Ph.D. Program in Economics

THESIS SUMMARY

for the Ph.D. dissertation

"Statistical methods in auditing, with special regard to financial audits"

authored by

Tamás Lolbert

Supervisor:

Dr. László Hunyadi

Professor

Budapest, 2008

Department of Mathematics, Corvinus University
and
State Audit Office of Hungary

THESIS SUMMARY

for the Ph.D. dissertation

"Statistical methods in auditing, with special regard to financial audits"

authored by

Tamás Lolbert

Supervisor:

Dr. László Hunyadi

Professor

Contents

I. Motivation

Contrary to some simpler models in economics, information is neither complete nor perfect in the real world; only a negligible minority of interpersonal or interorganizational transactions is free of information asymmetry, and almost every task-delegation is prone to the principal-agent problem.

A very large part of such problems is related to statements from third parties that serve as a basis for further decisions. If facts are misstated, decisions based on such statements may also be inadequate. Most decisions are robust for a given extent of misstatement: misstatements smaller than this extent do not influence the decision maker. This extent is called *materiality*. The users of the statements want to avoid material misstatements.

Statements mostly relate to an organization. Since the very first time of human history, people organize themselves to reach a higher utility for their community. The act of cooperation goes together with risks: factors that endanger the reaching of goals set by the members of the organization. These factors are called *inherent risks*.

Inherent risks can be mitigated in several ways. The first way, call it might and fear, is very effective but not very civilized. Another way, using the right personnel, can also work, but its use is seriously limited by the availability of such personnel. The third way was promoted in masses on the dawn of capitalism and it means forcing the organization to follow prescribed procedures. Using procedures may dehumanize the work, and in fact, machines are gradually taking over such tasks from humans. If the procedures themselves are well-designed, they are very effective in mitigating inherent risks, provided that they are followed. Failing to follow these procedures poses another risk that is called *control risk*.

Designing optimal controls is a very important part of enterprise risk management. An international organization named COSO sets the standards and best practice of this area for both public and private entities.

Inherent risks combined with control risks can make an organization issue a materially misstated statement. Its users want to be assured that neither of these particular risks arose, but mostly they are not in position to judge it. There is a need for someone who can give assurance on the validity of statements: these people are the *auditors*.

Auditing is always related to some (maybe implicit) statement or statements, and gives an assurance that these statements are free of material misstatement (or, in turn, that they are materially misstated). Depending on the nature of the audited statements, we can distinguish between four types of audit objectives:

- accounting estimation,
- attestation,
- compliance/propriety,
- value-for-money.

Most statements are in fact multiple statements, that is, during a single engagement auditors are confronted with at least two or three of the above objectives.

All audit activity aims the seizure of a reasonable assurance on validity of the statement. On the pursuit of this aim auditors can fail, so that they can accept materially misstated statements as well as reject materially not misstated statements. The first type of error is called *audit risk*.

By the widely used model of audit risk, it consists of three factors: inherent risk, control risk and *detection risk*, the latter meaning the conditional risk that an existing misstatement would not be detected by the auditor.

Procedures used in an audit are wide-ranging and among others include statistical methods. Statistical sampling is very useful since detection risk, which relates to sampling error in this case, can easily be quantified.

Using appropriate statistical techniques can make an audit much more efficient either by requiring smaller samples and less work or by increasing the level of assurance. Since auditing requires highly qualified and precious human resource, developing time-saving, parsimonious methods is very important.

The main aim of my Ph.D. thesis is to collect, unify in an integral framework and statistically-mathematically justify those procedures that can be used in relation to the audit objectives. On pursuit of this aim I will re-formulate, translate the procedures into the terminology of statistics and mathematics, and using the power of this formalism I analyze the actual procedures as well as develop new ones.

Audit procedures in this thesis are divided into three large groups: analytic procedures, attribute sampling (estimation of proportion) and variable sampling (estimation of the total value). The first group predominantly means here risk analysis and digital analysis, since other analytic procedures are either trivial or not closely related to statistics.

Analytic procedures

Assurance can be gained not only by testing but also by analyzing the internal structure of a statement and comparing parts to each other or to an external benchmark. Analytic procedures can include not only statistics, but several other methods as well. An important use of analytical procedures is risk analysis.

Risk analysis is used for the optimal allocation of audit resources. Focusing on high risk areas increases the chance that the auditor detects the material misstatement. Risk analysis in some cases requires testing of controls and the methodology of attribute sampling.

