

University of Groningen

Statistical Auditing and the AOQL-method

Talens, Erik

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2005

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Talens, E. (2005). *Statistical Auditing and the AOQL-method*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 2

Statistical Auditing

This chapter will give an overview of the statistical methods used in statistical auditing and the problems that auditors encounter in using these methods. A distinction is made between the methods used in compliance testing and the methods an auditor uses in a substantive test of details. This chapter will deal with methods that are well-known within auditing as well as more recently developed methods.

2.1 Compliance Testing

A compliance test aims to check whether the rules followed by the auditee (for instance an organisation, a department or a person under inspection) comply with the rules set by a certain higher authority (for instance government or management), hence the name compliance testing. Therefore, in compliance testing the auditor is interested in a certain characteristic of a population. This characteristic is uniquely defined and population items either possess this characteristic or they do not. For instance, such a characteristic can be a set of rules which have to be obeyed. Chapter 3 provides examples we used in our research at the IB-Groep, but one can also think of purchase and cost invoices that have to comply with certain rules. Sometimes it is very important that items do comply with this set of rules and no items are allowed that do not comply. The sampling method used to check whether a population contains no errors is called discovery sampling. Let us consider a population of N items. Suppose M is the number of incorrect items in this population that do not comply with a set of certain rules, then the

following hypothesis testing problem can be formulated: test the null-hypothesis

$$H_0 : M = 0$$

versus the alternative hypothesis

$$H_A : M > 0.$$

Notice that the probability of making a type I-error, i.e. incorrectly rejecting H_0 , equals zero because populations that contain no errors must result in samples that contain no errors. Let K denote the number of errors in the sample. We can use the sample size, n , to restrict the probability of making a type II-error, i.e. incorrectly not rejecting H_0 . In case all items of the population would be inspected, this probability would be zero, under the assumption that no mistakes are made while inspecting the items. But usually a certain probability of making a type II-error is accepted. Suppose $M > 0$ is the number of incorrect items in the population, then the probability of making a type II-error is given by

$$\beta = P\{K = 0 | n, M\}.$$

Dependent on the choice of the critical value M^* , we accept for all $M \geq M^*$ a probability of at most β_0 of making a type II-error. After we have chosen the values of β_0 and M^* , at least in principle we can calculate the sample size that fulfils the condition above. Here the underlying hypergeometric distribution can be used or we can use either binomial or Poisson approximations to calculate the sample size. In case of the hypergeometric distribution we can use an iterative procedure to calculate the sample size by finding the smallest value of n that satisfies the following condition

$$\frac{\binom{N-M^*}{n}}{\binom{N}{n}} \leq \beta_0.$$

Using the binomial approximation we get a sample size of

$$n \geq \frac{\log \beta_0}{\log(N - M^*) - \log N},$$

and using the Poisson approximation gives

$$n \geq -\frac{N \cdot \log \beta_0}{M^*}.$$

If any errors are found while performing the discovery sampling, then we have to assess the seriousness of the situation. We can calculate an exact upper confidence limit for the fraction p of errors in the population. Again the underlying hypergeometric distribution can be used or an approximation (binomial, Poisson or normal). Chapter 4 extensively describes how to find confidence sets for a proportion. For stratified samples the computations are more complicated. Wendell and Schmee (1996) propose a method for exact inference for proportions from a stratified finite population.

In discovery sampling we could use the sequential method that stops as soon as we find one error or we stop as soon as we have inspected n items, where n is determined in the way described above. This method will give the lowest expected sample size for all possible values of M^* and β_0 (Ghosh, 1970). In discovery sampling the value of M^* is usually small in practice, hence the expected sample size of this sequential procedure will be close to n . Therefore, in discovery sampling this sequential procedure will only be a little bit more efficient compared to the procedure with fixed sample size.

While performing discovery sampling also other (minor) errors may be found. Confidence intervals based on these errors provide insight in the number of such errors in the population. A separate check could be performed on these minor errors or other characteristics considering the following testing problem with the null-hypothesis

$$H_0 : M \leq M_l$$

versus the alternative hypothesis

$$H_A : M \geq M_u,$$

with $M_l \leq M_u$. Because minor errors occur more frequently than major errors and minor errors are probably always present it is not sensible to use the null-hypothesis of absence of errors in the population that we used in case of testing for major errors. We restrict the probability of the type I-error and the probability of the type II-error to α_0 and β_0 . Notice that here the null-hypothesis and the alternative hypothesis are not collectively exhaustive. For values of M between M_l and M_u , the so-called indifference region, we cannot make a type I-error or a type II-error but the probability of acceptance decreases for increasing values

of M and lies between β_0 and $1 - \alpha_0$. To perform this test we have to set the values of four parameters, namely M_l , M_u , α_0 and β_0 , and we have to choose n and the number of erroneous items allowed in the sample, c , in such a way that the restrictions on the probabilities of type I- and type II-error are satisfied. To give a meaningful basis for setting the values of these parameters is difficult in practice. That is why testing for minor errors is usually done in combination with the testing of major errors or by providing confidence intervals for the number of (or proportion of) minor errors in the population.

The method above can also be interpreted in terms of quality control (see among others Wetherill and Brown (1991)). The type I-error, i.e. incorrectly rejecting the null-hypothesis, is called the consumer risk and the type II-error, i.e. not rejecting the null-hypothesis if it is false, is called the producer risk. The level M_l is called the producer quality level and M_u is called the consumer quality level. For each sampling plan (n, c) we can plot the probability of not rejecting the null-hypothesis, $P\{K \leq c | n, c\}$, against the number of items in error in the population. By not rejecting the null-hypothesis, we conclude that the population is of sufficient quality. The curve we find is called the OC-curve (operating characteristic). A sampling plan (n, c) is determined by guaranteeing that the probability of accepting the population is at least equal to $1 - \alpha_0$ if M does not exceed M_l (the producer risk does not exceed α_0), the acceptance region, and by guaranteeing that the probability of accepting the population does not exceed β_0 if M equals the consumer quality level, M_u , or exceeds it (the consumer risk does not exceed β_0), the rejection region. If the number of items in error in the population lies between the producer quality level and the consumer quality level, the indifference region, then the probability of accepting the population decreases as M moves from M_l to M_u . A graphical summary of the above can be found in Figure 2.1.

The procedures considered above are often used to test for errors of a qualitative nature, but with these procedures we can also acquire insight in errors of a quantitative nature. For instance we could apply these procedures to errors that exceed a certain amount. If this error is considered to be a major error we can test for the occurrence of this error and else we could provide confidence intervals for the number of these quantitative errors in the population.

