulation and errors constrained to be positive. This work provides a theoretical basis for the widely used monetary unit sampling. Cox and Snell split the population mean prorated error into two independent variables: unconditional error rate, π, and conditional mean prorated error, μ. Assuming exponential distribution with mean μ for the prorated error, they considered conjugate gamma priors:

$$\pi \sim \Gamma(a/\pi_0, a) \text{ and } \mu \sim (b-1)\mu_0/\Gamma(1, b),$$

where $\Gamma(a, b)$ denotes the distribution with density $a^b x^{b-1}\exp(-ax)/\Gamma(b)$, and parameters a and b are the scale and shape parameters, respectively. Cox and Snell then obtained the density of a posterior distribution of the population mean prorated error, w, as:

$$f\left(w|m\bar{t}\right) = \left[\frac{(b-1)\,\mu_0 + m\bar{t}}{(a/\pi_0 + n)}\right]\left[\frac{a+m}{b+m}\right] F\left[2\left(a+m\right), 2\left(b+m\right)\right], \tag{3.15}$$

where $F\left[2\left(a+m\right), 2\left(b+m\right)\right]$ denotes the density of the F distribution with $2\left(a+m\right)$ degrees of freedom in the numerator and $2(b+m)$ degrees of freedom in the denominator, and n is the sample size. Given the assessed values for the prior parameters, the observed sample values m and $\bar{t}$, the $(1-\alpha)$ upper bound for w is given by

$$\widehat{w}(1-\alpha) = \left[\frac{(b-1)\,\mu_0 + m\bar{t}}{(a/\pi_0 + n)}\right]\left[\frac{a+m}{b+m}\right] F\left[1-\alpha; 2\left(a+m\right), 2\left(b+m\right)\right], \tag{3.16}$$

where $F\left[1-\alpha; 2\left(a+m\right), 2\left(b+m\right)\right]$ denotes the $100(1-\alpha)\,\%$ of the F distribution with $2\left(a+m\right)$ and $2(b+m)$ degrees of freedom. Hence the $(1-\alpha)$ upper bound for T_z, the total error amount is given as:

$$\widehat{T}_z(1-\alpha) = T_y\widehat{w}(1-\alpha). \tag{3.17}$$

The drawback of this method is that the sample prorated errors are assumed to be random observations from an exponential distribution; but as indicated in Section 3.2, the error distribution, in practice varies widely and is unlikely to be modelled by one standard distribution. In this case, the bound may not perform up to expectations if there are departures from model assumptions.

Godfrey and Neter (1984) analysed the Cox and Snell model. They noted that the Cox and Snell bound contains a number of simplifying assumptions, for example use of an exponential distribution for prorated errors, which does not recognize an upper bound of 1.0 for the overstatement prorated error, as well as failure to recognize the frequent occurrence of prorated error degeneracy at 1.0. Godfrey and Neter developed four modified models in which these and other assumptions were relaxed. They performed simulation studies to further investigate the modified versions of the Cox and Snell model. Their simulation results suggested that the modified bounds were tighter than Stringer and cell bounds.

Menzefricke and Smieliauskas (1984) also proposed a parametric Bayesian model which assumes that the density of non-zero prorated error is normal. Their model is a modification of the general model developed in Felix and Grimlund (1977). One shortcoming of this model is that there is no empirical evidence to suggest that the prorated errors are normally distributed (e.g. see Johnson, Leitch, and Neter 1981, Neter, Johnson, and Leitch 1985). This bound has sporadic coverage failures (Grimlund and Felix 1987).

McCray (1984) proposed a quasi-Bayesian model which gives a discrete posterior distribution on the expected total error amount. He assumed a multinomial sampling model for prorated errors. McCray selected those multinomial parameter values for a given total error amount that maximizes the likelihood of the observed prorated errors. He did not provide any theoretical justification for using a maximum likelihood. The model can evaluate any monetary unit sample with any number of overstatement and understatement errors. This model was assessed by comparing its upper bound with that of the multinomial (Fienberg, Neter, and Letich 1977, and Neter, Leitch and Fienberg 1978). McCray argued that these comparisons strongly suggest that his proposed model is a reasonable approach for evaluating monetary unit sampling.

3.7.2 Nonparametric Bayesian Approach

A nonparametric Bayesian model has also been suggested, as a standard distribution does not appear to work well for modelling the distribution of prorated errors.

Tsui, Matsumura, and Tsui (1985) proposed nonparametric Bayesian method for constructing confidence bounds for the population total error amount. Their approach is based on the multinomial sampling model suggested by Fienberg, Neter, and Letich (1977). The auditors' prior prediction of the discrete distribution of the prorated errors was modelled by a Dirichlet distribution. Using simulation, Tsui, Matsumura, and Tsui (1985) suggested certain parameters to be used as the prior setting for the upper bound. They argued that those parameters performed well under repeated sampling for a wide range of prorated error distribution.

Tamura (1988) pointed out that the error distribution was unlikely to be modelled by any standard parametric distributions. He proposed a nonparametric Bayesian approach using Ferguson's Dirichlet process to specify the prediction of the conditional distribution of the error. He modelled the error rate by a beta distribution and hypothesized that the auditor cannot predict the exact form of error distribution, but is able to describe the expected form. He argued that the auditor may either use any standard parametric model as his best prior prediction or base it on the past data. Tamura makes inferences on the expected aggregated error amount through numerical inversion of the characteristic function.

Laws and O'Hagan (2000) extended Tamura's nonparametric Bayesian model with an extra layer of modelling, by introducing a classification of errors according to prorated errors. Laws and O'Hagan make inference for the aggregated error amount through simulation.

The implementation of all the Bayesian methods, requires the assessment of prior probability distribution. A number of studies (e.g. by Corless 1972, Felix 1976, Crosby

1981) have suggested that auditors may have some difficulties in providing suitable information about prior.

3.8 Comparative Studies

Several authors have investigated the merits and drawbacks of the various estimators proposed in substantive tests of details.

In particular Menzefricke and Smieliaukas (1984) investigated the comparative performance of the Stringer and Cox and Snell bounds. Their analysis focused on the power of the performance of the bounds. The study confirmed that the Stringer bound tends to provide poor protection against rejecting populations whose total error amount is less than materiality.

Neter, Leitch, and Plant (1985) analysed the multinomial, Stringer, cell and stratified difference bounds. It was reported in their findings that the multinomial bound is substantially tighter than the Stringer, and somewhat tighter than the cell bound, when each bound is based on cell selection. It was also reported that the multinomial bound is not as tight as the stratified difference. However, the multinomial bound has coverage probabilities near or above the nominal level, whereas those for the stratified difference are below the nominal level. The coverage probabilities for the Stringer and cell bound were reported to be substantially higher than nominal level. They also observed that when the true aggregated error amount is substantially smaller than the materiality amount (tolerable error), the probability of rejection for the multinomial seems to be moderately low and at about the same level as those for stratified difference bound, whereas that for Stringer and cell are higher. Neter, Leitch, and Plante (1985) also reported that systematic or cell selection of the monetary units do not have a substantial effect on the sampling distribution of the multinomial, cell and Stringer bounds when line items are in random order.