Special attention should be devoted to the interrelation between inherent and control risks since presuming their independence may result in underestimated risks. It is mostly important if the auditor considers the possibility of fraud.

There are several methods for performing a risk analysis, but in auditing the dominant method of risk analysis is defining risk factors, scoring them and weighting them for each area.

The main problem with this method is the subjectivity of weighting and the method of aggregation.

Digital analysis is a non-standard way to perform risk analysis using the frequency of certain digits in the statement to be audited. It is based on Benford's law that is an empirical phenomenon observed on a wide range of data sets. It was found that the distribution of starting digits is not uniform but biased toward the smaller digits. Similar results apply to 2nd, 3rd etc. digits or groups of digits.

The mainstream way of performing a digital analysis is mostly to check the goodness-of-fit to Benford's law applying some of the following methods:

- visual inspection
- Kolmogorov-Smirnov goodness-of-fit test
- mean absolute deviation test
- chi-square goodness-of-fit test
- z-statistic test
- regression of theoretical values on empirical values
- summation test.

In some cases the benchmark distribution is not the one implied by Benford's law but other, company specific distribution. Digital analysis is mostly used by forensic auditors to detect fraud or manipulation.

The largest problem with digital analysis is Benford's law itself since it has been only an empirical phenomenon without having a formal proof of why it should work with real-life data sets. Since the method of digital analysis is also new to most auditors, a summary of key digital analysis techniques is always welcome.

Attribute sampling

Testing of controls is a basic audit procedure, either as a part of risk analysis or as a substantive procedure in case of compliance or propriety audits. By the recently issued Sarbanes-Oxley Act, public companies are required to regularly test and report on the adequacy of their internal control system. Controls are used to mitigate risks so noncompliance with rules, failing to follow procedures means that their risk benefits decrease. There is always a tolerable frequency of rule infractions; therefore the auditor aims to decide how the real frequency relates to the tolerable. Decision is mostly made via sampling of transactions.

There are two basic approaches in this context:

- classical Neyman-Pearson estimators, and
- Bayesian estimators.

In the context of sampling method itself, we can distinguish between single-stage, multi-stage and sequential sampling plans with the last one hardly ever used in audit applications.

The main problems related to estimation of proportion are:

- substantive testing of details can be very costly and time consuming that makes parsimonious methods highly desirable
- both tolerable and real error rates tend to be very small (<5%) in most situations which makes it very difficult to find appropriate (parsimonious) statistical methods
- the actual reliability of non-exact textbook estimators is yet unknown and must be checked under low error rate / small sample scenarios

Variable sampling

In the tighter classical sense auditing means the auditing of financial statements, which is mostly related to the attestation objective. Generally the most important part of a financial statement is a list of amounts ("line items") summed up. The materiality concerning this statement is mostly related to the stated sum ("total"), it means that the difference between the stated sum and its "real value" must not exceed a predefined material amount. The aim of the auditor is to decide how the stated sum relates to the real sum, or equivalently, how the difference between the stated and real sums relates to materiality. Decision is mostly made via sampling of line items.

There are several approaches in this context:

- simple random sampling (equal probability of being selected) with classical estimators, such as mean-per-unit estimator, difference estimator or ratio estimator
- Horvitz-Thompson estimator, or other classical estimators used with sampling designs of unequal selection probability
    - classical sampling designs with unequal selection probabilities, such as stratified, cluster etc. sampling
    - sampling with probability proportional to size, monetary unit sampling (MUS).
- Special methods developed directly for the variable sampling problem
    - combined attributes and variables (CAV) approach
    - a special CAV estimator called Stringer bound
    - cell bound
    - multinomial bound (with step down S).

The main problems related to value estimation are:

- materiality rarely exceeds the 2% of stated total, individual line items are very rarely in error if at all, error distribution pattern is nonstandard and unknown

- classical estimators are inappropriate for such situations
- most special methods are heuristic and are not formally proved to be reliable
- simulations show, however, that they are in fact overly reliable resulting in excess assurance but higher risk of rejecting a materially not misstated financial statement
- besides simulations, we need a way to compare the performance of these estimators.

II. Methodology

As noted some pages ago, the main aim of my Ph.D. thesis was to collect, unify and statistically-mathematically justify those procedures that can be used in relation to the audit objectives. Since the listed problems of the researched field are wide-ranging, I applied a wide range of standard and non-standard methods.