Of course also Bayesian methods can be used in compliance testing. A

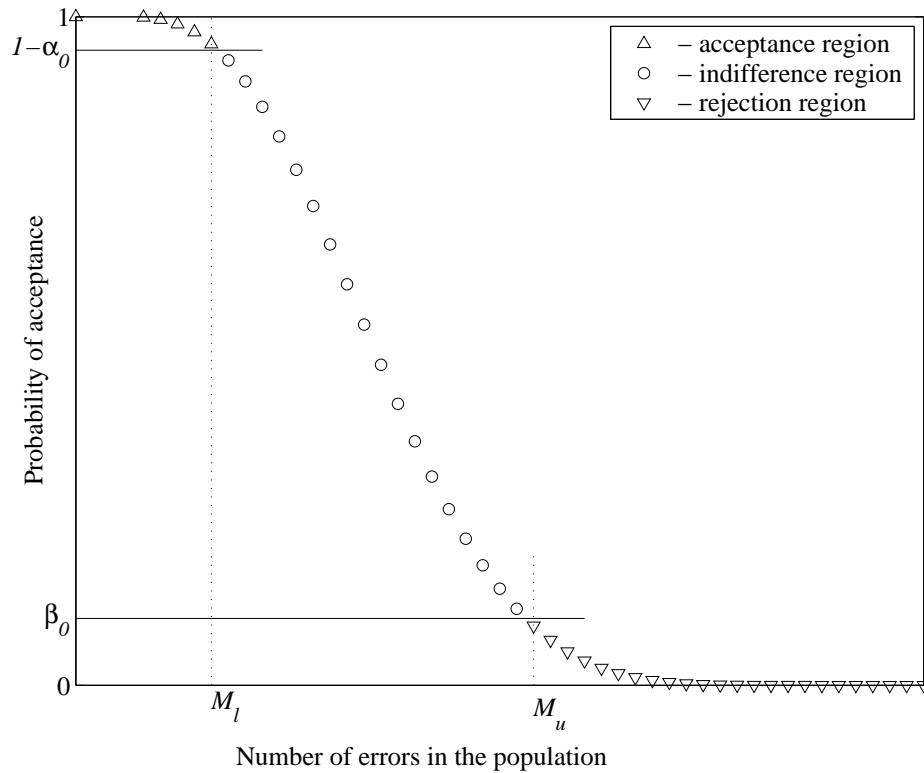


Figure 2.1. OC-curve.

comparison of classical and Bayesian interpretations of statistical compliance tests in auditing can be found in Johnstone (1997). Also Steele (1992) discusses Bayesian methods that can be used in compliance testing in auditing. Meeden (2003) uses a hierarchical Bayes model to find an upper confidence limit for the proportion of items that are in error in a stratified finite audit population.

Using the procedures above, the quality of a population can be established after the population was produced. Auditors often prefer procedures that enable them to inspect and improve the quality of the population while the population is being produced. A popular method that auditors use is the AOQL-method. This method guarantees that the population is of a certain predefined quality by taking samples and find and improve errors in the sample. If a sample contains too many errors, then the entire population is inspected and all errors in the population are corrected. Thus, after inspection of the entire population the population

contains no errors, where perfect inspection is tacitly assumed. The main part of this thesis discusses the AOQL-method. Chapter 3 discusses a practical case at the IB-Groep. Chapter 6 discusses some theoretical issues connected with the AOQL-method. How to find the optimal sample size without using a Poisson approximation for the underlying hypergeometric distribution plays an important role.

2.2 Substantive Test of Details

In a substantive test of details data is collected to find the aggregate monetary error amount. This error in the financial statements is considered to be material if knowledge of this error would affect a decision of a reasonable user of the statements. It is obvious that unambiguous quantitative criteria to find the materiality are not always available and subject to discussion. Materiality is a very important subject in auditing because it is used to decide on which methods are to be used to audit the financial statements and it is also used to evaluate potential and actual errors in the financial statements. If these errors exceed the materiality, then the auditor will ask the client to modify the financial statements.

2.2.1 Classical methods

This subsection will give a concise description of some classical methods that are used in auditing to find the aggregate monetary error; for a more detailed discussion see Cochran (1977). According to these methods items are sampled from the population. Each of the population items contains a certain number of monetary units which add up to a certain amount. This amount is called the book amount or recorded amount. The items in the sample are inspected by the auditor and the auditor establishes the number of monetary units this item should have consisted of. This amount is called the audited amount or correct amount. The difference between the book amount and the audited amount is called the error amount. The sum of the book values of all items in the population is denoted by Y and the sum of the audited values of these items is denoted by X . An auditor is interested in the total error amount found by taking the difference between the total recorded amount and the total audited amount. Therefore, this difference is

given by

$$D = Y - X. \quad (2.2.1)$$

Notice that the total book amount is known to the auditor. Finding point estimators and confidence intervals for the total audited amount will also supply us with point estimators and confidence intervals for the total error amount. The same notation in lower cases is used in denoting the audited amount, the book amount and the error amount of the k th item in the sample, namely, x_k , y_k , and d_k , respectively.

Mean-per-unit estimator

A sample of n items is taken and the auditor establishes the audited amount of these items, x_1, \dots, x_n . The sample mean of these audited amounts can be used as an estimator for the population mean. Hence, the audited amount of the population can be estimated by multiplying the number of items in the population, N , by the sample mean, \bar{x} . This gives the unbiased estimator

$$\hat{X} = N\bar{x}.$$

An unbiased estimator of the variance of \hat{X} is

$$s_{\hat{X}}^2 = \frac{N(N-n)}{n} s_x^2,$$

where s_x^2 is the sample variance of the x_i 's. A $(1 - \alpha)$ -confidence interval can be found by using the asymptotic normality of the sample distribution. This estimator is imprecise, because the variance of the audited amount can be large. Also a problem arises when the sample contains no errors. In this case all distinct samples of the same size will provide a different estimate and confidence interval. An auditor would like to see that every sample of the same size, containing no errors, gives the same results.