Smieliauskas (1986) studied the comparative performance of Bayesian and non-Bayesian methods in monetary unit sampling evaluation, using an approach pioneered by Leitch et al. (1982). The non-Bayesian bounds considered by Smieliauskas include the Stringer bound, the multinomial bound, the cell bound, and the moment bound. The Bayesian bounds he considered are the Cox and Snell bound and normal Bayesian (Menzefricke and Smieliauskas 1984). The general conclusion of his study was that the non-Bayesian bounds achieved coverage close to the nominal in most cases, which were accompanied by the lack of tightness whereas the Bayesian bounds had coverage failures but tighter than non-Bayesian bounds.

Dworin and Grimlund (1986) compared the performance of the moments bound with that of McCray's quasi-Bayesian bound using the uniform prior. The two bounds achieved coverage probability higher than the nominal level of 95% used for their study populations. Both bounds were considerably tighter than the Stringer bound.

Grimlund and Felix (1987) carried out simulation studies to compare four different monetary error confidence bounds: these are the Cox and Snell, moment, Bayesian-Normal and multinomial-Dirichlet bounds. It was reported that the multinomial-Dirichlet method achieved a coverage probability near the nominal level in most cases, the Cox and Snell bound failed to attain the nominal coverage in some cases, whereas the moment bound and Bayesian Normal bound had sporadic coverage failures.

Wurst, Neter, and Godfrey (1989) compared sieve sampling with random and cell sampling of monetary unit in terms of the behaviour of the Stringer and cell bounds. A simulation study based on two actual accounting populations used in Neter and Loebbeck (1975) was employed. It was reported that the method of sample selection has no substantial effect on the mean bound and the coverage of the bound. However, the sieve sample bounds were reported to be less variable than those for random selection for larger sample sizes and more variable than those for cell sampling for small sample sizes.

Matsumura et al. (1991) performed simulation studies to compare the lower multinomial bound and multinomial-Dirichlet lower bound. Their studies suggested that the coverage probability of the two bounds were quite similar for almost all the 24 cases they considered. As reported by the authors, computational considerations favour selection of the multinomial-Dirichlet bound as a reasonable procedure for lower-bound estimation.

Horgan (1996) carried out a comparison of unrestricted random, cell and sieve sampling of monetary unit in terms of the behaviour of the moment bound. It was reported that the moment bound is more precise with cell and sieve sampling than with unrestricted random sampling.

Swinamer et al. (2004) performed an extensive simulation study comparing 14 bounds used in monetary-unit sampling using both real and simulated data. These include Stringer bound, augmented variance bound, modified moment bound, Cox and Snell bound, multinomial-dirichlet bound, and Claytons combined bootstrap Hoeffding/bootstrap-t bound. Their results suggested that no one method is superior in sense of giving smaller and reliable upper confidence bound.

3.9 Summary

In this chapter we overview the characteristics of errors in accounting population and the methods proposed to estimate the total error amount, along with the problems involve. We noted a wide range of statistical methods is used in auditing. We first looked at the classical estimators such as Horvitz-Thompson (1952) estimator, the difference, ratio and regression, then followed respectively by the CAV bounds, likelihood ratio methods, methods based on asymptotic expansions, and Bayesian methods.

The various research studies suggest that the survey sampling estimators provide accurate point estimates but their confidence intervals based on assumptions of asymp-

Table 3.2: Some Previous Work on Problems in Statistical Auditing and their Limitations

Author(s)	**Approach**	**Limitations**
Stringer (1963)	Heuristics	overly conservative
Fienberg et al. (1977)	Multinomial	computationally intensive
Leslie et al. (1979)	Cell	conservative
Cox and Snell (1979)	Parametric Bayesian	unreliable
Garstka and Ohlson (1979)	Heuristic	coverage failure
Menzefricke and Smieliauskas (1984)	Beta-Normal	sporadic coverage failure
Dworin and Grimlund (1984)	Moment	sporadic coverage failure
Tsui et al. (1985)	Multinomial Dirichlet	unreliable in some cases
Tamura (1988)	Dirichlet Process	difficult to implement
Rohrbach (1993)	Augmented Variance	coverage failure

totic normality of the estimators fail to attain the nominal confidence level when used in sampling from a highly skewed population containing a substantial proportion of zero values.

The CAV bounds have provided a frame for the main stream of results in computing an upper confidence bound on the total error amount in the population. The performances of these methods vary, and no one approach clearly dominates.

The Bayesian method provides more structured approaches for combining subjective evidence, such as that obtained from the evaluation of compliance tests and analytical review, with objective evidence obtained from statistical sampling for substantive testing in computing confidence bounds on the total error amount in the auditing population. However the elicitation of subjective priors or quantification of expert knowledge poses a problem, this makes the implementation of the Bayesian methods difficult.

One problem that the methods reviewed in this chapter have in common is that the confidence level attained by the bounds constructed by these methods could differ substantially from the nominal coverage, depending on the distribution of the error amount. A brief summary of the properties of the estimators proposed is given in Table 3.2, which shows that none of the estimators is overwhelmingly superior or preferable. This leaves an open problem for further research and development.

An approach which has not been explored extensively is the use of asymptotic expansion techniques such as the Edgeworth expansion and saddlepoint approximation. We explore this approach in this thesis. We also propose to obtain confidence bounds that maintain their coverage probability over broad class of distributions.

In the remainder of this thesis we develop improved bounds for estimating the total error amount in substantive tests.

Chapter 4

Estimation of the Total Error Amount Using Asymptotic Expansion Techniques

In this Chapter we develop a new estimator for estimating the total error amount. The sum of non-zero prorated errors in auditing data leads naturally to a compound Binomial distribution. This is approximated by a compound Poisson distribution, and using an empirical Edgeworth expansion for a studentised compound Poisson sum, we construct confidence bounds on the total error amount.

As a first step, we present the model formulation in error analysis in audit data in Section 4.1. In Section 4.2, we apply the theory of compound Poisson sum to the error analysis in audit data, and illustrate how an asymptotic expansion for a compound Poisson distribution can apply. In Section 4.3, we construct confidence bounds on the total error amount. In Section 4.4, we look at a particular numerical example taken from Leitch et al. (1982), which is chosen to exhibit the comparative performance of our proposed estimator with other estimators in the literature. Summary and discussion is provided in the final section.

4.1 The Auditing Issue

We are interested in constructing accurate non-parametric confidence bounds on the total error amount, T_z. Throughout the following analysis, we suppose that samples are drawn with systematic PPS as defined in Section 2.6.4. Since the population size, N, is usually large this design can be thought of as being implemented through almost identical independent draws. We also assume that the probability of an item being in error is independent of its recorded value. Our statistical model is formulated in terms of the prorated errors.