Analytic procedures

In practical applications risk analysis is mostly performed in an ad-hoc way from a statistician's point of view. The possible risks of the usual methods for inherent and control risks are demonstrated with the help of probability theory using the simple definition of intersection of events.

The proof on the non-existence of a general Benford's law and the remarks stating that with certain restrictions most real-life distributions are close to Benford's law both use mathematical concepts, such as complex numbers, mantissa, sigma algebra, transformed random variable, characteristic function and gamma function.

Attribute sampling

The comparative analysis of existing rate estimators and the development of new the new sequential sampling method have been performed using the basic concepts of probability theory, such as hypergeometric-, binomial-, normal- and Poisson-distributions, mean, standard deviation, binomial coefficients.

In performing the comparisons, an exact method was implemented in Microsoft Excel. Sampling from finite populations, the distribution of items of certain characteristic will be hypergeometric, thus the exact probability for acceptance (the estimated interval includes the hypothesis) can be calculated with simple spreadsheet functions.

The development of the new sequential sampling method for audit purposes was both based on single-stage/multi-stage/sequential sampling theory and the development of computer software to calculate the appropriate probabilities.

Variable sampling

The analysis of variable sampling methods is a very complex topic. Though in the auditing literature the dominant way to "prove" the reliability of these special estimators is simulation, I opted to use the axiomatic method as a new approach to the problem (i.e. I applied the definitions directly). The analysis was performed both on the multinomial and Stringer bounds in order to determine their rejection regions in relation to other known rejection regions. To achieve the results some additional concepts were also used from mathematics:

- set theory: intersection, set inclusion, partition, monotone class
- calculus: continuity, contour line, level set, integral, incomplete beta function
- algebra: vectors, hyperplane
- probability theory: multinomial distribution, multinomial coefficients.

III. Results

New explanation for Benford's law

I prove that the general Benford's law on mantissae is non-existent, but if we set an upper limit on the possible numeral bases (e.g. 10, or 16), any distribution with density has a power-transform which is ”close-enough” to meet this restricted law. This is because benfordian behavior is closely related to the characteristic function, especially to its value at some dedicated points: if the characteristic function of the natural logarithm of a random variable satisfies $\varphi_{\ln\xi}\left(2\pi k/\ln b\right)=0$ for every k, ξ will obey Benford's law for base b. (see proof in Lolbert [2007])

Using this result we can give a new explanation why most sets of naturally generated numbers exhibit benfordian behavior: that's because

1. the characteristic function of most absolutely continuous distributions converges relatively fast to zero at both infinities.
2. Lévy continuity theorem ensures that pointwise convergence of characteristic functions means weak convergence (in this case to a distribution that satisfies Benford's law for base b)

Combining the two we get that a distribution with $\varphi_{\ln\xi}\left(2\pi k/\ln b\right)\approx 0$ will behave *almost like* one that obeys Benford's law.

Comparison of estimators in small sample and low error rate scenarios

I compared 6 different estimators to check their real level of confidence.

The first group of estimators (denoted by M1, M2, M3) is based on the Neyman-Pearson principle. Without a special computer program, it is very difficult to calculate exact interval, therefore most real life applications are still using approximate methods M1 and M2.

M1 is the textbook method in this context, based on binomial/normal approximation:

$$p \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{p\times(1\text{-}p)}{n}}$$

Additionally, in the case of finite populations, sometimes a correction factor is also used.

M2 is also an approximation but a more precise one:

$$\left[\lambda x+(1-\lambda)\cdot 0{,}5\right] \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{N-n}{N-1}}\sqrt{\lambda}\sqrt{\left[\lambda\frac{x(1-x)}{n}+(1-\lambda)\frac{0{,}5\cdot(1-0{,}5)}{n}\right]}$$

where $\lambda = \frac{n}{n+c^2}$ and $c = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{N-n}{N-1}}$.

Note that with $n\to\infty$ and $\frac{n}{N}\to 0$ M2 approximates M1.

M3 is the exact estimation based on the hypergeometric distribution. It can easily be calculated when using special computer programs.

The next group consists of the Bayesian estimators. Practically I used two types of prior: an informative one with Bin(N, P) and the non-informative uniform distribution.