Ratio estimator

To improve the precision of this mean-per-unit estimator we can include an auxiliary variable that is correlated with the audit value. Here, the book value is the most logical candidate. Not only are the book amount and the audited amount of

an item highly correlated, very often they are the same. We could use the ratio of the mean of the sampled audit values and the mean of the sampled book values to estimate the total audit amount of the population. This gives the so-called ratio estimator

$$\hat{X} = \frac{\bar{x}}{\bar{y}}Y = \hat{Q}Y,$$

which is generally biased. This bias, however, becomes negligible as n becomes large. If the relation between x_i and y_i is reflected by a straight line through the origin and the variance of x_i given y_i about this line is proportional to y_i , then it can be shown that the ratio estimator is a best linear unbiased estimator. The following biased estimator can be used to estimate the variance of the ratio estimator

$$s_{\hat{X}}^2 = \frac{N(N-n)}{n}(s_x^2 - 2\hat{Q}s_{xy} + \hat{Q}^2s_y^2),$$

in which s_y^2 equals the sample variance of the y_i 's, and s_{xy} is the sample covariance. Especially in relatively small samples the biases in $s_{\hat{X}}^2$ turn out to be more serious than the biases in \hat{X} . To reduce this bias the jackknife method can be applied. Frost and Tamura (1982) showed that when error rates are not too small the jackknife method gives a better performance.

Regression estimator

Suppose an approximate linear relation between book value and audit value exists in which the line does not go through the origin. In this case we could try an estimate based on the linear regression of x_i on y_i , instead of using the ratio of the two variables. This linear regression estimate is given by

$$\hat{X} = N\bar{x} + b(Y - N\bar{y}) \quad (2.2.2)$$

in which b is the estimate of the change in the audit value when the book value is increased by unity. If we choose the value of b beforehand, then this estimator is unbiased and an unbiased estimator of its variance is given by

$$s_{\hat{X}}^2 = \frac{N(N-n)}{n}(s_x^2 - 2bs_{xy} + b^2s_y^2).$$

For $b = 1$ the regression estimator yields the so-called difference estimator. If $b = 0$, then the regression estimator gives the mean-per-unit estimator and for

$b = \bar{x}/\bar{y}$ it coincides with the ratio estimator. For values of b between zero and one (2.2.2) is a weighted average of the mean-per-unit estimator and the difference estimator, and the weight assigned to the difference estimator equals b . The variance of the regression estimator is minimized by

$$b = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^N (y_i - \bar{Y})^2},$$

in which \bar{X} , and \bar{Y} are the population means of the audit values and the book values, respectively. An effective estimator is likely to be the familiar least squares estimate of b . This is given by

$$\hat{b} = \frac{s_{xy}}{s_y^2}.$$

By substituting this estimator in (2.2.2) a biased linear regression estimator is obtained. Standard linear regression theory assumes that the population regression of x on y is linear, that the population is infinite, and that the residual variance of x given y about the regression line is constant. Here, we are not willing to make these assumptions, and by only using large sample results the following biased estimator for the variance of the regression estimator can be found

$$s_{\hat{X}}^2 = \frac{N(N-n)}{n} s_x^2 (1 - r_{xy}^2),$$

where r_{xy} is equal to the sample correlation coefficient ($r_{xy} = \frac{s_{xy}}{s_x s_y}$). Cochran (1977) suggests multiplying this by $(n-1)/(n-2)$, because the factor $n-2$ is used in standard regression theory and gives an unbiased estimate of the residual variance. When the assumptions from standard linear regression theory are made, an expression of the estimated variance of the linear regression estimator can be found in Kleijnen, Kriens, Timmermans and Van den Wildenberg (1988). They also make a comparison between different ways of computing this variance. Among this methods is also the jackknife, which reduces the bias of the variance estimator. They show that jackknifing enhances the performance, but Frost and Tamura (1982) showed for the ratio estimator that the coverage of the confidence intervals computed with these variances is not acceptable if the error percentages are too low or the error distributions are too skew.

Stratification

Stratification can be used to improve the performance of the estimators described previously. After the auditor has decided on the number of strata and has determined the stratum boundaries, the auditor allocates the sample to the different strata. Auditors will use the recorded amounts of the items in the population in the stratification procedure. These amounts have the advantage of being known and are often closely related to the audited amounts. To allocate the sample items to the different strata, the auditor often uses stratification proportional to the sum of recorded amounts in the strata or Neyman allocation is used. When using stratification the items with a large book value are more likely to be in the sample than items with a smaller book amount. Auditors prefer this, because items with large book amounts are more likely to contain larger deviations. Very often all the items in the top stratum are included in the sample. Other sampling methods, for instance cell sampling and PPS, which also favour the larger book amounts, will be discussed later on in this chapter. Stratification dramatically improves the performance of the mean-per-unit estimator, but not the performance of the confidence interval estimators that use the audited amounts as auxiliary information.

PPS estimator

To favour the items with larger book values to be included in the sample also sampling proportional to book values can be used. This means that not all items have an equal probability of being selected in the sample, but items are included with a probability proportional to their book amount. This method of sampling is also known as probability-proportional-to-size (PPS) sampling. Auditors often refer to this method as monetary-unit-sampling or dollar-unit-sampling (MUS or DUS). MUS has the advantage that items which contain more monetary units have a greater chance of selection than items which contain less monetary units. This is from an auditor's point of view a pleasant characteristic. However, items with zero book values have no chance of being selected in the sample and should be inspected separately. Also items with audited values that exceed the book values have a smaller chance of selection than preferred. Therefore, if a population contains such items, MUS may not be the correct sampling procedure.

Using MUS, the auditor samples individual monetary units instead of items. To provide an estimator for the total audited amount of the population the auditor inspects the item in which the i th monetary unit of the sample falls. The auditor uses the proportion of recorded monetary units of the item that are in error to determine by how much this sampled monetary unit is tainted. This proportion is called the taint of this monetary unit. The taint of a sampled item, t_i , is given by

$$t_i = \frac{y_i - x_i}{y_i}.$$

Using these taints the following unbiased estimator can be found for the total audited amount of the population

$$\hat{X} = \frac{Y}{n} \sum_{i=1}^n \frac{x_i}{y_i} = Y(1 - \bar{t}).$$

Its variance can be estimated by

$$s_{\hat{X}}^2 = \frac{Y(Y - n)}{n} s_t^2,$$

with s_t^2 equal to the sample variance of the taints. If the number of taints that differ from one is not too small, then the sampling distribution is approximately normal. But also this estimator performs poorly in an auditing context. Rohrbach (1993) claims that variance estimates for auxiliary variable estimators consistently underestimate the variance and lead to unreliable upper confidence bounds in audit populations. According to Rohrbach this underestimation is caused by overestimating the correlation between book and audit values. This overestimating of the correlation is caused because audit populations usually contain few errors and if an audit population does have a relatively high error rate, then most errors are of an insignificant amount. Rohrbach proposes a variance augmentation method, which in a simplified form (Swinamer, Lesperance and Will, 2004) gives the following augmented variance estimate for the total audited amount of the population

$$s_{\hat{X}}^{*2} = \frac{Y(Y - n)}{n} \left(\frac{\sum (1 - t_i)^2}{n} - (2 - \Delta/n) \left(\frac{1}{2} \left(\frac{\sum (1 - t_i)^2}{n} - s_t^2 \right) \right) \right),$$

where Δ is an adjustment factor. Empirical study of Rohrbach determined $\Delta = 2.7$ as the smallest value that consistently provided nominal coverage for a certain error distribution.