Suppose $T_1, T_2, ..., T_n$ are independent and identically distributed real-valued random variables (representing the prorated error) with common distribution function F, of the form

$$F(t) = \pi G(t) + (1 - \pi)\delta_0, \tag{4.1}$$

where π is the population proportion of monetary units in error, G is the distribution of non-zero prorated errors and δ_0 is the one point distribution function concentrated at zero (representing the correct values). We can, therefore, express the prorated error T as:

$$T = \begin{cases} V & \text{with probability } \pi \\ 0 & \text{with probability } 1 - \pi \end{cases} \tag{4.2}$$

where V is a random observation from the distribution G. Then the population mean prorated error can be written (Aitchison 1955) as

$$w = \pi E(V). \tag{4.3}$$

From Cox and Snell (1979), with $E(V) = \mu$, we have the total error amount

$$T_z = T_y \pi \mu.$$

That is

$$T_z = T_y w. \tag{4.4}$$

The total book amount, T_y, is known. Therefore to estimate T_z in (4.4) we need to estimate w.

Suppose m errors are observed in the sample of size n. Let V_i be the ith non-zero prorated error observation then a point estimate of w is:

$$\widehat{w} = \frac{1}{n} \sum_{i=1}^{m} V_i.$$

The problem now becomes that of finding a $(1 - \alpha)$ upper confidence bound on w. This is considered in the next section.

4.2 The Statistical Model of the Non-Zero Prorated Errors

To estimate $w = \pi E(V)$ in (4.3) we need to construct a model incorporating two aspects of auditing populations: the distribution of error rate, π, and the distribution of non-zero prorated errors, G. Then, based on this model, we aim to construct upper confidence bounds on the population mean prorated error, w. A useful way of expressing this model is through a compound Poisson processes which are often useful as approximate models when describing rare events (Feller 1971, p181).

4.2.1 The Number of Errors Distribution

The number of errors, M, observed in the sample of size n, is a random variable. The exact distribution for M is the Hypergeometric. One can view the accounting population at the time of auditing as a random sample of size N from an infinite population, so that the distribution of M is approximated by the Binomial. Thus M has a Binomial distribution,

$$B_{n,\pi}(M = k) =: \binom{n}{k} \pi^k (1-\pi)^{n-k}, \;\; k = 0, 1, ..., n. \tag{4.5}$$

In audit population π and $n\pi^2$ are usually small, therefore the distribution of M can be further approximated by the Poisson distribution,

$$P_\lambda(k) =: \exp\left(-\lambda\right) \left(\lambda\right)^k / k!, \;\; k = 0, 1, ... \tag{4.6}$$

where $\lambda = n\pi$. Here, we make no use of the classical requirement that $\lambda = n\pi$ is fixed, while $\pi \to 0$ and $n \to \infty$. The Poisson approximation to Binomial works well provided that π and $n\pi^2$ are small (Barbour, Holst, and Janson 1992).

4.2.2 Distribution of the Sum of Non-Zero Prorated Errors

Assume that $V_1, ..., V_M$ are independent and identically distributed random variables, each distributed as a random variable V (representing the nonzero prorated error), where M is a number of errors observed in the sample of size n. We assume further that the sequence $\{V_i\}$ is independent of the variable M. Define a random variable S as

$$S = \begin{cases} \sum_{i=1}^{M} V_i & \text{if } M > 0 \\ 0 & \text{if } M = 0 \end{cases}; \tag{4.7}$$

obviously S gives the sum of non-zero prorated errors observed in a sample of size n. Then with the assumption that the error occurrence in auditing populations constitute a Poisson process with constant rate $\lambda = n\pi$, S is a compound Poisson variable with parameter λ and component distribution G. That is the cumulative distribution of S is

$$\Pr[S \le x] = \sum_{k=0}^{\infty} p_k G^{*k}(x), \;\; x \ge 0$$

where $G^{*k}(x)$ is kth convolution of G with itself, $G^{*0}(x) = 1$ for $x > 0$, i.e.

$$G^{*k}(x) = \Pr\left(V_1 + V_2 + ... + V_k \le x\right)$$

and $p_k = \Pr[M = k]$ and so the cumulative distribution of S is given (Feller 1971, p159-181) by

$$\Pr[S \leq x] = \sum_{k=0}^{\infty} \frac{e^{-\lambda} (\lambda)^k}{k!} \Pr(V_1 + V_2 + ... + V_k \leq x). \tag{4.8}$$

Let $M_S(r) = E[\exp\{rS\}]$ denote the moment generating function of S, then

$$\begin{aligned} M_S(r) &= E[\exp\{rS\}] \\ &= \sum_{k=0}^{\infty} E[\exp\{rS\} \mid M(n) = k] P\{M(n) = k\} \\ &= \sum_{k=0}^{\infty} E[\exp\{r(V_1 + V_2 + ... + V_k)\} \mid M(n) = k] e^{-\lambda} \frac{(\lambda)^k}{k!} \\ &= \sum_{k=0}^{\infty} (E[\exp r(V_1)])^k e^{-\lambda} \frac{(\lambda)^k}{k!}, \quad \text{from independence of } V_i's. \end{aligned} \tag{4.9}$$

Letting $M_{V_1}(r) = E[\exp r(V_1)]$ denote the moment generating function of V_1, we have from (4.9) that

$$\begin{aligned} M_S(r) &= \sum_{k=0}^{\infty} (M_{V_1}(r))^k e^{-\lambda} \frac{(\lambda)^k}{k!} \\ &= \exp\{\lambda(M_{V_1}(r) - 1)\}. \end{aligned} \tag{4.10}$$

Differentiating $M_S(r)$ in (4.10) with respect to r, it is easy to see that the mean and variance of S are given by:

$$E[S] = \lambda E(V) \tag{4.11}$$

and

$$Var[S] = \lambda E(V^2). \tag{4.12}$$

Explicit evaluation of the cumulative distribution function of the sum of non-zero prorated errors, S, presented in (4.8) is impossible because of the complicated nature of the convolutions. So we have to resort to approximations.

Remark: The motivation for taking a random sum of the non-zero prorated error is that the standard deviation is underestimated when the number of summands is fixed.

4.2.3 Approximation of the Cumulative Distribution of the Sum of the Non-Zero Prorated Errors

We denote the moments of the distribution of V about the origin and about the mean of order r respectively as

$$\mu'_r = E[V^r]$$

and

$$\mu_r = E\left[V-\mu\right]^r, r = 1, 2, ...$$

obviously $\mu'_1 = \mu$ and $\mu_2 = Var(V) = \sigma_v^2$. To approximate the distribution function of S consider a standardized statistic:

$$\Upsilon_n = \frac{S - E[S]}{\sqrt{Var[S]}} = \frac{(\sum_{i=1}^{M} V_i - \lambda\mu)}{\sqrt{\lambda\mu'_2}}.$$

In practice, the asymptotic variance $Var[S] = \lambda\mu'_2$ would be unknown therefore it is useful to consider a Studentised version:

$$\Upsilon_t = \frac{(\sum_{i=1}^{M} V_i - \lambda\mu)}{\sqrt{\sum_{i=1}^{M} V_i^2}} = \frac{(\sum_{i=1}^{M} V_i - n\pi\mu)}{\sqrt{\sum_{i=1}^{M} V_i^2}}, \tag{4.13}$$

since $\sum_{i=1}^{M} V_i^2$ is a natural consistent estimator of the variance of $(\sum_{i=1}^{M} V_i - \lambda\mu)$, i.e. for all $\varepsilon > 0$.