Ultimately a mixed method, **MxB1**, practically M1 with the prior error rate used to determine the standard deviation:

$$p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P_0 \cdot (1-P_0)}{n}}$$

The performance of interval estimators in under low-error-rate scenarios has been compared using an exact method which was implemented in Microsoft Excel. The main conclusion from the comparisons is the following:

- M1 is unreliable when $\min\{m, n-m\} \geq 10$ does not hold.
- Using $n = \frac{z^2 p(1-p)}{d^2}$ for determining the sample size (based on M1) is a wrong method in most auditing situations.
- M2 outperforms M1.
- Using M3, the exact bound is a real alternative if we have a computer.
- The Bayesian bound with non informative prior performs very similar to M2.
- Using Bayesian methods is very tempting if we can ensure that our priors are good.

Sequential testing of controls in the State Audit Office

I participated in a team of professionals that developed a new procedure for the audit of municipalities. During this engagement I developed a sequential testing plan that is different from Wald's optimal sequential test. Wald's optimal sequential test is inappropriate in auditing for several reasons, that's why former efforts to implement it in auditing failed. In addition, Wald's test did not incorporate important prior information on the actual distribution of bad performers/well performers.

The main task was to determine the maximal sample size and the stopping rules for acceptance and rejection. Once the auditor reaches a stopping point and certain conditions are met, he or she can abandon the testing and draw a conclusion without the need to check the rest of transactions. Negative stopping rule (for rejection) is a constant value same as in the single stage case. Positive stopping points (for acceptance) are calculated so that the total conditional probability of forming a negative opinion on a control that in fact is a bad performer is at least the expected level of assurance. Risk of type I error is divided uniformly among the stopping points.

Collecting and analyzing empirical results of application, we concluded that this method is still highly superior to the standard single stage methods in terms of actually used sample size, while having definite advantages over Wald's optimal sequential test in terms of interpretability towards both the auditors and auditees.

It is still theoretically possible to further optimize this method but increased numerical difficulties of optimization compared to possible gains were prohibitive.

Research on MUS estimators

My research reached the following individual results:

- There is no guarantee that the multinomial bound based on step down S set (as the set of “as extreme or more extreme” samples) has the right confidence level. It is possible, however, to define the sets of “as extreme or more extreme” samples so that the resulting collection of sets forms a monotone class, which in turn ensures the right level of confidence. Any form of exact multinomial estimators is still too computer intensive which keeps them impractical.
- A new method for determining the implicit rejection regions of Stringer bound has been outlined using the 100-dimension coordinate space of cent-taintings and a continuous integral approximation of the Stringer bound. An implicit equation for the contour-line of the rejection regions has also been derived.
- There can be, in fact, no optimal estimator in a variable sampling situation since the parameter to be estimated (population total monetary value of errors) is assigned to multiple, very different hypotheses. It is impossible to define the rejection regions so that their probability is nearly constant among all relevant hypotheses.

IV. Main references

ÁSZ MÓDSZERTANI KIADVÁNYOK – LÉVAI J. (szerk.) [2004]: A számvevőszéki ellenőrzés szakmai szabályai. Állami Számvevőszék. Budapest

BENFORD, F. [1938]: The law of anomalous numbers. Proceedings of the American Philosophical Society. Vol. 78. No. 4. pp. 551–572.

BICKEL, P. J. [1992]: Inference and auditing: The Stringer bound. International Statistical Review. Vol. 60. No. 2. pp. 197–209.

CASEWARE IDEA RESEARCH DEPARTMENT [2003]: Monetary unit sampling technical specification. http://www.caseware-idea.com.

CASEWARE IDEA RESEARCH DEPARTMENT [2003]: White papers on attribute sampling technical specification. http://www.caseware-idea.com.

COCHRAN, W. G. [1977]: Sampling Techniques. 3rd edition. Wiley. New York.

DAVID, H. A. [1981]: Order statistics. Wiley. New York.

DENKINGER G. [1990]: Valószínűségszámítás. Tankönyvkiadó. Budapest.

DRAKE, P. D. – NIGRINI, M. J. [2000]: Computer assisted analytical procedures using Benford's law. Journal of Accounting Education. Vol. 18. No. 2. pp. 127-146.

FELLER, W. [1971]: An introduction to probability theory and its applications. Wiley. New York.

FIENBERG, S. E. – NETER, J. – LEITCH, R. A. [1977]: Estimating the total overstatement error in accounting populations. Journal of the American Statistical Association. Vol. 72. pp. 295–302.