2.2.2 Non-standard mixture

The methods that are based on auxiliary information fail to provide confidence levels as planned. What causes this poor performance? In auditing populations there is a proportion of items in which the book value and audit value coincide. There is also a proportion of items p in which the book value and audit value do not coincide. The amount by which they differ can be modelled by a random variable Z . Therefore, the error d of an item can be modelled in the following way

$$d = \begin{cases} 0 & \text{with probability } 1 - p \\ Z & \text{with probability } p. \end{cases} \quad (2.2.3)$$

The distribution of the error amount is a so-called nonstandard mixture of a degenerate distribution and a continuous distribution. The poor performance of auxiliary information interval estimation is caused by this mixed nature of the audit population. The standardized statistic does not follow the t distribution for an auxiliary information estimator. The mixture causes the error distribution to be highly skewed, especially when there are few errors and errors are overstatements (the book value is larger than the audit value). This causes a poor performance of the auxiliary information interval estimator based on the asymptotic normality of the sampling distribution for the sample sizes usually used in auditing. The confidence limits for the total monetary error tend to be small using statistical techniques based on the approximate normality of the estimator of the total monetary error (Frost and Tamura, 1986). The upper confidence limit tends to be unreliable and the lower confidence limit usually is too conservative.

2.2.3 Deviating levels of confidence

In auditing practice, auditors are often more interested in obtaining lower or upper confidence limits than in obtaining two-sided confidence intervals. Independent public accountants are very often concerned in estimating the lower confidence bound for the total audited amount. An auditor wants to avoid overestimating this bound because of the potential legal liability that may follow from this. Giving a lower confidence limit for the total audited amount coincides with giving an upper confidence limit for the total error amount. Classical methods have the tendency to give an upper confidence limit which is too tight and thus

provides a level of confidence which is actually lower than the level supposed. This means that the auditor takes a greater risk than intended. Governmental agencies on the other hand are primarily concerned about the lower confidence limit of the total error amount, because they do not wish to overestimate the difference between the costs reported and what should have been reported. Classical methods tend to give a lower confidence limit which is too low. The true level of confidence is higher than it is supposed to be; the agency is assuming a lower risk than allowed by the policy. Internal auditors can be interested in providing a two-sided confidence interval, and they also would like to provide the nominal confidence level. The examples above show that research was and is still needed to provide better confidence bounds.

2.2.4 Methods using attribute sampling

Methods that take the mixed nature of audit populations into consideration have been developed in the last few decades. These methods use several approaches varying from attribute sampling theory, Bayesian inference, bootstrapping, modelling of the sampling distribution, inequalities for tail probabilities, or combinations of these methods. This subsection will give an overview of some methods that use attribute sampling.

We will begin to describe some methods that use attribute sampling. Under the assumption that the population only contains overstatements and the error amount never exceeds the book amount ($0 \leq D_i \leq Y_i$), taints will take on values between zero and one ($0 \leq T_i \leq 1$). Especially populations of accounts receivable are contaminated by overstatements. The maximum of all book values in the population is denoted by Y_{max} . An auditor is interested in giving an upper confidence bound for the total error amount of these overstatements. For this purpose a sample of n items is taken. The observed error for any item in the sample is given by (2.2.3), using this model we find for the total error amount

$$D = NpE(Z) \leq NpY_{max}.$$

This inequality holds because the error amount of an item cannot exceed the book value of this item. Suppose the number of errors in the sample is denoted

by K , then a $(1 - \alpha)$ -upper confidence limit is given by

$$D_u(K; 1 - \alpha) = Np_u(K, 1 - \alpha)Y_{max},$$

in which $p_u(K, 1 - \alpha)$ is a $(1 - \alpha)$ -upper confidence bound for p . This upper bound can be based on either the hypergeometric, binomial, Poisson or, even, normal distribution (see Chapter 4). In case of the hypergeometric distribution also the population size could be included as a parameter for the upper bound, but for the sake of notation we neglect this. Because the real outcomes of Z are not used in this upper bound for D , this bound can be too conservative and may be improved by using a stratification procedure on the book amount.

Instead of sampling the items, we could also take a sample by sampling monetary units. Stringer (1963) proposed the sampling of monetary units and the Stringer bound resulted from this, but Van Heerden (1961) was the first to propose a method of finding the total error amount using monetary unit sampling (MUS). He assumed that if a sampled monetary unit fell into the first X_i monetary units, which were considered to be correct, this sampled monetary unit was correct, and if it fell into the last D_i dollars, which were considered to be the dollars in error, it was in error.

Under the assumptions we previously made, we again use (2.2.3), but d is now considered to be the taint of a monetary unit instead of being the error amount of an item. Notice that $0 \leq Z \leq 1$, and p is the proportion of monetary units in error in the population. The total error amount is given by

$$D = YpE(Z) \leq Yp.$$

So, a $(1 - \alpha)$ -upper confidence limit is given by

$$D_u = Yp_u(K, 1 - \alpha),$$

with K equal to the number of non-zero taints in the sample. Although this bound is less conservative than the bound we found by sampling items, it still has the tendency to be too conservative, because it assumes all taints to be equal to one.

2.2.5 The cell bound

Methods that try to use the information that not all taints are equal to one, are called combined attributes and variables (CAV) estimation methods. An example

of such a bound is the cell bound, which was introduced by Leslie, Teitlebaum and Anderson (1980). This method assumes that errors are overstatements and that the sample is taken by using MUS with cell sampling. What is cell sampling? Instead of random sampling of n monetary units, auditors often divide the population in parts, cells, consisting of Y/n monetary units, and they select randomly from each cell one monetary unit. By doing so the minimal distance between two consecutive sampled monetary units is one, and the maximum distance is $(2Y/n) - 1$. Sampling items which are certain to be sampled, items of $2Y/n$ monetary units or larger, are inspected separately. This so-called cell sampling gives a more even distribution of the sampled monetary units over the population. Auditors prefer this. It also helps to find the smartest fraud, because the smartest fraud will minimize the chance of discovery by spreading the errors evenly over the population. The smartest auditor maximizes this minimal chance by spreading the sample evenly over the population. Cell sampling has the disadvantage of not being one sample of n items, but in fact when using cell sampling the auditor takes n random samples each consisting of one item. Hoeffding (1956) showed that the probability distribution of sample errors can still be approximated by a binomial distribution.

