stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK (DEPARTMENT OF MATHEMATICAL STATISTICS)

SN 12/83 AUGUSTUS

R.D. GILL

THE SIEVE METHOD AS AN ALTERNATIVE TO DOLLAR-UNIT SAMPLING: THE MATHEMATICAL BACKGROUND

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, Kruislaan 413, Amsterdam, The Netherlands.

The Mathematical Centre, founded 11 February 1946, is a non-profit institution for the promotion of pure and applied mathematics and computer science. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics-subject classification: 62D05, 62P99.

The sieve method as an alternative to dollar-unit sampling: the mathematical background

by

R.D. Gill

ABSTRACT

This note describes the mathematical background to sieve sampling, a new method for audit sampling developed by C. Rietveld of Klynveld, Kraayenhof & Co, Accountants. This work has been done as part of a long term consultation project with KKC.

KEY WORDS & PHRASES: *Sieve sampling, cell sampling, Hoeffding inequalities.*

CONTENTS

1. TECHNIQUES FOR DOLLAR-UNIT SAMPLING AND DOLLAR-VALUE SAMPLING

1.1. Introduction

The sieve method is an alternative way of taking a dollar-unit sample in which full advantage is taken of the physical composition of an accounting population in order to make sample selection more efficient and flexible than in classical random sampling. In fact the method is so simple that it is easy to implement with a programmable pocket calculator or a microcomputer. At the same time, evaluation of a sieve sample is exactly the same as for a classical random sample, using the well known simple tables based on the Poisson distribution.

The total amount of an accounting population is generally made up of a number of sub-totals, themselves again divided and sub-divided at several levels. At the bottom level one arrives at items of various monetary values. The classical method of dollar-unit sampling completely ignores this structure. A $ 1,000,000 population is considered simply as 1,000,000 units of 1 dollar each, of which a few are bad or in error. The random sample is obtained by selecting, completely at random and independently of one another, a number of dollars from this population. These dollars are investigated and an upper error limit (a confidence upper bound for the population error rate) at the desired β-risk (one minus the confidence level) is computed from the usual Poisson tables on the basis of sample size, desired β-risk, and the number of bad dollars actually found in the sample.

We assume that this classical procedure is familiar to the reader, and that it is understood how the Poisson tables (see Appendix) are tailor made to the way the sample is chosen. We give a resumé of the background theory in the rest of this paragraph. The fact that the dollars are chosen at random and independently of one another means that the (random) number of bad dollars (errors) found in the sample has the hypergeometric sampling distribution with parameters N (population size), n (sample size) and K (number of bad dollars in the population). This sampling distribution is very close to the binomial distribution with parameters n and $p = K/N$ (error rate). In fact we would have had exactly a binomial distribution if the sample had been taken with replacement. This distribution in turn is very close to the Poisson distribution with parameter $\lambda = np$ (at least, for small values of p as are met with in practice). The tables for a standard dollar-unit sample

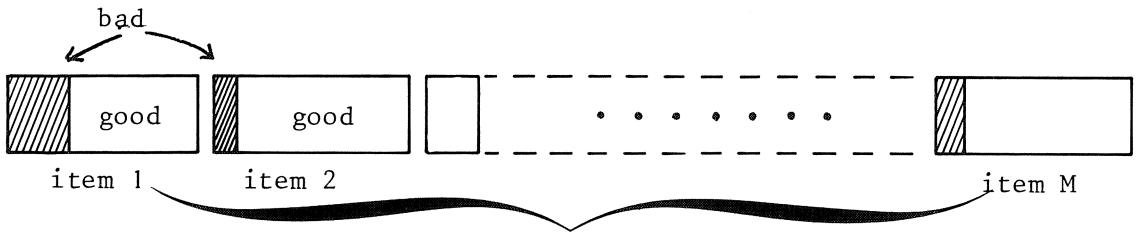
evaluation are derived from the Poisson distribution. The tables give, for various values of risk β and number of errors found in the sample k, the value of $\lambda = np$ such that the chance of finding k or less errors (in a sample of size n from a population with error rate p) equals β. Of course p is not known. However, if for a chosen value of risk β one adheres to the following behaviour rule: "if k bad dollars are found in the sample then state that np is less than the value found in the table under k errors and risk β" then the chance is at most β that one will obtain a sample with so few errors that the statement made by following the rule is untrue, whatever the actual value of p may be. Note that the upper error limit is influenced by chance through its dependence on the number of errors which chance puts in the sample, while the actual population error rate p is some unknown, fixed number. One runs a risk of β or less that the upper error limit will be lower than p, whatever p may be. For more information see any basic probability and statistics text; for auditing applications see any of the books in the reference list. Lower error limits are sometimes required too and are calculated similarly.

We shall describe sieve sampling in Subsection 1.4. However, we emphasize now that the number of errors found in a sieve sample does *not* generally have a hypergeometric, binomial, or Poisson distribution; *not even by approximation*. We shall in fact show an example in which *every* sieve sample from a certain population contains *exactly* the same number of errors; whereas for a Poisson distribution any number of errors has a chance of being found. *Yet the ordinary Poisson based evaluation is valid for sieve sampling too*. (At least, with the proviso that it may be conservative: one may run a lower risk than the chosen risk β of making an untrue statement about the true error rate. But one never runs a larger risk.) As we mentioned before, sieve sampling takes full advantage of the physical composition of an accounting population so that the selection of the random sample is much more easily made than in classical random sampling. We shall start then by describing classical random sampling in a little more detail, pinpointing its practical difficulties. Then we will describe *systematic sampling*, a popular but very dangerous way of getting a sample more easily. Then we move to *cell sampling*, a technique which superficially resembles systematic sampling quite strongly, but which does, as we shall

explain, allow the usual evaluation to be made validly. One more easy step brings us to *sieve sampling*. Rather than considering the accounting population as a collection of *dollar-units*, each of which has an *equal* chance of being selected, we now consider the population as a collection of *items*, each of which has a *different* chance (propotional to its value) of being selected. Thus the method could also be called "dollar-value sampling". We only describe the *mathematical essentials* of the method. The various ways in which it can be applied, making it an extremely versatile and efficient tool for auditing, are described in Rietveld (1978, 1979, 1984).