$$\lim_{n\to\infty} \Pr\left(\left|\sum_{i=1}^{M} V_i^2 - \lambda\mu'_2\right| > n\varepsilon\right) = 0.$$

It is impossible to present the exact distribution of the statistic Υ_t in a tractable form. To construct a confidence bound on w we need to know the distribution of Υ_t. It is common practice to approximate the distribution of Υ_t using traditional large-sample techniques based on the central limit theorem. However, the distribution of the V_i is highly skewed in auditing populations, therefore for any sample size the sampling distribution of $\sum_{i=1}^{M} V_i$ will be skewed, as well as that of Υ_t. As a result, appropriate sample sizes for the application of the central limit theorem are likely to be larger than desirable in the audit context. The normal approximation or Student t distribution in general does not perform well even for moderately large sample sizes such as 50,100, or 200, when the skewness is severe, see Sutton (1993) and the references therein. In such situations, the confidence bounds constructed on the basis of the Student t distribution would have inflated significance level for the lower bounds and loss of power for the upper bounds (Sutton 1993).

It therefore seems natural to try to obtain a higher order approximation in the hope that this will give sufficiently good approximation to the distribution of function of Υ_t. We propose a transformation of Υ_t which removes the skewness effect of the auditing population. This transformation would be used as the basis for constructing accurate confidence bounds for the population mean prorated error, w. To do this, we require a one-term Edgeworth expansion for Υ_t (Abramovitch and Singh 1985 and Hall 1992a, 1992b), which is obtained in the next section.

4.2.4 The Edgeworth Expansion of the Cumulative Distribution of the Aggregated Non-Zero Prorated Errors

The Edgeworth expansion is an asymptotic expansion as $n \to \infty$ of the distribution of the standardised or studentized variable, $\Pr(\Upsilon_t \leq x)$, in terms of the corresponding normal distribution and in powers of $n^{-1/2}$ (see appendix B). A detailed treatment of the univariate Edgeworth expansion is given in Bardorff-Nielsen and Cox (1989, Chapter 4) and Hall (1992b). The comprehensive treatment of the classical theory of the Edgeworth expansion for the sample mean can be found in Chapters 4 and 5 of Bhattacharya and Rao (1976). The one-term Edgeworth expansion for a studentized statistic is usually of the form

$$\Phi(x) + n^{-1/2}p(x)\phi(x) + o(n^{-1/2}), \tag{4.14}$$

where $p(x)$ is a polynomial whose coefficients depend on the first few population moments and $\phi(x)$, $\Phi(x)$ denote standard normal density and distribution functions respectively. The $o(n^{-1/2})$ is a standard notation for some unspecified function f having the property that $\lim_{n\to\infty} \sqrt{n}f = 0$. That is $o(n^{-1/2})$ is negligibly small compared with $n^{-1/2}$ as $n \to \infty$. Asymptotic expansions for compound Poisson processes can be found in von Chossy and Rappl (1983) and Hipp (1985). Indeed, using Edgeworth expansions of distribution function of Υ_t (4.13) far more accurate approximations may be obtained for the distribution of Υ_t. However, the Edgeworth expansion for distribution function of Υ_t is not readily available. To obtain a one-term Edgeworth expansion for the distribution function of Υ_t Theorem 2.1 of Babu, Singh, and Yang (2003), which is stated as the following lemma is needed.

Lemma 1(Babu, Singh, and Yang 2003, Theorem 2.1): *Let $\{N(t), t > 0\}$, be a Poisson process with rate $\lambda > 0$. Suppose $X_1, X_2, ...$ are independent and identically distributed random variables independent of $\{N(t), t > 0\}$ and that $E(X_1{}^6) < \infty$ and the distribution of X_1* has *continuous component. Then uniformly in x, as $t \to \infty$,*

$$\Pr\left(\frac{(\sum_{i=1}^{N(t)} X_i - \lambda\mu t)}{\sqrt{\sum_{i=1}^{N(t)} X_i^2}} \leq x\right) = \Phi(x) + \frac{1}{6\nu^3\sqrt{\lambda t}}[\mu_3(1+2x^2) - \mu^3(x^2-1) + 3\mu\left(\sigma^2 + x^2\sigma^2 + x^2\nu^2\right)]\phi(x) + o(t^{-1/2}) \tag{4.15}$$

where $\mu = E(X_1), \sigma^2 = Var(X_1), \mu_3 = E(X_1-\mu)^3, \nu^2 = E(X_1^2)$.

Theorem 1 below provides the Edgeworth approximation of the distribution function of Υ_t up to terms of order $o(n^{-1/2})$.

Theorem 1: *Assume that* $M(n)$,*the number of errors observed in the sample of size* n, *follows a Poisson distribution with parameter* $\pi > 0$. *Suppose* $V_1, V_2, ...$ *are independent and identically distributed random variables independent of* $M(n)$ *and that* $E(V_1{}^6) < \infty$ *and the distribution of* V *has a non-degenerate absolutely continuous component. Then*

$$\begin{aligned}\Pr(\Upsilon_t \le x) &= \Phi(x) + \frac{1}{6(\mu_2')^{3/2}\sqrt{n\pi}}\{\mu_3(1+2x^2) - \mu^3(x^2-1) \\ &\quad +3\mu\left(\sigma_v^2 + x^2\sigma_v^2 + x^2\mu_2'\right)\}\phi(x) + o(n^{-1/2}) \qquad (4.16)\end{aligned}$$

uniformly in x, *as* $n \to \infty$.

Proof: Let $\{M(t), t > 0\}$, be a Poisson process with rate $\pi > 0$. Suppose $V_1, V_2, ...$ are independent and identically distributed random variables independent of $\{M(t), t > 0\}$ then by Lemma 1

$$\begin{aligned}\Pr\left(\frac{(\sum_{i=1}^{M(t)} V_i - \pi\mu t)}{\sqrt{\sum_{i=1}^{M(t)} V_i^2}} \le x\right) &= \Phi(x) + \frac{1}{6(\mu_2')^{3/2}\sqrt{\pi t}}\{\mu_3(1+2x^2) - \mu^3(x^2-1) \\ &\quad +3\mu\left(\sigma_v^2 + x^2\sigma_v^2 + x^2\mu_2'\right)\}\phi(x) + o(t^{-1/2}) \qquad (4.17)\end{aligned}$$

Assume that we observe the process $S = \sum_{i=1}^{M(t)} V_i$ at equally spaced time points $h, 2h, 3h \cdots$, where $h > 0$. Then we have at time $t = hn$, where $n = 1, 2, \cdots$.

$$\begin{aligned}\Pr\left(\frac{(\sum_{i=1}^{M(nh)} V_i - \pi\mu nh)}{\sqrt{\sum_{i=1}^{M(nh)} V_i^2}} \le x\right) &= \Phi(x) + \frac{1}{6(\mu_2')^{3/2}\sqrt{nh\pi}}\{\mu_3(1+2x^2) - \mu^3(x^2-1) \\ &\quad +3\mu\left(\sigma_v^2 + x^2\sigma_v^2 + x^2\mu_2'\right)\}\phi(x) + o((nh)^{-1/2}) \qquad (4.18)\end{aligned}$$

Taking $h = 1$, and recalling that

$$\Upsilon_t = \frac{(\sum_{i=1}^{M(n)} V_i - \pi\mu n)}{\sqrt{\sum_{i=1}^{M(n)} V_i^2}},$$

we obtain the result. □

We next use the Edgeworth expansion of Υ_t in (4.16) to remove the skewness effect of audit data so that confidence bounds on the population mean prorated error, w, can be more accurately computed.