After taking a MUS sample with cell sampling, the K taints of the n sampled monetary units are ordered in a decreasing order, $1 \geq Z_{(1)} \geq \dots \geq Z_{(K)} > 0$. An upper limit for the total error amount in the population is calculated by

$$D_u = \frac{Y}{n} \text{UEL}_K,$$

in which UEL_K , the upper error limit factor for the K th taint is calculated by an iterative procedure. Starting at $\text{UEL}_0 = \lambda_u(1 - \alpha; 0)$ the upper error limit factors for the i th taint can be found with

$$\text{UEL}_i = \max \left(\text{UEL}_{i-1} + Z_{(i)}, \frac{\lambda_u(1 - \alpha; i)}{i} \sum_{j=1}^i Z_{(j)} \right),$$

in which $\lambda_u(1 - \alpha; i)$ denotes the upper limit for the Poisson distribution parameter λ when i errors are observed. See Leslie et al. (1980) for a more detailed description of this bound.

2.2.6 The Stringer bound

A CAV method frequently used is the Stringer bound, which was introduced by Stringer (1963) and elaborated by, among others, Leslie et al. (1980). This heuristic procedure, which assumes that errors are overstatements, has never been satisfactorily explained and not even an intuitive explanation can be found in the literature. The first step in constructing the Stringer bound is to order the K taints in a decreasing order, $1 \geq Z_{(1)} \geq \dots \geq Z_{(K)} > 0$, then the Stringer bound is given by

$$D_u = Yp_u(0, 1 - \alpha) + \sum_{j=1}^K Z_{(j)} \left(p_u(j, 1 - \alpha) - p_u(j - 1, 1 - \alpha) \right).$$

However, several simulation studies show that the Stringer bound is too large, i.e. the actual level of confidence exceeds the nominal confidence level. Indications of this conservatism of the Stringer bound can be found, for example, in Leitch, Neter, Plante and Sinha (1982), Plante, Neter and Leitch (1985), and Reneau (1987). Also Lucassen, Moors and Van Batenburg (1996) present a simulation study that confirms this conservatism, and they also examine several modifications of the Stringer bound based on different rankings of the taints. They examine methods in which the taints are ordered in an increasing order (ITO), a random order (RTO), and according to the corresponding error amounts (ATO) instead of ordering the taints in a decreasing order. They conclude that ITO does not give a proper confidence bound and ATO and RTO provide bounds that are less conservative than the original Stringer bound. Since ATO is a natural method of expressing the auditor's ideas about the severeness of the misstatements, they prefer the ATO method. According to the authors this method is recommended as an alternative to the Stringer bound in the Deloitte Touche Tohmatsu International manual on Audit Sampling. Bickel (1992) gives some weak fixed sample support to the conservatism of the Stringer bound and claims that the Stringer bound is asymptotically always too large. Pap and Van Zuijlen (1996) show that the Stringer bound is asymptotically conservative for confidence levels $(1 - \alpha)$, with $\alpha \in (0, 1/2]$, and asymptotically it does not have the nominal confidence level for $\alpha \in (1/2, 1)$. They also propose a modified Stringer bound which asymptotically does have the nominal confidence level.

Here we only consider overstatements, but suppose that also understatements are present. Meikle (1972) and Leslie et al. (1980) present methods that provide upper confidence limits that take understatements into account. Meikle suggests to subtract the lower confidence limit for the understatements from the upper confidence limit for the overstatements. Leslie et al. suggest to subtract the estimated mean value of the understatements from the upper confidence limit for the overstatements. Grimlund and Schroeder (1988) found the adjustment by Leslie et al. to be reliable and uniformly more efficient than the adjustment of Meikle.

2.2.7 The multinomial bound

The Stringer bound calculates an upper bound for the total monetary error which takes the magnitude of the errors into account and does not require any assumptions about the error distribution. Confidence bounds based on the multinomial distribution also have these properties but these bounds also have distributional properties which are fully known. This bound was first described by Fienberg, Neter and Leitch (1977). They assumed errors to be overstatements and each sampled dollar was placed in one of 101 unit categories corresponding to the magnitude of the error (in cents). Each taint is rounded upward to the nearest whole percent and classified according to its value in cents. If the item has no error, then the sampled dollar is placed in the category of zero cents. This gives categories of 0, 1, . . . , 99, and 100 cents, which coincides with taints of 0, 0.01, . . . , 1. More generally, there will be $r + 1$ categories labelled 0, 1, . . . , r . Let p_i denote the population proportions of the i th category, with $0 \leq p_i < 1$ and $\sum p_i = 1$, and W_i the observed number of sampled monetary units in category i . If we sample with replacement, then the distribution of $W = (W_0, \dots, W_r)$ will be multinomial with parameters (n, \mathbf{p}) , $\mathbf{p} = (p_0, \dots, p_r)$. The distribution is approximately multinomial if sampling is without replacement and the sample size is small compared to the total book value. In the original setup of Fienberg et al. the total error amount in the population is given by

$$D = \frac{Y}{100} \sum_{i=1}^{100} i p_i, \quad (2.2.4)$$

and a point estimator for D is given by

$$\hat{D} = \frac{Y}{100} \sum_{i=1}^{100} i \frac{W_i}{n}.$$

To find an upper confidence bound for D a set S is defined in the following way. It contains all possible outcomes which are at least as extreme as the sample outcome. There are many ways to define this set S , but taking computational simplicity into account, Fienberg et al. defined this so-called step down set S such that the number of errors associated with an element of S does not exceed the number of errors in the sample outcome and each error amount does not exceed any observed amount. Using S , a $(1 - \alpha)$ joint confidence set can be found for \mathbf{p} . This confidence set consists of all \mathbf{p} such that

$$\sum_S P\{W_0 = v_0, \dots, W_r = v_r\} \geq \alpha, \quad \sum v_i = n.$$

The confidence upper limit for D is acquired by maximizing (2.2.4) over this confidence set for \mathbf{p} . Although the true confidence level of this bound is unknown, Plante et al. (1985) showed that it was tighter than the Stringer and cell bounds. Plante et al. also showed that if the line items are in random order, then cell sampling does not have serious effect on the location and symmetry of the sampling distributions of the Stringer and multinomial bound, but it does tend to reduce the variability of the distributions. For the multinomial bound cell sampling tends to give improved coverage.