1.2. True random sampling

An accounting population from which samples are drawn consists of a large number of monetary units, say dollars, a few of which are bad and many good. The aim is to make a statement or to come to a decision concerning the fraction of bad dollars in the population. Generally the dollars in the population are not physically present as separate elements, but are grouped together in *items* of various sizes. These items are often present in some physical sequence. After investigation of an item one can determine that a certain number of dollars in the item is good and the rest is bad. One can use some convention as to which dollars are which; we shall say that the lowest numbered dollars in each item are the bad ones, see figure 1.



monetary units of population arranged in sequence according to sequence of items

Figure 1. Random sampling

In true random sampling one proceeds as follows. Suppose there are N dollars in the population and one wants to take a sample of size n. Then by means of random number tables, a computer or pocket calculator programmed random number generator one selects n random numbers independently of one another between 1 and N inclusive. Next one sorts these n numbers into ascending order. Then one starts adding the values of the items in the population together till the cumulative total first exceeds the lowest random

number. One knows then that the first random number falls in the last selected item and by a simple computation one can determine the rank number of the selected dollar within the item. Similarly the item number and within the item the rank number of the following n-1 selections can be determined. One proceeds to investigate the items selected. Completely good items yield zero errors, completely bad items yield one error each, and for partially bad items it depends on whether a low or a high dollar in the item has been selected. (The so-called tainting evaluation in which *each* dollar in a partially bad item is considered partially bad is described in Subsection 2.5). When the procedure has to be carried out without the use of a computer, several of the steps are rather time consuming. The sorting of n random numbers into sequence especially is rather tedious when n is of the order of several hundreds, as will often be the case. Can this be avoided?

We mention several alternatives:

(1) One can construct a random number generator which produces the numbers *in sequence*. So the first number must have the sampling distribution of the smallest of n independent random numbers; given the first, the second must have the sampling distribution of the second conditional on the value of the first, etc. This reduces to the following quite simple technique: Let $y_1, y_2, \ldots$ denote a stream of (unordered) random numbers (strictly) between zero and one. Define $s_0 = 0$. Then n *ordered* random numbers between zero and one, $x_1 < \ldots < x_n$, are defined recursively for $i = 1, \ldots, n$ by: $s_i = s_{i-1} + \frac{1}{n-i+1} \log_e(y_i)$, $x_i = 1 - \exp(s_i)$. (One can also replace $\log_e(.)$ and $\exp(.)$ by $\log_{10}(.)$ and 10 to the power (.) respectively). However many users may find this too sophisticated.

(2) *Systematic sampling*. A very common approach is to select a dollar at random from the first $\frac{N}{n}$ dollars, and then to select every $\frac{N}{n}$th dollar from then on. This is certainly very easy. But is it appropriate? Certainly, each dollar has an equal chance of being selected. But we do not have the independence between the numbers of different selected dollars which was of crucial importance right at the beginning of our chain of reasoning from the hypergeometric or binomial distribution to statistical procedures based on the Poisson.

Recall it was important that the number of bad errors found has

approximately a Poisson distribution with parameter np, where p is the population fraction of bad dollars. It is easy to conceive of populations for which the Poisson distribution is a very bad approximation. For instance, if each *item* is of size $\frac{N}{n}$ and each has the same fraction p of bad errors, then with probability p one will find a bad dollar in *every* item and with probability 1-p one will find no bad dollars at all. As a consequence, using a standard evaluation based on the Poisson distribution can lead to acceptance of a very bad population with a very large probability. Basically, the sample of n dollars gives no more information than a sample of 1 dollar.

This is a rather farfetched example but it is clear that accounting populations will often show regularities and recurring patterns of errors, which will make systematic sampling give less information about the population than true random sampling. Since this loss of information is not taken into account by a standard evaluation, one can run far higher risks than one wishes.

It is often claimed that if the items can be considered to be in random order, then systematic sampling does lead to valid evaluation. This is certainly true if each item is of size 1, so that the dollars themselves are in random order (and if probability statements refer not only to the random starting point of the procedure but also to the random order). However in general this claim has not yet been mathematically proven.

1.3. Cell sampling

By taking just a little more trouble than in systematic sampling one can still save all the time consuming sorting of random numbers without jeopardizing the statistical validiy of the evaluation. However, we will need a new mathematical result in order to justify the procedure. The idea is quite simply to select a new random number for each *cell* of $\frac{N}{n}$ dollar-units. We consider the dollar-units as forming n cells of equal size:

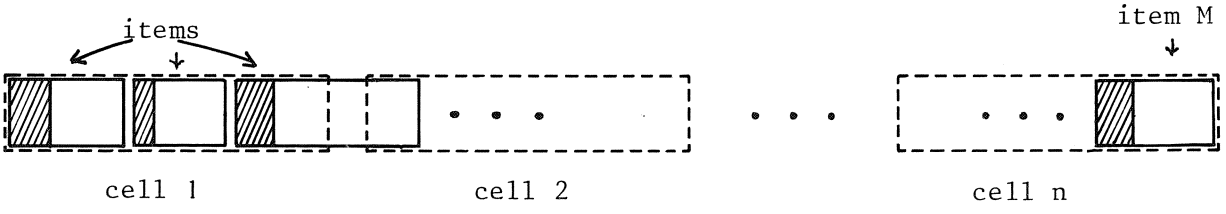


Figure 2. Cell sampling

We draw n random numbers independently from 1 to N/n inclusive; and these numbers indicate which dollar has been selected in each cell. We then

go on to make the usual evaluation.

First we should note that this is definitely *not* the same as true random sampling. Each dollar in the population has the same chance of being selected, but they are not selected independently: given one dollar has been selected, all other dollars in the same cell now have chance zero of being selected (as in systematic sampling!). Again, the sampling distribution of the number of errors found is not generally the same as in true random sampling. But whereas in systematic sampling there is an increased risk of not finding errors, in cell sampling there is a *decreased* risk. Let us consider two extreme situations. In situation 1) there is an equal number of bad dollars in each cell. A bad dollar is found in different cells independently of one another. Hence the sampling distribution of the number of bad dollars found is quite simply the binomial distribution of the number of successes in n independent Bernoulli trials, each with success chance p. This sampling distribution is exactly the same as in a true random sample with replacement from the whole population and corresponds therefore to the chosen form of evaluation. The opposite situation 2) is when all bad dollars are concentrated in one or more cells. We are now certain to find a bad dollar in each "bad" cell and none in the good ones so we have a completely different sampling distribution of number of errors found. But note: in the sample we find the same fraction of errors as in the population. For the usual confidence levels the usual (Poisson-based) computed upper error limit is larger than the observed sample fraction. (In subsection 2.4 we show that the precise bound on β is 37%. For a lower limit another bound applies; see subsection 2.5). Therefore for such a population the usual evaluation leads to a confidence interval for p which is *always* correct, hence has a risk less than the stated or nominal level β.