GOODFELLOW, J. L. – LOEBECKE, J. K. – NETER, J. [1974]: Some perspectives on CAV sampling plans I-II. CA Magazine. October, pp. 23–30., November, pp. 46–53.

HALDENE, J. B. S. [1945]: On a method of estimating frequencies. Biometrika. Vol. 33. pp. 222–225.

HILL, T. P. [1995]: Base-invariance implies Benford's law. Proceedings of the American Mathematical Society. Vol. 123. No. 3. pp. 887–895.

HILL, T. P. [1996]: A statistical derivation of the significant-digit law. Statistical Science. Vol. 10. No. 4. pp. 354–363.

HILL, T. P.– SCHÜRGER, K. [2005]: Regularity of digits and significant digits of random variables. Journal of Stochastic Processes and their Applications. Vol. 115. No. 10. pp. 1723–1743.

HORVITZ, D. G. – THOMPSON, D. J. [1952]: A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association. Vol. 47. No. 12. pp. 663–685.

HUNYADI L. [2001]: Statisztikai következtetéselmélet közgazdászoknak. Központi Statisztikai Hivatal. Budapest.

INTERNATIONAL FEDERATION OF ACCOUNTANTS (IFAC) [2006]: Handbook of international auditing, assurance, and ethics pronouncements. http://www.ifac.org

LESLIE, D. A. – TEITLEBAUM, A. D. – ANDERSON, R. J. [1980]: Dollar-Unit Sampling-A practical guide for auditors. Pitman. London.

MEDVEGYEV P. [2002]: Valószínűségszámítás. Aula. Budapest

NETER, J. – KIM, H. S. – GRAHAM, L. E. [1984]: On combining Stringer bounds for independent monetary unit samples from several populations. Auditing. Vol. 4. No. 1 pp. 74–88.

NEWCOMB, S. [1881]: Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics. Vol. 4. No. 1. pp. 39–40.

NEYMAN, J. [1934]: On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistician Society. Vol. 97. pp. 558–606.

NIGRINI, M. J. – MITTERMAIER, L. [1997]: The use of Benford's law as an aid in analytical procedures. Auditing: A Journal of Practice and Theory. Vol. 16. No. 2. pp. 52–67.

Panel on Nonstandard Mixtures of Distributions – TAMURA, H. ET AL. [1989]: Statistical models and analysis in auditing. Statistical Science. Vol. 4. No. 1. pp. 2–33.

PAP GY. – VAN ZUIJLEN, M. C. A. [1996]: On the asymptotic behaviour of the Stringer bound. Statistica Neerlandica. Vol. 50. No. 3. pp. 367–389.

RAIMI, R. A. [1976]: The first digit problem. American Mathematical Monthly. Vol. 83. No. 7. pp. 521–538.

STRINGER, K. W. [1979]: Statistical sampling in auditing. The state of art. Annual Accounting Review. Vol. 1. pp. 113–127.

WRIGHT, T. [1991]: Exact confidence bounds when sampling from small finite universes. Springer. New York.

WRIGHT, T. [1997]: A simple algorithm for tighter exact upper confidence bounds with rare attributes in finite universes. Statistics & Probability Letters. Vol. 36. pp. 59–67.

WALD, A. [1945]: Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics, Vol. 16. No. 2. pp. 117-186.

V. Publications

LOLBERT T. [2003]: Az "objektív" kockázatelemzési módszer. Ellenőrzési Figyelő. No. 2003/4. pp. 42-45.

LOLBERT T. [2004]: A sokasági arány meghatározására irányuló statisztikai eljárások véges sokaság és kis minták esetén. Statisztikai Szemle. Vol. 82. No. 12. pp. 1053–1076.

LOLBERT T. [2006a]: A sokasági értékösszeg becslése a könyvvizsgálatban. Statisztikai Szemle. Vol. 84. No. 3. pp. 225–248.

LOLBERT T. [2006b]: Digital analysis: Theory and applications in auditing. Statisztikai Szemle. Special Number 10. pp. 148–170.

LOLBERT T. [2007]: On the non-existence of a general Benford's law. Mathematical Social Sciences. Accepted, article in press.

LOLBERT T. [2008]: Control testing with stopping points: an efficient way to comply with SOx 404. Draft working paper.