A lower bound for the overstatement error can be obtained by minimizing (2.2.4), under the same restrictions as above, with respect to a step-up set S which contains outcomes that are as extreme as or more extreme than the observed outcome. This set is the set of outcomes for which the total number of errors is at least as large as the observed number of errors and any individual error cannot be less than the corresponding observed individual error (Plante, Neter and Leitch, 1984).

The multinomial bound can also be used to give a lower confidence bound for understatement errors. The only difficulty is setting the maximum value of the understatement error for a monetary unit. This is far more difficult than setting the maximum value of the overstatement error for a monetary unit. But

once this maximum has been set the procedure is the same as for calculating the multinomial bound for overstatement errors (Neter, Leitch and Fienberg, 1978).

Calculating the multinomial bound is a computational complex and intensive process and soon becomes unmanageable as the number of errors increases. A solution is to make a categorisation which is less refined than the one described here above. Each tainting within such a cluster will be regarded as a tainting with the maximum possible value within this cluster. This will reduce the computational complexity considerably and still the multinomial bound will be tighter than the Stringer bound (Leitch et al., 1982). Leitch et al. propose a clustering that performs very well. Their clustering is a modification of an optimal clustering procedure developed by Fisher (1958).

2.3 Bayesian methods

Many empirical studies have been carried out to discover more about the error distributions of various accounting populations, see e.g. Johnson, Leitch and Neter (1981). Such results, in combination with other knowledge the auditor possesses, can be used by an auditor to make a prediction about the error distribution of certain audit populations. Bayesian inference provides a useful framework to incorporate this knowledge.

2.3.1 Felix and Grimlund method

We will first discuss the parametric Bayesian model of Felix and Grimlund (1977). They applied this method to item sampling, but Menzefricke and Smieliauskas (1984) applied this method also to dollar unit sampling. Suppose a MUS-sample is collected according to the model given by (2.2.3), then the density of the non-zero taints Z is assumed to be normal with mean μ_Z and variance σ_Z^2 . The precision h is defined as σ_Z^{-2} . The population proportion of monetary units p is assumed to follow a prior beta distribution with parameters $n_0 p_0$ and $n_0(1 - p_0)$. The precision follows a gamma prior distribution with parameters $\tau_0/2$ and $\tau_0 \psi_0/2$. The mean of the non-zero taints μ_Z has a normal prior distribution with mean μ_0 and variance $(hr_0)^{-1}$. The prior expected values of respectively μ_z , h , and p are equal to μ_0 , ψ_0 , and p_0 . The measure of confidence in the choices of these prior expected values is reflected by the values of r_0 , τ_0 and

n_0 . The prior joint distribution of (μ_Z, h) is a normal gamma distribution and the prior distribution of p is independent of (μ_Z, h) . The posterior distribution of (μ_Z, h) is again a normal gamma distribution and the posterior distribution of p is again a beta distribution and independent of the posterior distribution of (μ_Z, h) . The marginal distribution of μ_Z is obtained by integrating out h . This gives a Student t distribution. Because we can write the mean error amount per item μ_d as $\mu_d = p\mu_Z$, the transformation $\mu_Z = \mu_d/p$ can be substituted into the marginal distribution of μ_Z and by integrating out p the posterior distribution of μ_d is obtained. This integration does not lead to an explicit solution and has to be done numerically. Expressions for the expectation and variance can be found (Menzefricke and Smieliauskas, 1984), and by using a Student t distribution an approximate upper bound for μ_d can be found (Swinamer et al., 2004). Notice that this method can also deal with understatements. This method has the disadvantage that only certain prior parameter values can be used if a sample does not contain any errors.

2.3.2 Cox and Snell method

Cox and Snell (1979) used an exponential prior distribution for Z , with parameter $1/\mu_Z$. Since the exponential distribution can only take on positive values, this method can only deal with overstatements, but it also can handle the situation when no errors are contained in the sample. It is assumed that $1/\mu_Z$ has a gamma density with parameters b and $(b - 1)\mu_0$, where μ_0 is the prior mean of μ_Z and b specifies the variance of the prior distribution. The probability that a monetary unit is in error, p , has a gamma distribution with parameters a and p_0/a . The parameter p_0 is the prior expected value of p and a controls the variance of the prior distribution. Unlike the previous model, where the number of errors in the sample has a binomial distribution with parameters n and p , here a Poisson distribution is used with parameter np . The posterior distributions of μ_d can be shown to be a scalar transformation of an F distribution. Suppose (z_1, \dots, z_k) are the taints that are found in the sample, then a $(1 - \alpha)$ upper bound for D is calculated by

$$Y \frac{(k + a)}{(k + b)} \frac{(k\bar{z} + (b - 1)\mu_0)}{(n + a/p_0)} F_{2(k+a), 2(k+b); 1-\alpha},$$

where $\bar{z} = \sum z_i/k$ and $F_{u,v;1-\alpha}$ is the $(1 - \alpha)$ quantile of the F distribution with u and v degrees of freedom. Neter and Godfrey (1985) show that this bound is very sensitive to the choice of the prior parameters and this sensitivity does not disappear for larger sample sizes. They also show that it is possible to set these values conservatively in such a way that this bound is still reliable and that it is still significantly tighter than the Stringer bound.

2.3.3 Multinomial-Dirichlet bound

Tsui, Matsumura and Tsui (1985) described a Bayesian nonparametric method which has the same setup as the multinomial bound. Here, also only overstatements are assumed and each taint is rounded and classified according to its value in cents (0 to 100 cents). Tsui et al. assume a Dirichlet prior distribution for \mathbf{p} , namely $\text{Dir}(K\alpha_0, \dots, K\alpha_{100})$ with $K > 0$, $\alpha_i > 0$, $i = 0, \dots, 100$ and $\sum \alpha_i = 1$. Appendix 2.A gives more detailed information and some properties of the Dirichlet distribution. The α_i represent the best prediction of the auditor of the unknown p_i , and large values of K imply that the prior guesses are fairly sharp. K is usually chosen to be much smaller than the sample size to reflect that the sample information is more reliable than the best predictions of the auditor. The posterior distribution of \mathbf{p} is again a Dirichlet distribution with parameters $K'\alpha'_0, \dots, K'\alpha'_{100}$, where

$$K' = K + n \text{ and } \alpha'_i = \frac{K\alpha_i + W_i}{K'}.$$

The exact form of the posterior distribution of D can be obtained but is very complicated to work with. Appendix 2.A shows how \mathbf{p} can be simulated by simulating 101 independent gamma distributed variables, with the i th variable X_i having a gamma distribution with shape parameter $K'\alpha'_i$ and scale parameter equal to 1. By using this property the empirical distribution of D can be obtained by simulating a large number (10.000) of these \mathbf{p} 's and for each of these the total error amount D can be calculated by

$$D = \frac{Y}{100} \sum_{i=1}^{100} i p_i = \frac{Y}{100} \sum_{i=1}^{100} i \frac{X_i}{\sum_{j=0}^{100} X_j}.$$