For populations intermediate between these two extremes one can show (Hoeffding's theorem, a proof is given in Section 3) that the sampling distribution of number of errors found is in a certain sense also intermediate between what it is in these two extreme situations. One can use this fact to go on and prove that the usual evaluation applied to a sample got by cell sampling is, if anything, *conservative;* i.e. overstates the actual β-risk. An upper error limit exceeds the true error rate p with a chance of

more than 1-β (this is proved in Section 2). This is in complete contrast to systematic sampling which if anything is anti-conservative; the upper error limit can have an appreciably larger chance than β of being lower than the true value of p.

The mathematical proof of these statements about cell sampling in intermediate populations is surprisingly hard even though the intuitive idea is so clear (we hope!). Hoeffding (1956) gives a complete description of the sampling distribution of number of errors found in intermediate populations and Anderson & Samuels (1967) apply these results to upper error limits. In Sections 2 and 3 we give precise mathematical statements and simplified proofs of these results.

1.4. Sieve method

We have now cleared up the difficulty of getting our random numbers in order, but still we have the complications of referring dollars in the population to dollars in the items. Can we orientate the whole sampling method better to the physical collection of items? The answer is yes, using a technique discovered in 1955 and further developed in the seventies by C. Rietveld of Klynveld Kraayenhof & Co. (see Rietveld 1978, 1979, 1984).

This technique has many possibilities for extension and elaboration which make it a very powerful and versatile method, but here we just concentrate on the bare bones. Suppose an item is of total book value a. The chance that a dollar is selected in this item (in cell sampling) is precisely $a/(N/n) = na/N$, the number of dollars in the item divided by the number of dollars in the cell. (We suppose for the moment that $a \leq N/n$ and that the item falls completely in one cell.) One could achieve the same probability of selection if one drew a random number from 1 to N/n for this specific item alone, and selected the item if the random number, say X, is less than or equal to a. Looking for bad dollars in a selected item, we would add 1 to our total of errors found if the Xth dollar is one of the bad ones. We can do this for each item *independently*. One could say that we are considering cells each of size $\frac{N}{n}$, one for each item, and generally largely consisting of imaginary dollars. We draw a random dollar from each cell, and, if it is not imaginary, see if it is a good or a bad dollar. (Alternatively a tainting evaluation could be used; see subsection 2.5).

An alternative way of visualizing this process is as follows. We discard completely the idea that the items consist of individual dollars. Rather, we imagine the item in its entirety as being laid on a sieve with random mesh

size, uniformly and continously distributed between zero and N/n. If the item is of size a, then the chance that it remains lying on the sieve is a/(N/n). In a second round we imagine just the error in the item as lying on the same sieve, so if the error amount is e then the chance it remains and is hence discovered is e/(N/n). This sieve-idea turns out to be extremely fruitful since it naturally leads to various important extensions and modifications of the method.

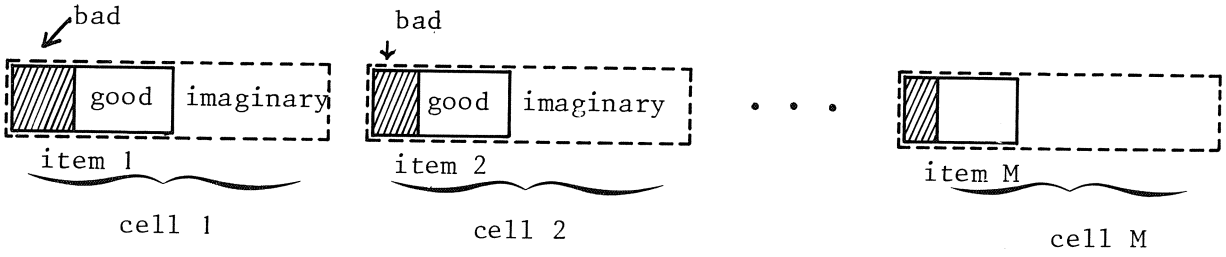


Figure 3. Sieve sampling

The question is now: what can we say about the sampling distribution of the number of errors found? Suppose the population consists of M > n items each of size less than or equal to N/n: population size divided by "sample size" (the procedure must be modified if larger items are present. These are generally subjected to complete investigation, or split into smaller items). Again we consider two extreme situations. (1) The bad dollars are evenly distributed through the population. If the fraction of bad dollars is p, there are pN bad dollars in total and $\frac{pN}{M}$ bad dollars per item. For each of the items, the chance a bad dollar is found equals $\frac{pN}{M} / \frac{N}{n} = \frac{pn}{M}$. There are M items. So the number of bad dollars found is binomially distributed with parameters M and $\frac{pn}{M}$; so nearly Poisson distributed with parameter $M \times \frac{pn}{M} = pn$. We get the same sampling distribution of number of errors found as in the "true random sampling case", for which our evaluation is tailor made. So the evaluation is correct.

In the other extreme (2) the bad dollars are concentrated in a few items. Supposing again a fraction p of bad dollars, the most extreme case is that pN bad dollars are contained in $\frac{pN}{N/n} = pn$ completely bad items of size $\frac{N}{n}$ each. We find then precisely pn bad dollars in our sample. Since our evaluation is based on a sample size n, the fraction of bad dollars with respect to this sample size is p. At the usual confidence levels (precisely, at confidence levels larger then 63%) our computed upper error limit is with probability one larger than p.

Thus we have exactly the same extremes as in cell sampling. Since math-

ematically the sieve method *is* a form of cell sampling exactly the same result holds: intermediate situations lead to intermediate (in a certain sense) sampling distributions of the number of errors found, and the upper error limit *based on the usual Poisson method and sample size n* has probability β *or less* of being less than the true value of p (at least, for β less than 37%. For lower limits a slightly different rule applies, see 2.5).

2. APPLICATION OF HOEFFDING'S THEOREM TO THE STATISTICAL EVALUATION OF A SIEVE SAMPLE OR CELL SAMPLE

2.1. Introduction

In this section we give a precise statement of part of Hoeffding's theorem in the context of the sieve method and show how it can be applied to confidence intervals and hypothesis tests. As we said in Section 1, the validity of the method depends on the fact that concentration of errors in a small number of large items leads to a large chance of discovery of errors. Conversely, if the errors are evenly spread throughout the population, one will not discover many errors. Thus the "evenly spread" case is the most dangerous one; but in this case the usual Poisson based evaluation of ordinary dollar unit sampling is appropriate. In other cases with increased chance of error detection the usual evaluation is conservative. We will only consider the part of Hoeffding's theorem which is relevant to upper limits. For lower limits and two-sided limits the other part is needed, which is discussed briefly in subsection 2.5.

2.2. Statement of Hoeffding's theorem.

We use the same notation as in Section 1, supplemented by some further symbols.