4.2.5 Removal of Skewness Effect of Auditing Data.

Procedures that reduce the effect of population skewness on the distribution of standardised or studentised statistics in general terms have been suggested by many authors such as Johnson (1978), Hall (1983, 1992a), and Abramovitch and Singh (1985), among others. These procedures seek to convert an asymmetric statistic into a symmetric one, then approximate by a standard normal distribution. Here we use the Edgeworth correction for skewness discussed by Abramovitch and Singh (1985) to transform our statistic Υ_t. That is

Lemma 2 (Abramovitch and Singh 1985, Theorem 1): *Suppose that a pivotal statistic $S = (\widehat{\theta}_n - \theta)/\widehat{\sigma}_\theta$ admits an Edgeworth expansion*

$$\Pr(S \le x) = \Phi(x) + n^{-1/2}\rho(F,x)\phi(x) + o(n^{-1/2})$$

uniformly in x, where $\rho(F,x)$ is a polynomial in x whose coefficients possibly depend upon the underlying distribution F. Let $\widehat{p}_n$ be an estimator of $\rho(F,S)$ which satisfies the condition that for all $\varepsilon > 0$

$$\Pr(|\widehat{p}_n - \rho(F,S)| > \varepsilon) = o(n^{-1/2}) \tag{4.19}$$

as $n \to \infty$. Then $S_1 = S + n^{-1/2}\widehat{p}_n$ has the following Edgeworth expansion

$$\Pr(S_1 \le x) = \Phi(x) + o(n^{-1/2}). \tag{4.20}$$

Remark 1 (Remark 1 of Abramovitch and Singh 1985): If $\rho(F,x)$ is a polynomial in x whose coefficients depend on F only through its first r moments, if these coefficients as a function of these moments have bounded partial derivative in a neighbourhood of the true moments of F, and $\int |X|^{\frac{3r}{2}+\delta}\, dF < \infty$ for any $\delta > 0$, then

$$\Pr(|\rho(F_n,S) - \rho(F,S)| > \varepsilon) = o(n^{-1/2})$$

for any $\varepsilon > 0$.

Our transformed statistic is formally stated in the following theorem. Suppose $m > 0$ errors are observed in a sample of size n, and let

$$\widehat{\mu} = m^{-1}\sum_{i=1}^{m} V_i, \quad \widehat{\mu}_2' = m^{-1}\sum_{i=1}^{m} V_i^2,$$

$$\widehat{\sigma}_v^2 = \frac{1}{m-1}\sum_{i=1}^{m}(V_i - \widehat{\mu})^2, \quad \widehat{\mu}_3 = m^{-1}\sum_{i=1}^{m}(V_i - \widehat{\mu})^3.$$

Then define

$$\widehat{p}_n = \frac{1}{6(\widehat{\mu}_2')^{3/2}\sqrt{m/n}}\{\widehat{\mu}_3(1+2\Upsilon_t^2) - \widehat{\mu}^3(\Upsilon_t^2 - 1) + 3\widehat{\mu}\left(\widehat{\sigma}_v^2 + \Upsilon_t^2\widehat{\sigma}_v^2 + \Upsilon_t^2\widehat{\mu}_2'\right)\}. \tag{4.21}$$

Theorem 2: *Assume that Υ_t admits an Edgeworth expansion as given in Theorem 1. Define*

$$\Upsilon_t' = \Upsilon_t + n^{-1/2}\widehat{p}_n, \tag{4.22}$$

then the distribution of Υ_t' has *an Edgeworth expansion*

$$\Pr\left(\Upsilon_t' \leq x\right) = \Phi(x) + o(n^{-1/2}) \tag{4.23}$$

uniformly in x.

We point out that equation (4.23) means that the transformed statistic, Υ_t', if approximated by the standard normal, will only be in error by $o(n^{-1/2})$.

Proof of Theorem 2: By Theorem 1, one-term Edgeworth expansion for Υ_t could be expressed as

$$\Pr\left(\Upsilon_t \leq x\right) = \Phi(x) + n^{-1/2}\rho(G,x)\phi(x) + o(n^{-1/2})$$

where

$$\rho(G,x) = \frac{1}{6(\mu_2')^{3/2}\sqrt{n\pi}}\{\mu_3(1+2x^2) - \mu^3(x^2-1) + 3\mu\left(\sigma_v^2 + x^2\sigma_v^2 + x^2\mu_2'\right)\}.$$

Using Lemma 2 and (4.16) of Theorem 1, we modify Υ_t as:

$$\Upsilon_t' = \Upsilon_t + n^{-1/2}\rho(G_n, \Upsilon_t) \tag{4.24}$$

where G_n is the empirical distribution of V. Since $\widehat{p}_n = \rho(G_n, \Upsilon_t)$, an estimator of $\rho(G, \Upsilon_t)$, by Remark 1 satisfies the condition that for all $\varepsilon > 0$

$$\Pr\left(|\rho(G_n,\Upsilon_t) - \rho(G,\Upsilon_t)| > \varepsilon\right) = o(n^{-1/2}), \tag{4.25}$$

It follows from Lemma 2 that Υ_t' has the Edgeworth expansion

$$P_r(\Upsilon_t' \leq x) = \Phi(x) + o(n^{-1/2}) \tag{4.26}$$

as claimed. □

Clearly, the transformation (4.22) reduces the effect of extreme skewness inherent in auditing population on our statistic Υ_t. As it will become clear in the next section, we can obtain an accurate asymptotic confidence bound on the population mean prorated error, w, on the basis of the transformed statistic, Υ_t'.