The percentiles of this empirical distribution can be used as the true percentiles of the posterior distribution of D . As an upper bound the $100(1 - \alpha)$ percentile can be used. The associated interval is said to have credibility $(1 - \alpha)$. Notice that this does not mean that this interval automatically is a $100(1 - \alpha)$ confidence interval from a classical point of view. Tsui et al. show that if a prior Dirichlet distribution is used with $K = 5$, $\alpha_0 = 0.8$, $\alpha_1 = \alpha_2 = \dots = \alpha_{99} = 0.001$, and $\alpha_{100} = 0.101$, then the acquired bounds have good repeated sampling performance, with good frequentist properties. Tsui et al. also show that a good approximate posterior distribution of $\mu_d = D/Y$, the mean dollar taint in the population, is obtained by taking a beta distribution with mean and variance equal to the theoretical posterior distribution of μ_d . The mean of the theoretical posterior distribution of μ_d can be shown to be equal to

$$E(\mu_d) = \frac{1}{100} \sum_{i=1}^{100} i\alpha'_i,$$

and

$$\text{Var}(\mu_d) = \frac{\sum_{i=1}^{100} i^2\alpha'_i - \left(\sum_{i=1}^{100} i\alpha'_i\right)^2}{10^4(K' + 1)}.$$

This leads to an approximation of the theoretical posterior distribution of μ_d by a beta distribution $B(a, b)$ with parameters

$$a = E(\mu_d) \left(\frac{E(\mu_d)(1 - E(\mu_d))}{\text{Var}(\mu_d)} - 1 \right), \quad (2.3.1)$$

and

$$b = (1 - E(\mu_d)) \left(\frac{E(\mu_d)(1 - E(\mu_d))}{\text{Var}(\mu_d)} - 1 \right). \quad (2.3.2)$$

An upper bound for D is obtained by finding the $100(1 - \alpha)$ percentile of this beta distribution and multiplying this with the total book value Y .

Matsumura, Tsui and Wong (1990) extended the multinomial-Dirichlet model to situations in which both understatement and overstatement errors are possible. They assume that the maximum over- and understatement error in a dollar-unit is 100 percent. For overstatement errors this is by the definition of a tainting almost always the case, but for understatement errors this is somewhat different. Leslie et al. (1980) claim that in practice understatement errors rarely

exceed 100 percent. Matsumura et al. treat understatement errors that do exceed 100 percent as 100 percent errors. The model could also be modified to allow for larger errors. Here, errors are classified in one of 201 cells, namely $-100, -99, \dots, 0, \dots, 99, 100$ percent errors. The distribution of the vector $W = (W_{-100}, \dots, W_{100})$ is again assumed to be multinomial with parameters (n, \mathbf{p}) , $\mathbf{p} = (p_{-100}, \dots, p_{100})$. A Dirichlet prior distribution for \mathbf{p} is assumed, namely $\text{Dir}(K\alpha_{-100}, \dots, K\alpha_{100})$ with $K > 0$, $\alpha_i > 0$, $\sum \alpha_i = 1$, $i = -100, \dots, 100$. Matsumura et al. recommend a Dirichlet distribution as prior distribution for \mathbf{p} with $K = 1.5$, $\alpha_{-100} = \dots = \alpha_{-1} = 0.05/100$, and $\alpha_0 = \dots = \alpha_{100} = 0.95/100$. The posterior distribution of \mathbf{p} is again a Dirichlet distribution with parameters $K'\alpha'_{-100}, \dots, K'\alpha'_{100}$, where

$$K' = K + n \text{ and } \alpha'_i = \frac{K\alpha_i + W_i}{K'}.$$

The exact form of the posterior distribution of μ_d can be obtained but is very complicated to work with. Since $(1 + \mu_d)/2$ takes values between 0 and 1, we can use a beta distribution $B(a, b)$ with expectation and variance that match the expectation and variance of the true posterior distribution of $(1 + \mu_d)/2$. The values of a and b can be found by replacing $E(\mu_d)$ by $0.5(1 + E(\mu_d))$ and $\text{Var}(\mu_d)$ by $0.25\text{Var}(\mu_d)$ in equations (2.3.1) and (2.3.2). If the $100(1 - \alpha)$ percentile of this beta distribution is denoted by $B_{1-\alpha}$ then the approximate $100(1 - \alpha)$ percentile of the posterior distribution of D is given by $Y(2B_{1-\alpha} - 1)$. This approximate percentile has a credibility of $100(1 - \alpha)$ percent, but Matsumura et al. state that further research is necessary to evaluate the achieved confidence levels of this procedure.

In auditing it can also be of importance to provide a lower bound for the overstatement error. Matsumura, Plante, Tsui and Kannan (1991) studied the performance of the multinomial bound and the multinomial-Dirichlet bound. They concluded that if computational considerations are not important, then the multinomial bound should be used, otherwise the multinomial-Dirichlet method should be used with a prior Dirichlet distribution with $K = 5$, $\alpha_0 = 0.8$, $\alpha_1 = \dots = \alpha_{99} = 0.001$, and $\alpha_{100} = 0.101$.

Another nonparametric Bayesian method that uses the multinomial distribution as the data-generating model can be found in McGray (1984).

2.3.4 A distribution-free method

The Bayesian parametric approaches above specify a parametric model for the distribution function F_Z of Z for specification of the prior distribution for μ_Z . Tamura (1988) treated $F_Z(z)$ as distribution-free with Ferguson's Dirichlet process with parameter $\alpha(z)$ as the prior. Instead of using an exact form of the conditional error distribution as prior distribution, this method allows the auditor to describe the expected form of the error distribution. Let $F_{0,z}$ be the auditor's best prior prediction of the conditional distribution of the error. This prediction can vary from a standard parametric distribution to observations without any formal structure. Using the Dirichlet process with parameter $\alpha(z) = \alpha_0 F_{0,z}$ as prior for F_Z , it follows that $P\{Z \leq z\} = F_Z(z)$ has a beta distribution $B(\alpha(z), \alpha_0 - \alpha(z))$. The auditor uses the finite weight α_0 to reflect the uncertainty about his prediction. The prior expectation of $F_Z(z)$ is then given by $F_{0,Z}(z)$. Suppose k errors are found, say $v = z_1, \dots, z_k$, then given these observations the posterior distribution of $F_Z(z)$ is again a Dirichlet process with parameter $\alpha(z|v) = \alpha(z) + kF_{k,Z}(z)$, where $F_{k,Z}(z)$ is the empirical distribution function of z . The distribution of μ_d can be derived numerically from the distribution of μ_Z , which can also be derived numerically. The exact derivation can be found in Tamura (1988, pp. 4-5). An upper bound can be found by multiplying Y with the $100 \cdot (1 - \alpha)$ quantile of the distribution of μ_d . Laws and O'Hagan (2000) extend the model by splitting up the error region into a number of error categories and a Dirichlet-multinomial model is used for the rates of errors. Independent Dirichlet process models are now used for the values of taints in these error categories. Using the distribution of the book values, Monte-Carlo simulation is applied to find the distribution of the total error amount. Laws and O'Hagan (2002) adapt this model for multilocation auditing.