M	Number of items
a_i	Size of item i, i = 1, ..., M
e_i	Amount of error in item i, i = 1, ..., M
$N = \sum_{i=1}^{M} a_i$	Population size
$K = \sum_{i=1}^{M} e_i$	Total error amount
$p = K/N$	Error rate

n Nominal sample size

$C = N/n$ Effective cell size

$p_i = e_i/C$ Probability of error detection in item i, i = 1, ..., M

$\lambda = \sum_{i=1}^{M} p_i = K/C = np$ Expected number of errors found

To avoid unnecessary complications, we suppose that all items are smaller than the effective cell size (larger items are set apart and evaluated separately) $0 \le e_i \le a_i \le C$ and $0 \le p_i \le 1$. In the sieve method, independently of one another, in each item i

- with probability $e_i/C = p_i$ a bad dollar (error) is found
- with probability $(a_i - e_i)/C$ a good dollar is found
- with probability $(C - a_i)/C$ the item is discarded.

Let $\underline{x}_i$ be the random variable (random quantities are underlined) which takes the value 1 if an error is found in item i, zero otherwise. So $\underline{x}_1, \ldots, \underline{x}_M$ are independent Bernoulli variables with $\Pr[\underline{x}_i = 1] = 1 - \Pr[\underline{x}_i = 0] = p_i$. Let $\underline{x} = \sum_{i=1}^{M} \underline{x}_i$ denote the total number of errors found.

Hoeffding's theorem states that the sampling distribution of $\underline{x}$ lies, in a certain sense, between the distributions it has in the two extreme cases

(1) evenly spread errors: $e_1 = \ldots = e_M$

(2) maximal concentration of errors: all nonzero e_i's (except possibly one) equal to the maximal value C.

Let $[\lambda]$ and $\{\lambda\}$ denote the whole part (entier) and fractional part of λ respectively; so $[\lambda]$ is a whole number, $0 \le \{\lambda\} < 1$, and $\lambda = [\lambda] + \{\lambda\}$. Then if we keep the total error K and total number of items M fixed, the two extreme cases are

(1) $$p_1 = \ldots = p_M = \lambda/M$$

(2) $$p_1 = \ldots = p_{[\lambda]} = 1;\ p_{[\lambda]+1} = \{\lambda\};\ p_{[\lambda]+2} = \ldots = p_M = 0.$$

Let x be a whole number, $0 \le x \le \lambda - 1$ or equivalently $0 \le x < [\lambda]$. Write $\Pr[\underline{x} \le x \mid p_1, \ldots, p_M]$ for the probability that the random number of errors found $\underline{x}$ takes on a value less than or equal to the fixed number

x when $p_1,\ldots,p_M$ are the probabilities of finding an error in each item. We have in case (1)

$$\Pr[\underline{x} \le x \mid \lambda/M,\ldots,\lambda/M] = \sum_{y=0}^{x} \binom{M}{y} \left(\frac{\lambda}{M}\right)^y \left(1-\frac{\lambda}{M}\right)^{M-y}$$

$$\simeq \sum_{y=0}^{x} \lambda^y e^{-\lambda}/y! \ .$$

In case (2) we have

$$\Pr[\underline{x} \le x \mid 1,\ldots,1,\{\lambda\},0,\ldots,0] = 0$$

since we will find $[\lambda]$ errors with probability $1-\{\lambda\}$ and $[\lambda]+1$ errors with probability $\{\lambda\}$, but never x or less errors since $x < [\lambda]$. Hoeffding's theorem states exactly that, for any $p_1,\ldots,p_M$ with $\sum_{i=1}^{M} p_i = \lambda$ and any $x < [\lambda]$,

$$0 \le \Pr[\underline{x} \le x \mid p_1,\ldots,p_M] \le \sum_{y=0}^{x} \binom{M}{y} \left(\frac{\lambda}{M}\right)^y \left(1-\frac{\lambda}{M}\right)^{M-y}$$

$$\simeq \sum_{y=0}^{x} \lambda^y e^{-\lambda}/y! \ .$$

The two "≤" signs are in fact "<" unless $p_1,\ldots,p_M$ correspond to case (1) or case (2). Moreover one can use Hoeffding's theorem itself to show that the last "$\simeq$" can also be replaced by "<" (see subsection 2.5). We prove the theorem in Section 3. But now we show how it applies to the testing problem and the confidence interval problem. Except in subsection 2.5, we only consider one-sided upper testing and estimation. In each case β will denote the risk ($1-\beta$ is the confidence level). In view of the condition $x < [\lambda]$ it turns out that β cannot be arbitrarily chosen; one must require $\beta \le e^{-1} = 0.3679$. However this is no restriction in practice.

2.3. The testing problem

Suppose we wish to test the null hypothesis $\lambda \ge \lambda_0$ against the alternative $\lambda < \lambda_0$ at confidence level $1-\beta$ for some given λ_0. Since $\lambda = K/C$ and C, the effective cell size, is known, this corresponds to the null hypothesis

that the total error K exceeds the amount $C\lambda_0$ (materiality). Using the standard Poisson evaluation means that we will reject the null hypothesis (and accept the population) if $\underline{x}$ takes on some value less than or equal to x_0, where x_0 is determined by

$$\sum_{y=0}^{x_0} \lambda_0^y \, e^{-\lambda_0}/y! = \beta.$$

We must show that if in fact $\lambda \geq \lambda_0$, then the chance of rejecting the null hypothesis is less than or equal to β, whatever the actual values of $p_1,..,p_M$. Now note that for the usual values of β we always have $x_0 \leq \lambda_0 - 1$ (see the standard Poisson table in the appendix). The largest admissable value of β is $\beta = e^{-1} = 0.3679$, corresponding to the case $x_0 = 0$, $\lambda_0 = 1$. Suppose indeed that $\lambda = \Sigma_{i=1}^{M} p_i \geq \lambda_0$. Then $x_0 \leq \lambda_0 - 1 \leq \lambda - 1$ and hence

$$\begin{aligned} \Pr[\underline{x} \leq x_0 | p_1,\ldots,p_M] &< \sum_{y=0}^{x_0} \lambda^y \, e^{-\lambda}/y! \quad \text{(Hoeffding)} \\ &\leq \sum_{y=0}^{x_0} \lambda_0^y \, e^{-\lambda_0}/y! \quad \text{(because } \lambda \geq \lambda_0\text{)} \\ &= \beta \qquad\qquad\qquad \text{(by definition of } x_0\text{)} \end{aligned}$$

and the required result has been proved. At the second step we used the well known fact that if $\underline{y}$ has the Poisson (λ) distribution, then $\Pr[\underline{y} \leq x_0]$ decreases as λ increases.

