Statistical Sampling Methods for Auditing

Klaus J. Miescke  
University of Illinois at Chicago, Department of Mathematics, Statistics, & Computer Science  
851 South Morgan Street  
Chicago, IL 60607-7045, U.S.A.  
Klaus@uic.edu

1. Extrapolation based on SRS

To determine overpayments to health care providers, government agencies make use of statistical sampling methods whenever the number N of units in a universe is too large for a complete audit. A universe is determined by a provider and an audit period, and its units are recipients or services. For brevity, they are represented here by U={1,…,N}. Because of the large numbers of providers, many of them with thousands of claims per year, this is the only way agencies, with their limited resources, can control overpayments. A provider, reluctant to return a requested amount, may challenge its correctness by questioning the validity of parts of the audit, including random number selection, data documentation, and medical necessity. The purpose of this paper is to describe the basic statistical sampling method used for such audits, and to discuss pros and cons regarding validity and accuracy that are raised by statisticians working for the two sides.

The standard audit method, according to Arkin (1984), is extrapolation based on simple random sampling (SRS). Here, a simple random sample \( S \) of size \( n \) is created with a generally accepted random number generator, and the seed number is kept on records for proof. Auditors then determine the overpayment amounts \( y_i, i \in S \). Their mean \( \bar{y} \) and standard deviation \( s \) are calculated, and the total overpayment amount \( \sum_{i=1}^{N} y_i \) is estimated by \( n\bar{y} \). This is the agency’s claim for recoupment. Usually, \( n \) is sufficiently large for the sample mean to be approximately normal (central limit theorem), and the following confidence interval is presented along with the claim.

\[
\bar{y} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}
\]

where the factor \( \frac{s}{\sqrt{n}} \) is the finite population correction. The sample size \( n \) is usually determined by a precision requirement, such as having the half-width of the confidence interval not larger than a fraction \( R \) of the sample standard deviation. The latter leads to the smallest \( n \) satisfying

\[
\frac{\frac{1}{n}}{\sqrt{\frac{1}{n} - \frac{1}{N}}} z_{1-\alpha/2} \leq R
\]

A typical scenario is a universe of size \( N = 5,000 \), with a highly skewed distribution of unknown dollar overpayment amounts: 80% zeroes, 19% between 5 and 150, and 1% in the range of several hundreds. A 95% confidence level is used, i.e. \( \alpha = .05 \) and \( z_{.025} = 1.96 \). \( R = .1 \) is chosen, and (2) leads to \( n = 357 \). The result of the audit turns out to be \( N\bar{y} = 110,520 \), with the half-width of the confidence interval equal to 63,786. The agency’s statistician emphasises that this method is statistically valid, and that the recoupment amount is based on an unbiased estimator. On the other hand, the provider’s statistician questions its accuracy and claims that certain alternative methods, while also being unbiased, would have been more accurate. This will be discussed below.

2. Alternative Methods
The standard deviation of y-values in U is large due to skewness, and thus s and the width of the confidence interval in (1) are usually large as well. A reduction of the width can be achieved by choosing a smaller value for \( \alpha \) and/or \( R \), which leads to a larger \( n \). Regarding the accuracy of the normal approximation, the rule by Cochran (1977) of \( n > 25G_1^2 \), where \( G_1 \) is Fisher’s measure of skewness, leads to an even more excessive value of n. However, to achieve broad coverage of providers with audits, an agency cannot afford sample sizes far beyond of 350.

If confined to a SRS of size \( n \), one may search for an unbiased estimator that is better than \( N\bar{y} \). Utilizing the known payment amounts \( x_{i}, i \in U \), the ratio estimator would be a promising candidate if the x- and y-values were sufficiently strong correlated. Contrary to intuition, however, such a correlation has not been observed so far. Another promising candidate would be a weighted average estimator. Hidiroglou, and Srinath (1981) have studied estimators of this type for a universe U that contains \( T < n \) large y-values, but without incorporating into U a large proportion of zeroes that is typically found in audit cases. Because of the latter, it remains questionable if any of these estimators is better than \( N\bar{y} \) in an agency’s audit. Finally, the prior knowledge of about 80% zeroes in U could be utilized with a Bayes estimator, but that may still be hard to defend in court, despite the growing importance of Bayes estimation in recent years.

Instead of an SRS, stratified sampling could be utilized whenever there is a good predictor of the y-values available. It consists of values \( v_{i}, i \in U \), that are highly correlated with the y-values. Such a predictor is usually hard to identify beforehand. If one exists at all, it may not show up before the audit of an SRS has been completed. Such hindsight can not be used as an argument against the present audit, but it can help to adjust future sampling plans for audits toward some more efficient stratification. Sometimes each \( y_{i}, i \in U \), is the sum of one ore more overpayment amounts due to multiple services provided to the same recipient. The ratio-to-size estimator for an SRS of clusters does not appear to be better than \( N\bar{y} \). Switching from recipients to a universe of services, a good predictor may then be available for stratification, but that at the expense of an inflated universe size that is counterproductive with respect to precision.

To achieve an effective and accurate audit result, the strongest tool of an agency is the choice of the universe itself. Since every payment to a provider, and every fraction of it, can be subjected to examination, proper choices usually do exists before others. Cases that are very likely to have high overpayments should be audited separately, and cases that are very likely to have zero overpayments should be excluded. Obviously, cases for which no payments have been made should never be in the universe. Overpayments are not always just the result of plain errors. Sometimes they may be intended. To guard against fraud, which capitalizes on past audit experiences, agencies also have to maintain a flexible sampling strategy to keep it to some extent unpredictable.

REFERENCES


FRENCH RÉSUMÉ

La méthode-standard d'échantillonnage en cas d'un audit est extrapolation en base d'un échantillon rigoureusement aléatoire. Cette méthode est décrite en détail. Problèmes de la validité et de la précision de la méthode sont discutés. En plus, méthodes alternatives sont considérées.