4.3 Confidence Bounds for the Mean Prorated Error

In this section an approximate non-parametric confidence bound on the population mean prorated error, w, with a coverage probability $(1-\alpha)$, based on Abramovitch and Singh (1985, Theorem 2) is derived. That is :

Lemma 3 (Abramovitch and Singh 1985, Theorem 2): *Suppose a statistic* $S = (\widehat{\theta}_n - \theta)/\widehat{\sigma}_\theta$ *admits an Edgeworth expansion*

$$\Pr(S \leq x) = \Phi(x) + n^{-1/2}\rho(F, x)\phi(x) + o(n^{-1/2})$$

uniformly in x, *and the conditions of remark 1 hold. Let* $S_1 = S + n^{-1/2}\rho(F_n, S)$. *Then*

$$\begin{aligned} \Pr\left(\theta < \widehat{\theta}_n - a_1\widehat{\sigma}_\theta\right) &= P\left(S_1 > z_{\alpha/2}\right) + o(n^{-1/2}) \\ &= \alpha/2 + o(n^{-1/2}) \end{aligned}$$

and

$$\begin{aligned} \Pr\left(\theta > \widehat{\theta}_n - a_2\widehat{\sigma}_\theta\right) &= P\left(S_1 < -z_{\alpha/2}\right) + o(n^{-1/2}) \\ &= \alpha/2 + o(n^{-1/2}). \end{aligned}$$

where $\alpha \in (0, 1/2)$, $-z_{\alpha/2} = \Phi^{-1}(\alpha/2)$, *so that* $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ *and*

$$\begin{aligned} a_1 &= z_{\alpha/2} - n^{-1/2}\rho(F_n, z_{\alpha/2}) \\ a_2 &= -z_{\alpha/2} - n^{-1/2}\rho(F_n, -z_{\alpha/2}) \end{aligned}$$

and F_n *denotes the empirical cumulative distribution function.*

Application of Lemma 3 to our statistic Υ_t yields the following results which are formally stated in Theorem 3.

Theorem 3: *Assume that (4.16) of Theorem 1 holds. Define the quantities*

$$\begin{aligned} \eta_1 &= z_\alpha - \frac{1}{6(\widehat{\mu}_2')^{3/2}\sqrt{m}}[\widehat{\mu}_3(1+2z_\alpha^2) - \widehat{\mu}^3(z_\alpha^2 - 1) \\ &\quad +3\widehat{\mu}\left(\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\mu}_2'\right)], \qquad (4.27) \\ \eta_2 &= -z_\alpha - \frac{1}{6(\widehat{\mu}_2')^{3/2}\sqrt{m}}[\widehat{\mu}_3(1+2z_\alpha^2) - \widehat{\mu}^3(z_\alpha^2 - 1) \\ &\quad +3\widehat{\mu}\left(\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\mu}_2'\right)]. \qquad (4.28) \end{aligned}$$

Then

$$\begin{aligned} \Pr\left(w < \widehat{w}_n - \frac{\eta_1}{n}\sqrt{\sum_{i=1}^m V_i^2}\right) &= \Pr\left(\Upsilon_t' > z_\alpha\right) + o(n^{-1/2}) \qquad (4.29) \\ &= \alpha + o(n^{-1/2}) \qquad (4.30) \end{aligned}$$

and

$$\Pr\left(w > \widehat{w}_n - \frac{\eta_2}{n}\sqrt{\sum_{i=1}^{m} V_i^2}\right) = \Pr\left(\Upsilon_t' < -z_\alpha\right) + o(n^{-1/2}) \qquad (4.31)$$

$$= \alpha + o(n^{-1/2}). \qquad (4.32)$$

where $\alpha \in (0, 1/2)$, m = *number errors observed in a sample of size* n, *and* $-z_\alpha = \Phi^{-1}(\alpha)$, *so that* $z_\alpha = \Phi^{-1}(1-\alpha)$, w *is the population mean prorated error.*

Proof of Theorem 3: By Theorem 1, one-term Edgeworth expansion for Υ_t is given by

$$\Pr(\Upsilon_t \le x) = \Phi(x) + n^{-1/2}\rho(G,x)\phi(x) + o(n^{-1/2})$$

where

$$\rho(G,x) = \frac{1}{6(\mu_2')^{3/2}\sqrt{n\pi}}\{\mu_3(1+2x^2) - \mu^3(x^2-1) + 3\mu\left(\sigma_v^2 + x^2\sigma_v^2 + x^2\mu_2'\right)\}.$$

Let G_n be empirical distribution of V. Then for any $\varepsilon > 0$

$$\Pr(|\rho(G_n,\Upsilon_t) - \rho(G,\Upsilon_t)| > \varepsilon) = o(n^{-1/2}).$$

Thus Lemma 3 applies. Let $\Upsilon_t' = \Upsilon_t + n^{-1/2}\rho(G_n,\Upsilon_t)$ and define

$$\begin{aligned}\eta_1 &= z_\alpha - n^{-1/2}\rho(G_n, z_\alpha),\\ \eta_2 &= -z_\alpha - n^{-1/2}\rho(G_n, -z_\alpha).\end{aligned}$$

Then it follows by Lemma 3 that

$$\begin{aligned}\Pr\left(w < \widehat{w}_n - \frac{\eta_1}{n}\sqrt{\sum_{i=1}^{m} V_i^2}\right) &= P\left(\Upsilon_t' > z_\alpha\right) + o(n^{-1/2})\\ &= \alpha + o(n^{-1/2})\end{aligned}$$

and

$$\begin{aligned}\Pr\left(w > \widehat{w}_n - \frac{\eta_2}{n}\sqrt{\sum_{i=1}^{m} V_i^2}\right) &= P\left(\Upsilon_t' < z_\alpha\right) + o(n^{-1/2})\\ &= \alpha + o(n^{-1/2}).\end{aligned}$$

□

This theorem says that a confidence bound for the population mean prorated error, w, constructed on the basis of transformed statistic Υ_t' has a coverage error of $o(n^{-1/2})$.

Remark: We may use a bootstrap approach to compute the empirical $100\,(1-\alpha)$ percentile of the Υ_t' instead of the normal approximation (Hall 1992). The coverage error can be reduced to $o(n^{-1})$ by bootstrapping Υ_t' (Abramovitch and Singh 1985).

Clearly, Theorem 3 can be used to construct interval for w. We next construct confidence bounds for w , when (1) two or more errors are observed and (2) exactly one error is observed in a sample of size n.

4.3.1 Treatment for Samples with Two or More Errors

We first consider the case where the number of errors observed in the sample is two or more. By Theorem 3 we have

$$\Pr\left(w > \widehat{w}_n - \frac{\eta_2}{n}\sqrt{\sum_{i=1}^{m} V_i^2}\right) = \alpha + o(n^{-1/2}) \quad \text{for } m \geq 2 \tag{4.33}$$

provided $\alpha \in (0, 1/2)$. It follows that $100(1-\alpha)\%$ upper confidence bound on w is:

$$\begin{aligned}
\widehat{w}_u(1-\alpha) \quad = \quad & \frac{1}{n}\sum_{i=1}^{m} V_i \\
& + \frac{1}{n}\{\frac{1}{6(\widehat{\mu}_2')^{3/2}\sqrt{m}}[\widehat{\mu}_3(1+2z_\alpha^2) - \widehat{\mu}^3(z_\alpha^2 - 1) \\
& + 3\widehat{\mu}\left(\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\mu}_2'\right)] + z_\alpha\}\sqrt{\sum_{i=1}^{m} V_i^2}
\end{aligned} \tag{4.34}$$