2.4. The estimation problem

In the usual evaluation, if we observe $\underline{x} = x$ then the upper confidence limit $\underline{\lambda}_u$ for λ takes the value $\underline{\lambda}_u = \lambda_u$ which satisfies

$$\sum_{y=0}^{x} (\lambda_u)^y \, e^{-\lambda_u}/y! = \beta$$

where $1-\beta$ is the chosen confidence level. We then make the statement "$\lambda < \lambda_u$". We must prove that

$$\Pr[\underline{\lambda}_u \leq \lambda | p_1,\ldots,p_M] \leq \beta;$$

i.e. the chance of making a false statement about $\lambda = \Sigma_{i=1}^{M} p_i$ is less than or equal to β, whatever the values of $p_1,\ldots,p_M$. (Again, an upper bound for λ is equivalent to an upper bound for the total error K since the effective cell size C is known). Now we have: $\underline{\lambda}_u \leq \lambda$ if and only if $\underline{x}$ takes on a value less than or equal to x_0, where x_0 satisfies

$$\sum_{y=0}^{x_0} \lambda^y e^{-\lambda}/y! \leq \beta \quad \text{but} \quad \sum_{y=0}^{x_0+1} \lambda^y e^{-\lambda}/y! > \beta.$$

(If we observe $\underline{x} = x_0$ or less then the true value λ gives a very small chance of finding such a small number of errors, and our upper limit λ_u is set lower than λ). Notice again that for the usual values of β ($\beta \leq e^{-1}$) we have $x_0 \leq \lambda - 1$ (see the Poisson table in the appendix). So

$$\begin{aligned} \Pr[\underline{\lambda}_u \leq \lambda \mid p_1,\ldots,p_M] &= \Pr[\underline{x} \leq x_0 \mid p_1,\ldots p_M] \\ &< \sum_{y=0}^{x_0} \lambda^y e^{-\lambda}/y! \qquad \text{(Hoeffding)} \\ &\leq \beta \qquad \text{(by definition of } x_0). \end{aligned}$$

Again the required result has been proved.

2.5. Extensions of the sieve method; lower and two-sided limits; tainting; conservatism of the Binomial-Poisson approximation

We have only treated the simplest version of the sieve method here. Variants of the method, such as level-wise subselection and adjustable sample size using sieve boundaries (see Rietveld, 1978, 1979, 1984) and its application when some items are larger than C and when the population size N is unknown prior to sampling, reduce mathematically to the simple case treated here. Note too that cell-sampling corresponds, mathematically, to sieve sampling with M = n; i.e. each cell of size C is considered as a separate item.

So far we have only considered upper error limits. For lower and two-sided limits one needs the second part of Hoeffding's theorem, which concerns x larger than λ. Referring back to subsection 2.2 note that if $x \geq \lambda$ (so $x \geq [\lambda]$ if $\{\lambda\} = 0$, $x \geq [\lambda] + 1$ if $\{\lambda\} > 0$) then in case (1)

$$\Pr[\underline{x} \le x \mid \lambda/M, \ldots, \lambda/M] = \sum_{y=0}^{x} \binom{M}{y} \left(\frac{\lambda}{M}\right)^y \left(1-\frac{\lambda}{M}\right)^{M-y} \simeq \sum_{y=0}^{x} \lambda^y e^{-\lambda}/y!$$

while in case (2) we now have

$$\Pr[\underline{x} \le x \mid 1, \ldots, 1, \{\lambda\}, 0, \ldots, 0] = 1.$$

The second part of Hoeffding's theorem states precisely that for $x \ge \lambda$, and for any $p_1, \ldots, p_M$ with $\sum_{i=1}^{M} p_i = \lambda$, $\Pr[\underline{x} \le x \mid p_1, \ldots, p_M]$ lies betwen these two extremes. Since the event $\underline{x} \ge x$ is the complement of the event $\underline{x} \le x - 1$ one can state the theorem also as: for any $x \ge \lambda + 1$ and any $p_1, \ldots, p_M$ with $\sum_{i=1}^{M} p_i = \lambda$,

$$0 \le \Pr[\underline{x} \ge x \mid p_1, \ldots, p_M] \le \sum_{y=x}^{M} \binom{M}{y} \left(\frac{\lambda}{M}\right)^y \left(1-\frac{\lambda}{M}\right)^{M-y}$$

$$\simeq 1 - \sum_{y=0}^{x-1} \lambda^y e^{-\lambda}/y! \; .$$

In fact the last " $\simeq$ " can also be replaced by " $<$ " . The inequalities are also trivially true in the case $x = 0$. In the Poisson table (lower limits) the condition $x \ge \lambda + 1$ or $x = 0$ is always satisfied for the usual confidence levels, *except* that in the row $x = 1$ it is *always* violated. For in this row $\lambda = -\log(1-\beta) > 0$, hence $x < \lambda + 1$. As the risk β increases, the condition $x \ge \lambda + 1$ or $x = 0$ is first violated in the row $x = 2$ at $\lambda = 1$, $\beta = (1-e^{-1}-1e^{-1}) = 1 - 2 \times 0.3679 = 0.2642$. This leads to the following rules ensuring that confidence bounds are valid:

(1) For a confidence upper limit the Poisson evaluation is conservative for all risks $\beta \le e^{-1} = 0.3679$.

(2) For a confidence lower limit the Poisson evaluation is conservative for all risks $\beta \le 1 - 2e^{-1} = 0.2642$ *provided* the lower limit is taken as zero when one error is observed.

(3) For two-sided, equal-tails limits the Poisson evaluation is conservative for all risks $\beta \le 2 \times 0.2642 = 0.5284$ under the same proviso for $x = 1$ as in (2).

Similar rules can be made up for one and two sided testing, but the easiest rule is: check the condition $x_0 \le \lambda_0 - 1$ or $x_0 \ge \lambda_0 + 1$ (whichever is appropriate).

We have not discussed the common "tainting evaluation" of dollar unit sampling (see e.g. Leslie, Teitlebaum & Anderson, 1980) which could of course also be applied to sieve sampling. Given that an item i has been selected, rather than going on to select an individual dollar (the "my dollar right or wrong" approach) one would instead calculate the degree of tainting of the item $t_i = e_i/a_i$ and then apply the standard method to the non-zero taintings found. If $\underline{t}_{(1)} \geq \ldots \geq \underline{t}_{(\underline{x})}$ denote the ordered non-zero taintings found, and $\lambda_u(x)$ denotes the usual $1-\beta$ confidence upper limit to λ when x errors are found, then the tainting evaluation is to state

$$"\lambda < \lambda_u(0) + \sum_{y=1}^{\underline{x}} \underline{t}_y(\lambda_u(y) - \lambda_u(y-1))".$$

Unfortunately mathematically very little is known about tainting evaluation, whether for true random sampling, cell-sampling or sieve-sampling. Much empirical evidence exists strongly suggesting that the evaluation is conservative (see e.g. Fienberg, Neter & Leitch (1977) or Cox & Snell (1979)) but no proof has as yet been found.