2.3.5 Other methods

Dworin and Grimlund (1984) approximate the sampling distribution by a gamma distribution. To estimate the parameters of this gamma distribution the method of moments is applied. They introduce a hypothetical taint which is treated as an additional observed taint. The value of this hypothetical taint depends on the nature of the audit population. Using this hypothetical taint, the bound they

calculate tends to be larger than the bound calculated without using this hypothetical taint. Dworin and Grimlund state that this conservatism compensates for the lack of information about the error distribution. This method, which can deal with both over- and understatement errors, provides a confidence level close to the stated one and is about as tight as the multinomial bound. Another parametric bound is introduced by Tamura and Frost (1986). They model the tainting distribution by the power function density and they use a parametric bootstrap to find the upper bound of the total error amount.

A solution to gain additional information about the error distribution is bootstrapping. This method was used by Talens (1998) at the Informatie Beheer Groep to find the total error amount in an audit population which was related to organising exams in the Netherlands. This method gives very tight bounds but unreliability is an issue. Another solution is a reduction of the problem to inequalities for tail probabilities of certain relevant statistics. Bentkus and Van Zuijlen (2001) present results concerning this topic. According to them, Hoeffding (1963) inequalities are currently the best available. The bounds they find can be extended to non-i.i.d. cases and to settings with several samples. This method gives very reliable bounds but also this method tends to be conservative. Inequalities that give a better approximation of the tail probabilities will lead to tighter bounds. A combination of the bootstrap and the use of Hoeffding inequalities can be found in Howard (1994). Helmers (2000) developed a method, as part of a research project with the Statistical Audit Group of PricewaterhouseCoopers in Amsterdam, that provides a new upper confidence bound for the total error amount in an audit population, where line-item sampling is appropriate. His method consists of two stages, the first stage uses an empirical Cornish-Fisher expansion, and the second stage uses the bootstrap to calibrate the coverage probability of the resulting interval estimate.

Statistical analysis in auditing very often presumes that the auditor does not make any errors while inspecting. Despite the professional skill of auditors this presumption may not always hold. In a case where only qualitative errors were the subject of interest, an item in an audit population is either correct or incorrect and no interest is being paid to the possible error amount, Raats and Moors (2004) and Wille (2003) dealt with the problem of finding estimators and upper confidence limits for the error fraction in the audit population. They did assume

that the first check was followed by a second check by an expert auditor who is infallible. By using Bayesian decision theory, but using different parameterizations, they found estimators and upper confidence limits for the error fraction in the audit population.

2.4 Conclusions

This chapter gives an overview of statistical methods used in auditing. The 1989 National Research Council's panel report on Statistical Models and Analysis in Auditing already mentioned "...the generally scattered and ad hoc nature of the existing methodology". Today this description still is very appropriate and confirmed by Swinamer et al. (2004). They compared 14 different bounds currently used in statistical auditing with each other and no bound was superior in terms of reliability and efficiency. The fact that every profession, including auditing, experiments, either motivated or not, with a variety of methods can possibly be blamed for the generally scattered and ad hoc nature of the existing methodology. This way a system with lack of structure evolves, but by feedback of experiences this system seems to suffice in most situations. Of course, it is essential to examine if the methods do not only seem to suffice, but if they really do suffice. Therefore, a good dialogue between auditor and statistician is very important. Also the 1989 National Research Council's panel report on Statistical Models and Analysis in Auditing gives an outline of statistical problems in auditing that need attention. At present many of these problems still need attention and further research. Moreover, it stays a challenge for researchers to find the method that is superior to all other methods.

2.A Dirichlet distribution

If $\mathbf{p} = (p_0, \dots, p_r)$ follows a Dirichlet distribution $\text{Dir}(K\alpha_0, \dots, K\alpha_r)$ with parameters $K\alpha_0, \dots, K\alpha_r$, with $K > 0$, $\alpha_i > 0$, $i = 0, \dots, r$ and $\sum \alpha_i = 1$, then the probability density function of \mathbf{p} has the form

$$f(\mathbf{p}) = \frac{\Gamma(K)}{\Gamma(K\alpha_0), \dots, \Gamma(K\alpha_r)} \prod_{i=0}^r p_i^{K\alpha_i-1}.$$

The distribution of p_i is a beta distribution $B(K\alpha_i, K(1 - \alpha_i))$ with

$$E(p_i) = \alpha_i \text{ and } \text{Var}(p_i) = \frac{\alpha_i(1 - \alpha_i)}{K + 1}.$$

The family of Dirichlet distributions is a class of conjugate priors for the multinomial distribution. Let $W = (W_0, \dots, W_r)$ have a multinomial distribution (n, \mathbf{p}) and the prior distribution for \mathbf{p} is Dirichlet $\text{Dir}(K\alpha_0, \dots, K\alpha_r)$, then the posterior distribution of \mathbf{p} is again a Dirichlet distribution with parameters $K\alpha_0 + W_0, \dots, K\alpha_r + W_r$. This can be rewritten as a Dirichlet distribution with parameters $K'\alpha'_0, \dots, K'\alpha'_r$, where

$$K' = K + n \text{ and } \alpha'_i = \frac{K\alpha_i + W_i}{K'}.$$

Suppose $X_i, i = 0, \dots, r$ has a gamma distribution with shape parameter $K\alpha_i$ and scale parameter equal to 1. The joint distribution of the proportions

$$p_i = \frac{X_i}{\sum_{j=0}^r X_j}, \text{ for } i = 0, \dots, r$$

is the Dirichlet distribution $\text{Dir}(K\alpha_0, \dots, K\alpha_r)$. For further results we refer to e.g. Fang and Zhang (1990).