The method (4.34) provide $100(1-\alpha)\%$ upper confidence bound for the population mean taint, w. To obtain a bound for the population total error amount, T_z, multiply (4.34) by the total recorded value T_y. Thus the $(1-\alpha)$ upper confidence bound on T_z is:

$$\widehat{T}_z(1-\alpha) = T_y\widehat{w}_u(1-\alpha). \tag{4.35}$$

Similarly the $100(1-\alpha)\%$ lower confidence bound on w, $\widehat{w}_l(1-\alpha)$, is derived using Theorem 3:

$$\begin{aligned}
\widehat{w}_l(1-\alpha) \quad = \quad & \frac{1}{n}\sum_{i=1}^{m} V_i \\
& + \frac{1}{n}[\frac{1}{6(\widehat{\mu}_2')^{3/2}\sqrt{m}}\{\widehat{\mu}_3(1+2z_\alpha^2) - \widehat{\mu}^3(z_\alpha^2 - 1) \\
& + 3\widehat{\mu}\left(\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\sigma}_v^2 + z_\alpha^2\widehat{\mu}_2'\right)\} - z_\alpha]\sqrt{\sum_{i=1}^{m} V_i^2}
\end{aligned} \tag{4.36}$$

where z_α satisfies $\Phi(z_\alpha) = \alpha$. Therefore $(1-\alpha)100$ lower confidence bound on T_z is:

$$\widehat{T}_z(1-\alpha) = T_y \widehat{w}_l(1-\alpha). \tag{4.37}$$

4.3.2 Treatment for Samples with One Error

With one error, the variance is zero, then the statistics Υ'_t in (4.22), can be written as

$$\Upsilon'_t = \Upsilon_t + \frac{1}{6}(2\Upsilon_t^2 + 1).$$

and η_2 in Theorem 3 as:

$$\eta_2 = -z_\alpha - \frac{1}{6}(2z_\alpha^2 + 1),$$

Therefore we have a $(1-\alpha)$ upper confidence bound on T_z as:

$$\widehat{T}_z = T_y \frac{V_1 + [z_\alpha + \frac{1}{6}(2z_\alpha^2 + 1)]V_1}{n}. \tag{4.38}$$

The bounds (4.35) and (4.38) shall be referred to as Studentised compound Poisson (SCP) bound. So far we have seen that, for any auditing data, the use of equations (4.35) and (4.38) in computing upper confidence bounds of the error amount are theoretically justifiable.

Clearly the definition of the statistic Υ'_t in (4.22) depends on $M = m$ and if $m = 0$, where m is the number of errors observed in a sample of size n, Υ'_t is undefined.

4.4 Numerical Evaluation of the Upper Bound

Before going on to an extensive examination of the bound developed in this Chapter (SCP) in Chapter 7, we first carry out a preliminary investigation of its performance following an approach pioneered by Leitch et al. (1982). This approach involves generating, in a single sample of given size, a replica of a particular population prorated error model for a specified error rate. Various bounds are computed based on this single sample to provide representative information about the bounds behaviour. However, this does not provide sampling information about the bounds.

Typical tainting patterns were created in Leitch et al. (1982) as follows:

If k error taintings were to be created for a sample of n, corresponding to an error rate of $\pi = k/n$, the smallest error tainting was set to be equal to the $1/2k-$ quantile of the probability distribution, the next smallest error tainting was set to the $3/2k$ -quantile and so on. For example, for the χ^2 distribution with three degrees of freedom and $k = 10$

errors, the smallest error tainting would be set equal to the $100[1/2(10)] = 5th$ percentile, which is 0.0352 (scaled down). The other error taintings were created in corresponding fashion.

Four prorated error models were utilised in Leitch et al. (1982). These are the

(i) J: χ^2 distribution with one degree of freedom (χ^2_1),

(ii) J-100: χ^2 distribution with one degree of freedom with 20% of 100 percent overstatement,

(iii) Unimodal: χ^2 distribution with three degrees of freedom (χ^2_3) and

(iv) Uniform distribution from 0 to 1.

Error rates of 6, 10, 15, 20 and 25% were also utilised in the study. For the χ^2 distribution, the actual total error seeded in the population is given

$$T_z = T_y \pi [100(1 - q) + 10dq]/100,$$

where π = error rate, q =fraction of non-100 percent overstatement, and d =number of degrees of freedom of the chi-squared variate. The typical tainting patterns used in study here are given in the appendix 4.

The SCP bound in (4.35) is applied to benchmark data sets, "typical sample error tainting patterns" given in Leitch et al. (1982). The results are given in Table 4.1, in the last column.

For convenience and for purpose of comparison, the reported results for the Stringer bound (SB), Modified Multinomial (MM) from Leitch et al. (1982), Moment bound (MB) from Dworin and Grimlund (1984, Table 3), Multinomial-Dirichlet bound (MD) from Tsui, Matsumura, and Tsui (1985, Table 4) and Augmented Variance bound (AV) from Rohrbach (1993, Table 1) based on "typical tainting patterns" described in Leitch et al. (1982) are included. These are summarised in Table 4.1, columns 4 through 8.

4.4.1 Performance of the SCP Bound

The last column in Table 4.1 gives the 95% upper confidence bounds obtained by the SCP bound, using typical tainting patterns described in Leitch et al. (1982).

As can be seen from Table 4.1 the SCP bound is tighter (much closer to the true population mean taint, at a specified confidence level) than all the five bounds for the J- error model. For the J-100, the SCP bound is tighter than all except the Augmented Variance bound. It is tighter than the Stringer, Modified Multinomial, and Moment

Table 4.1: Comparisons of Upper Bounds Based on a Typical Tainting Pattern in One Sample n=100, 95% nominal confidence, Recorded Value=1 million Dollars

Error Model	Line Item Error Rate	Exact	SB	MM	MB	MD	AV	SCP
J	6%	.006	.039	.031	.025	.028	.025	.014
	10%	.010	.045	.033	.031	.033	.030	.021
	15%	.015	.052	.037	.038	.039	.036	.028
	20%	.020	.059	.043	.045	.044	.042	.035
	25%	.025	.066	.049	.051	.050	.047	.042
J-100	6%	.017	.054	.047	.042	.045	.040	.042
	10%	.028	.060	.050	.047	.049	.044	.046
	15%	.042	.080	.068	.068	.068	.063	.068
	20%	.056	.087	.075	.073	.073	.069	.072
	25%	.070	.106	.093	.092	.090	.087	.092
unimodal	6%	.018	.057	.039	.042	.043	.041	.038
	10%	.030	.074	.051	.059	.059	.055	.054
	15%	.045	.094	.072	.080	.076	.074	.075
	20%	.060	.114	.098	.098	.093	.091	.094
	25%	.075	.133	.117	.115	.110	.108	.112
Uniform	6%	.030	.076	.057	.067	.064	.059	.062
	10%	.050	.104	.083	.095	.088	.084	.088
	15%	.075	.136	.116	.126	.116	.114	.120
	20%	.100	.167	.151	.156	.144	.143	.151
	25%	.125	.198	.183	.185	.171	.171	.180

Table 4.2: Paired t test for SCP and other Upper Bounds Based on a Typical Tainting Pattern in One Sample n=100, 95% nominal confidence, Recorded Value=1 million Dollars. Ho: mean(SCP-other bounds) = mean(diff) = 0

Bounds	Average	Std. Dev.	diff(SCP- bound)	p-value
Stringer	.09005	.0423	-.01835	0.0000
Multinomial	.07465	.0412	-.00295	0.0173
Moment	.07675	.0427	-.00505	0.0000
Multinomial-Dirichlet	.07415	.0380	-.00245	0.0450
Augmented Variance	.07115	.0389	.00055	0.6614
SCP	.07170	.0437		

bound, and compares well with the Multinomial-Dirichlet bound for the unimodal and uniform distribution.