Finally we show how Hoeffding's theorem implies that the usual Poisson approximation to the binomial distribution is conservative. We need to prove the following result: suppose $\underline{x}^{(M)}$ is binomially $(M,\lambda/M)$ distributed and $\underline{y}$ is Poisson (λ) distributed. Then for $x \leq \lambda - 1$, $\Pr[\underline{x}^{(M)} \leq x] < \Pr[\underline{y} < x]$. We prove this by noting that according to Hoeffding's theorem, for any $M' < M$, x and λ such that $x \leq \lambda - 1$ and $\lambda/M' < 1$

$$\Pr[\underline{x} \leq x \mid \lambda/M', \ldots, \lambda/M', 0, \ldots, 0] < \Pr[\underline{x} \leq x \mid \lambda/M, \ldots, \lambda/M].$$

Thus $\Pr[\underline{x}^{(M')} \leq x] < \Pr[\underline{x}^{(M)} \leq x]$. This shows that the sequence $\Pr[\underline{x}^{(M)} \leq x]$, $M = [\lambda] + 1, [\lambda] + 2, \ldots$ is strictly increasing in M. We know that the sequence approaches $\Pr[\underline{y} \leq x]$ as M tends to infinity. Therefore we must have $\Pr[\underline{x}^{(M)} \leq x] < \Pr[\underline{y} \leq x]$ for all M.

3. SIMPLIFIED PROOF OF HOEFFDING'S THEOREM

Write $\tilde{p} = (p_1, \ldots, p_M)$ for the vector of M components whose i'th component is p_i and define

$$f_x^M(\tilde{p}) = f_x^M(p_1,\ldots,p_M) = \Pr[\underline{x} \leq x \mid p_1,\ldots,p_M].$$

Here $\underline{x}$ is the sum of M independent Bernoulli variables $\underline{x}_i$ with $\Pr[\underline{x}_i = 1] = 1 - \Pr[\underline{x}_i = 0] = p_i$. Throughout this section we consider a fixed value of M, a fixed value of $\lambda = \Sigma_{i=1}^M p_i$ and a fixed value of $x \leq \lambda - 1$. We wish to show that for any $\tilde{p}$ with $\Sigma_{i=1}^M p_i = \lambda$,

$$0 = f_x^M(\underbrace{1,\ldots,1}_{[\lambda]\text{ components}}, \{\lambda\}, 0,\ldots,0) \leq f_x^M(\tilde{p}) \leq f_x^M(\lambda/M,\ldots,\lambda/M)$$

Since the first equality and inequality are trivially true, we need only prove the last inequality. Equivalently we must show that $f_x^M(\tilde{p})$ is maximized over all $\tilde{p}$ with $\Sigma_{i=1}^M p_i = \lambda$ by $\tilde{p} = (\lambda/M,\ldots,\lambda/M)$. We shall do this in two steps: first we will show that at the maximum value of $f_x^M(\tilde{p})$, each p_i equals zero, one, or one other value (at least, we show that $\tilde{p}$ can be so chosen to give the maximum value). After that we use the condition $x \leq \lambda - 1$ to show that the values zero and one can be excluded.

One more piece of notation is needed: write $\tilde{p}_{ij}$ for the vector obtained from $\tilde{p}$ by deleting the i'th and j'th components; i.e. when $i < j$

$$\tilde{p}_{ij} = (p_1,\ldots, p_{i-1}, p_{i+1},\ldots,p_{j-1}, p_{j+1},\ldots,p_M)$$

and similarly when $i > j$.

If we distinguish two items i and j we can split up the event "x or less errors found" according to whether or not errors are found in items i and j (with probabilities p_i and p_j) and correspondingly finding x - 2 errors or less, x - 1 errors or less, or x errors or less, in the remaining M - 2 items. This gives us

$$\begin{aligned} f_x^M(\tilde{p}) &= p_i p_j\, f_{x-2}^{M-2}(\tilde{p}_{ij}) + p_i(1-p_j) f_{x-1}^{M-2}(\tilde{p}_{ij}) + (1-p_i)p_j\, f_{x-1}^{M-2}(\tilde{p}_{ij}) \\ &\quad + (1-p_i)(1-p_j) f_x^{M-2}(\tilde{p}_{ij}) \\ &= p_i p_j\,(f_{x-2}^{M-2}(\tilde{p}_{ij}) - 2\, f_{x-1}^{M-2}(\tilde{p}_{ij}) + f_x^{M-2}(\tilde{p}_{ij})) \\ &\quad + (p_i+p_j)(f_{x-1}^{M-2}(\tilde{p}_{ij}) - f_x^{M-2}(\tilde{p}_{ij})) + f_x^{M-2}(\tilde{p}_{ij}) \end{aligned}$$

$$= p_i p_j A + B.$$

(Of course A and B both depend on the choice of i and j). Here

$$A = f_{x-2}^{M-2}(\tilde{p}_{ij}) - 2f_{x-1}^{M-2}(\tilde{p}_{ij}) + f_x^{M-2}(\tilde{p}_{ij})$$

$$= (f_x^{M-2}(\tilde{p}_{ij}) - f_{x-1}^{M-2}(\tilde{p}_{ij})) - (f_{x-1}^{M-2}(\tilde{p}_{ij}) - f_{x-2}^{M-2}(\tilde{p}_{ij}))$$

$$= g_x^{M-2}(\tilde{p}_{ij}) - g_{x-1}^{M-2}(\tilde{p}_{ij})$$

if we define $g_y^{M-2}(\tilde{p}_{ij})$ as the probability of finding *exactly* y errors in the remaining M - 2 items, since $f_x^{M-2}(\tilde{p}_{ij}) = \Sigma_{y=0}^{x} g_y^{M-2}(\tilde{p}_{ij})$.

Note that if $\tilde{p}_{ij}$ is held fixed, and we vary p_i and p_j but keep their sum fixed, then in the equation

$$f_x^M(\tilde{p}) = p_i p_j A + B$$

only $p_i p_j$ varies; A and B remain fixed. Suppose we can replace p_i by $p_i + \varepsilon$ and p_j by $p_j - \varepsilon$ for some small quantity ε (if p_i or p_j equals zero or one, then this may not be possible at all or only possible for positive or negative ε). Then the only change in $f_x^M(\tilde{p})$ is that $p_i p_j$ is replaced by $p_i p_j + \varepsilon(p_j - p_i) - \varepsilon^2$. Suppose to begin with that $p_i < p_j$. For ε close enough to zero, ε^2 is much smaller than $\varepsilon(p_j - p_i)$. So if A is positive, taking ε small and positive makes $f_x^M(\tilde{p})$ increase; if A is negative taking ε small and negative makes $f_x^M(\tilde{p})$ increase; if A is zero then $f_x^M(\tilde{p})$ does not change at all.