Paired t-tests to investigate whether the SCP bound is significantly tighter than the other bounds estimates, based on the typical tainting pattern in a sample of 100 were carried out. The results are given in Table 4.2.

It can be seen that SCP bound is significantly tighter than the Stringer, Modified Multinomial, Moment bound and Multinomial-Dirichlet estimates with p-value = 0.000, 0.035, 0.000 and 0.045 respectively. As shown in Table 4.2. the SCP bound is not significantly tighter than Augmented Variance bound, based on the typical tainting pattern in a sample of 100. These tests are regarded as approximate since the populations are not normally distributed. Table 4.23 gives the summary of statistics of upper bounds based on a typical tainting pattern

Large-scale simulation studies are carried out in Chapter 7 to further assess the performance of the SCP bound using real and simulated populations in terms of coverage probability, relative efficiency and relative advantage over the Stringer bound.

4.5 Summary and Discussion

In this chapter, we have developed a new method for constructing confidence bounds for the total error amount in an auditing population, when the underlying distribution is highly skewed and contains substantial number of zero errors. We have reduced the problem of estimating total error amount in auditing to compound Poisson sum and using the Edgeworth expansion for a compound Poisson distribution, we removed the effect of a extreme skewness inherent in auditing data. One interesting feature of the SCP bound is that it can handle both overstatement and understatement errors and also could be used to construct both upper and lower confidence bounds.

Using an approach pioneered by Leitch et al. 1982, we compared the performance of the new bound with other methods in the literature. Our initial results presented in this chapter suggest that the new bound compares favourably to other methods in the literature. The generalizability of the results is limited by the representativeness of the audit conditions considered here, the typical tainting patterns of Leitch et al. (1982). Further numerical investigation is carried out in Chapter 7 to investigate the reliability of the bound and also to give better indications on when our proposed estimator is appropriate.

Chapter 5

On the Stringer Bound

In this chapter, we obtain a bound on F, the common cumulative distribution function of the prorated errors, which we then use to give an alternative form of the Stringer bound. With this form of the Stringer bound, we were able to use Bolshev's recursion to obtain a lower bound on its coverage probability, and showed that, for $n \leq 2$, this lower bound is greater than or equal to the stated coverage probability. We illustrate numerically that the Stringer bound is reliable when (n, α) falls into a number of ranges; specifically $n \leq 11$ and $\alpha \in (0, 0.05)$; $n \leq 10$ and $\alpha \in (0, 0.1)$; $n \leq 9$ and $\alpha \in (0, 0.20)$; $n \leq 8$ and $\alpha \in (0, 0.40)$; and $n \leq 7$ and $\alpha \in (0, .5)$. We also proposed an extension to the Stringer method based on Rom's adjusted significance levels, and illustrated numerically the reliability of the extended Stringer bound for values of α in the range .05 to .5, and for $n = 1$ to $n = 20$.

5.1 Introduction

Writing $T_1, T_2, \cdots, T_n$ for the sample prorated errors, we assume that they are independent and identically distributed random variables with common cumulative distribution function (cdf) F. We also assume that the T_i are non negative and that $\Pr(0 \leq T_i \leq 1) = 1$. We also write $t_{i:n}$ as the i^{th} order statistics of $(T_1, T_2, \ldots, T_n)$, so that $0 \leq t_{1:n} \leq t_{2:n} \leq t_{3:n} \leq \cdots \leq t_{n:n} \leq 1$.
One of the most widely used upper confidence bounds for estimating the error in the population is the Stringer bound (Stringer, 1963), which attempts to provide an upper $1 - \alpha$ confidence bound for the mean error amount in the population:

$$\overline{\mu}_{st} = p_n(0; 1-\alpha) + \sum_{i=1}^{n} \{p_n(i; 1-\alpha) - p_n(i-1; 1-\alpha)\} t_{n-i+1:n}, \tag{5.1}$$

where $p = p_n(i; 1-\alpha)$ is a unique solution to equation

$$\sum_{k=0}^{i} \binom{n}{k} p^k (1-p)^{n-k} = \alpha, \tag{5.2}$$

for $i = 0, 1, \cdots n-1$,and $\alpha \in (0,1)$. By convention $p_n(n; 1-\alpha) = 1$. There is inequality

$$0 < p_n(0 : 1-\alpha) < ... < p_n(n-1 : 1-\alpha) < 1.$$

The formulation of the Stringer bound has never been satisfactorily explained (Panel on Nonstandard mixtures of Distributions, 1989). However there is a common belief in auditing practice that the bound works in the sense that its coverage probability is at least the nominal confidence level. That is the Stringer bound, $\overline{\mu}_{st}$, satisfies

$$\Pr[\overline{\mu}_{st} \geq \mu(F)] \geq 1-\alpha \tag{5.3}$$

for all n and for α typically less than 0.5, where F is the common cumulative distribution function of the prorated errors with mean $\mu(F)$. While, there is no general mathematical proof of (5.3) (Pap and Van Zuijlen 1996), some asymptotic results have been obtained by Bickel (1992) and Pap and Van Zuijlen (1996).

Bickel (1992) studied the asymptotic behaviour of the Stringer bound, and showed that (5.3) for $n = 1$, while

$$\Pr[\overline{\mu}_{st} \geq \mu(F)] \geq (1-\alpha)^{n+1} \qquad (n \geq 2)$$

under certain conditions on the distribution F. Bickel also showed that

$$\overline{\mu}_{st} = n^{-1} \sum_{i=1}^{n} T_i + C(F) z_{1-\alpha}/n^{1/2} + o_p(n^{-1/2}) \tag{5.4}$$

where $z_{1-\alpha}$ is $(1-\alpha)$ percentile of the standard normal distribution, and

$$C(F) = \int_0^1 F^{-1}(t) \frac{2t-1}{2\sqrt{t(1-t)}} dt, \tag{5.5}$$

and $C(F)^2 \geq \sigma^2(F) = Var(T)$ with equality only when F concentrates on at most 2 points. The $o_p(n^{-1/2})$ is a standard notation for some unspecified random variable ν_n having the property that $n^{1/2}\nu_n$ converges in probability to zero as $n \to \infty$.

Pap and Van Zuijlen (1996) extended this work to demonstrate the asymptotic conservatism of the Stringer bound. They showed that (5.4) converges almost surely and