Suppose $\tilde{p}$ maximizes $f_x^M(\tilde{p})$ (at least one maximizing value does exist, but there may be several values of $\tilde{p}$ yielding the same maximum). If any two components of $\tilde{p}$ exist not equal to one another and not equal to zero or one, say $0 < p_i < p_j < 1$, then the corresponding term A must be zero or we could increase $f_x^M(\tilde{p})$ still further. Since the term A is zero we can replace both p_i and p_j by their average $\frac{1}{2}(p_i + p_j)$ without changing the value of $f_x^M(\tilde{p})$. If we repeat this procedure infinitely often, at each step choosing for i and j the items with (currently) the smallest nonzero p_i and the largest nonone p_j, then $f_x^M(\tilde{p})$ stays at the same (maximum) value while all nonzero and nonone components of $\tilde{p}$ get closer and closer to one another. In

the limit all these components are equal and $f_x^M(\tilde{p})$ still has the same maximum value.

This completes the first step of the proof: we now know that the maximum value of $f_x^M(\tilde{p})$ is attained by a $\tilde{p}$ whose nonzero and nonone components are all equal. In fact Hoeffding shows that no other values of $\tilde{p}$ maximize $f_x^M(\tilde{p})$, but that does not concern us.

Now we go on to the second step. Suppose that $\tilde{p}$ maximizes $f_x^M(\tilde{p})$ and that $\tilde{p}$ has r components equal to zero, s components equal to one, and therefore M - r - s components equal to $a = (\lambda-s)/(M-r-s)$ where $0 < a < 1$ ($\tilde{p}$ must satisfy $\Sigma_{i=1}^M p_i = \lambda$). We show first that we must have $M - r-s > 0$. For suppose that $M - r - s = 0$. Then it must be true that $\lambda = [\lambda] = s$. For $x \leq \lambda - 1$ this gives $f_x^M(\tilde{p}) = 0$, which cannot be the maximum value. So $M - r - s = 0$ leads to a contradiction.

Recall that $f_x^M(\tilde{p}) = p_i p_j A + B$ where $A = g_x^{M-2}(\tilde{p}_{ij}) - g_{x-1}^{M-2}(\tilde{p}_{ij})$. Recall also that at a maximizing value $\tilde{p}$, if $p_i < p_j$ then $A \leq 0$ since otherwise $f_x^M(\tilde{p})$ can be increased by moving p_i and p_j towards one another a little.

For the time being suppose that $x < \lambda-1$ (we treat the case $x = \lambda-1$ later). We are now going to compute the exact value of A for two choices of p_i and p_j, to show that $r > 0$ and $s > 0$ each leads to a contradiction. Suppose $\tilde{p}_{ij}$ contains u zeros, v ones and $M - 2 - u - v$ a's. The numbers u and v depend on r, s and the choice of p_i and p_j. Since $g_x^{M-2}(\tilde{p}_{ij})$ is the chance of finding x errors in the remaining M - 2 items, it equals the chance of finding x-v errors in the M-2-u-v items corresponding to the value a. So we have for $x > v$

$$A = g_x^{M-2}(\tilde{p}_{ij}) - g_{x-1}^{M-2}(\tilde{p}_{ij}) = \binom{M-u-v-2}{x-v} a^{x-v}(1-a)^{M-u-x-2} - \binom{M-u-v-2}{x-v-1} a^{x-v-1}(1-a)^{M-u-x-1}$$

$$= \left(a(M-u-x-1) - (1-a)(x-v)\right)\left(\frac{(M-u-v-2)!}{(x-v)!(M-u-x-1)!} a^{x-v-1}(1-a)^{M-u-x-2}\right).$$

Note that the second term in large brackets here is positive.

For x = v we have

$$A = g_x^{M-2}(\tilde{p}_{ij}) = (1-a)^{M-2-u-x} > 0$$

and for $x < v$ we have $A = 0$.

(i) If $r > 0$ we can choose i and j with $p_i = 0$ and $p_j = a$. So we can take $u = r - 1$ and $v = s$. Because $\tilde{p}$ gives a maximum of $f_x^M(\tilde{p})$ we must have $A \leq 0$. If $x = v$ this is immediately a contradiction. If $x < v = s$ then $f_x^M(\tilde{p}) = 0$, which is again a contradiction (strictly positive values of $f_x^M(\tilde{p})$ are possible). If $x > v$ then $A \leq 0$ implies that

$$a(M-u-x-1) - (1-a)(x-v) \leq 0$$

$$\therefore a(M-r+1-x-1) - (1-a)(x-s) \leq 0$$

$$\therefore a(M-r-x+x-s) \leq x - s$$

$$\therefore a \leq \frac{x-s}{M-r-s} < \frac{\lambda-s}{M-r-s} = a \qquad \text{(since } x < \lambda\text{)},$$

a contradiction. So $r > 0$ is impossible.

(ii) If $s > 0$ we can choose i and j with $p_i = a$ and $p_j = 1$. We now have $u = r$ and $v = s - 1$. Again we must have $A \leq 0$. If $x \leq v = s-1$ then $f_x^M(\tilde{p}) = 0$, which is immediately a contradiction. If $x > v$ then $A \leq 0$ implies that

$$a(M-u-x-1) - (1-a)(x-v) \leq 0$$

$$\therefore a(M-r-x-1) - (1-a)(x-s+1) \leq 0$$

$$\therefore a(M-r-x-1+x-s+1) \leq x - s + 1$$

$$\therefore a \leq \frac{x-s+1}{M-r-s} < \frac{\lambda-s}{M-r-s} = a \qquad \text{(since } x < \lambda - 1\text{)},$$

a contradiction. So $s > 0$ is impossible.

This shows that when $x < \lambda - 1$ we must have $r = s = 0$. Thus for $x < \lambda - 1$, $f_x^M(\tilde{p})$ is maximized by $\tilde{p} = (\lambda/M,\ldots,\lambda/M)$. By letting λ approach arbitrarily close to x+1 from above, the same is also true in the limit when $x = \lambda - 1$; and this concludes the proof of the theorem.

The second part of Hoeffding's theorem, namely that for $x \geq \lambda$ and for any $\tilde{p}$ with $\Sigma_{i=1}^M p_i = \lambda$ we have $f_x^M(\tilde{p}) \geq f_x^M(\lambda/M,\ldots,\lambda/M)$, can be proved by reversing the role of $\underline{x}$ and $M - \underline{x}$. In more detail, note that

$$f_x^M(\tilde{p}) = Pr[\underline{x} \le x \mid p_1,\ldots,p_M] = 1 - Pr[\underline{x} \ge x+1 \mid p_1,\ldots,p_M]$$

$$= 1 - Pr[M - \underline{x} \le M - x - 1 \mid p_1,\ldots,p_M]$$

$$= 1 - Pr[\underline{x} \le M - x - 1 \mid 1-p_1,\ldots,1-p_M]$$

$$\ge 1 - Pr[\underline{x} \le M - x - 1 \mid 1 - \lambda/M,\ldots,1 - \lambda/M]$$

$$\text{if } M - x - 1 \le \sum_{i=1}^{M} (1-p_i) - 1.$$

This condition is equivalent to $x \ge \Sigma_{i=1}^{M} p_i$. Running back through the chain of equalities with p_i replaced by λ/M gives the required result.

4. APPENDIX: POISSON TABLES

Upper limits for $\lambda = np$

Products of sample size n and error rate p when:

- x errors are found
- β equals the accepted risk of making an incorrect statement

x \ β	0.001	0.01	0.05	0.37 (*)	0.50
0	6.91	4.60	3.00	1.00	0.69
1	9.23	6.64	4.74	2.15	1.68
2	11.23	8.41	6.30	3.26	2.67
3	13.06	10.05	7.75	4.35	3.67
4	14.79	11.60	9.15	5.43	4.67
5	16.45	13.11	10.51	6.51	5.67
6	18.06	14.57	11.84	7.58	6.67
7	19.63	16.00	13.15	8.64	7.67
8	21.16	17.40	14.43	9.70	8.67
9	22.66	18.78	15.71	10.75	9.67
10	24.13	20.14	16.96	11.81	10.67
11	25.59	21.49	18.21	12.86	11.67
12	27.03	22.82	19.44	13.90	12.67
13	28.45	24.14	20.67	14.95	13.67
14	29.85	25.45	21.89	16.00	14.67
15	31.24	26.74	23.10	17.04	15.67
16	32.62	28.03	24.30	18.08	16.67
17	33.99	29.31	25.50	19.12	17.67
18	35.35	30.58	26.69	20.16	18.67
19	36.80	31.85	27.88	21.20	19.67
20	38.04	33.10	29.06	22.24	20.67

The value of λ given in the table is the solution to the equation

$$\sum_{y=0}^{x} \lambda^y e^{-\lambda}/y! = \beta.$$

Boxed values violate the requirement $x \leq \lambda - 1$, or satisfy $x = \lambda - 1$.

For two-sided limits add the risks corresponding to lower and upper limits.

(*) $e^{-1} = 0.3679$.

Lower limits for $\lambda = np$.

Products of sample size n and error rate p when:

- x errors are found
- β equals the accepted risk of making an incorrect statement

x \ β	0.001	0.01	0.05	0.26 (*)	0.50
0	0.00	0.00	0.00	0.00	0.00
1	[0.00]	[0.01]	[0.05]	[0.31]	0.69
2	0.05	0.15	0.36	[1.00]	1.68
3	0.19	0.44	0.82	1.78	2.67
4	0.43	0.82	1.37	2.60	3.67
5	0.74	1.28	1.97	3.45	4.67
6	1.11	1.79	2.61	4.31	5.67
7	1.52	2.33	3.29	5.18	6.67
8	1.97	2.91	3.98	6.06	7.67
9	2.45	3.51	4.70	6.95	8.67
10	2.96	4.13	5.43	7.84	9.67
11	3.49	4.77	6.17	8.74	10.67
12	4.04	5.43	6.92	9.65	11.67
13	4.61	6.10	7.69	10.56	12.67
14	5.20	6.78	8.46	11.47	13.67
15	5.79	7.48	9.25	12.39	14.67
16	6.41	8.18	10.04	13.31	15.67
17	7.03	8.89	10.83	14.23	16.67
18	7.66	9.62	11.63	15.15	17.67
19	8.31	10.35	12.44	16.08	18.67
20	8.96	11.08	13.25	17.01	[19.67]

The value of λ given in the table is the solution to the equation

$$1 - \sum_{y=0}^{x-1} \lambda^y e^{-\lambda}/y! = \beta \qquad (x = 1,2,\ldots);$$

λ is exactly zero in the row $x = 0$.

Boxed values violate the requirement $\lambda + 1 \leq x$ or $x = 0$, or satisfy $\lambda + 1 = x$.

For two-sided limits add the risks corresponding to lower and upper limits.

(*) $1 - 2e^{-1} = 0.2642$.

REFERENCES

[1] ANDERSON, T.W. & S.M. SAMUELS (1967), *Some inequalities among binomial and Poisson probabilities*, Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability 1 1-12.

[2] ARENS, A.A. & J.K. LOEBBECKE (1981), *Application of statistical sampling to auditing*, Prentice Hall.

[3] COX, D.R. & E.J. SNELL (1979), *On sampling and the estimation of rare errors*, Biometrika 66 123-132.

[4] FIENBERG, S.E., J. NETER & R.A. LEITCH (1977), *Estimating the total overstatement error in accounting populations*, Journal of the American Statistical Association 72 295-302.

[5] GUY, D.M. (1981), *An introduction to statistical sampling in auditing*, John Wiley & Sons.

[6] HOEFFDING, W. (1956), *On the distribution of the number of successes in independent trials*, Annals of Mathematical Statistics 27 713-721.

[7] KRIENS, J. & A.C. DEKKERS (1979), *Steekproeven in de accountantscontrole*, Stenfert Kroese, Leiden.

[8] LESLIE, D.A., A.D. TEITLEBAUM & R.J. ANDERSON (1980), *Dollar-unit sampling*, Pitman Publishing Ltd. London.

[9] NIvRA Geschrift no. 25 (1982), *Accountantscontrole en steekproef*, Kluwer, Deventer.

[10] RIETVELD, C. (1978, 1979), *The sieve method as selection method for statistical samples in audit practice*, parts I to IV (in Dutch), Compact 15, 16 and 17, Klynveld Kraayenhof & Co., Amsterdam.

[11] RIETVELD, C. (1984), *The sieve method as selection method for statistical samples in audic practice* (in Dutch), Maandblad voor Accountancy en Bedrijfshuishoudkunde 54.

[12] ROBERTS, D.M. (1978), *Statistical auditing*, AICPA 1978, New York.