Publ. Math. Debrecen
57 / 1-2 (2000), 163–183

Modified Stringer bounds

By G. PAP (Debrecen) and M. C. A. VAN ZUIJLEN (Nijmegen)

Abstract. The asymptotic theory of some alternatives proposed by Lucassen, Moors and Van Batenburg for the Stringer bound has been developed. It indicates that these alternatives cannot compete with the Stringer bound. Moreover, it is proved that the so called "all or nothing method" is superiour.

1. Introduction

Monetary unit sampling is a widely used technique in auditing accounts (Van Batenburg and Kriens [1], Tamura [13]), which makes it possible to obtain an upper bound for the total misstatement percentage in an accounting population. One of these upper bounds is the Stringer bound, which is a linear combination of the ordered taintings. These taintings are defined as the relative misstatement in the sample items, i.e., the quotient of the difference between the book value and the audited value, and the book value. Simulation studies (Burdick and Reneau [3], Leitch et al. [5], Plante et al. [10], Reneau [11]) have indicated that the Stringer bound is conservative, which means that the actual confidence level achieved by the Stringer bound exceeds the nominal conficence level α (say). Pap and Van Zuijlen [9] have proved the conservatism for large sample sizes n in case $\alpha \leq \frac{1}{2}$ and the anticonservatism for $\alpha > \frac{1}{2}$. Moreover they proposed a modified Stringer bound which has the right confidence level for large samples. See also Bickel [2]. In Lucassen, Moors and

Mathematics Subject Classification: 62G30, 62G20.
Key words and phrases: Stringer bound, "all or nothing method", "my dollar right or wrong method", confidence upper bound, monetary unit sampling, taintings.
Research supported by the Limperg Institute, in the Netherlands, and by OTKA T 032361.

Van Batenburg [6] several other modifications of the Stringer bound are proposed with the aim to decrease the conservatism. Some of these modifications are based on their conjecture that the coefficients in the Stringer bound are decreasing for $\alpha < 0.28$. This statement and also some of the proposed alternatives for the Stringer bound are supported by their simulation study.

In this paper we will investigate analytically the monotonicity statement mentioned above and the behavior of the proposed alternatives (and others) for the Stringer bound. We conclude that in general these alternatives cannot be considered as competitors of the Stringer bound. The results are presented in Section 2. In Section 3 we prove the superiority of the "my dollar right or wrong method" (also called the "all or nothing method") as compared with the use of the Stringer bound which is based on the taintings. It confirms a conjecture of G.B. Broeze (1998, personal communication). This somewhat suprising result is important for practical purposes and throws quite a different light on the effectiveness of procedures based on the Stringer bound. The proofs of the theorems are given in Section 4.

2. Results

Let $T_1, T_2, \ldots, T_n$ be independent, identically distributed random variables with distribution function F on $[0,1]$ and let $\mu := \mathbb{E}T_1$. Our aim is to find confidence upper bounds for μ of level $1-\alpha$. Therefore let

$$0 =: T_{0:n} \le T_{1:n} \le \cdots \le T_{n:n} \le T_{n+1:n} := 1$$

be the order statistics of $T_1, \ldots, T_n$. Moreover, for $\alpha \in (0,1)$ and $j = 0, 1, \ldots, n-1$, let $p_n(j; 1-\alpha)$ be the unique solution in p of the equation

$$\sum_{k=j+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \mathbb{P}(\mathrm{Bin}(n,p) \ge j+1) = 1-\alpha,$$

and let $p_n(n; 1-\alpha) := 1$, $p_n(-1, 1-\alpha) := 0$. We define for $j = 0, 1, \ldots, n$ (omitting the α in the notation)

$$c_n(j) := p_n(j) - p_n(j-1).$$

The definition of the well-known Stringer bound μ_{1n} is

$$\begin{aligned}\mu_{1n} &:= p_n(0) + \sum_{j=1}^{n}(p_n(j) - p_n(j-1))T_{n-j+1:n} \\ &= \sum_{j=0}^{n} c_n(j)T_{n-j+1:n}.\end{aligned}$$

There has been a common belief in the auditing literature that the Stringer bound works well in the sense that the real confidence probability is at least the nominal confidence level ($1-\alpha$, say). Or in mathematical terms

$$\mathbb{P}\left\{\overline{\mu}_{ST}^{(n)} \geq \mu\right\} \geq 1-\alpha \tag{1}$$

for all $\alpha \in (0,1)$, for all $n \geq 1$ and for all underlying distributions F of the taintings T_i. However, simulation studies indicated that the Stringer bound is rather conservative which could mean that the difference of the probability in (1) and $1-\alpha$ can be rather large.

Bickel [2] has proved that

$$\mathbb{P}\left\{\overline{\mu}_{ST}^{(n)} \geq \mu\right\} \geq \begin{cases} (1-\alpha)^{n+1} & \text{for } n \geq 2 \\ (1-\alpha) & \text{for } n = 1 \end{cases}$$

under certain conditions on the distribution of the taintings T_i. Let Φ denote the standard normal distribuiton function and let $z_{1-\alpha} := \Phi^{-1}(1-\alpha)$. In Pap and Van Zuijlen [8] the following asymptotic expansion has been obtained for the Stringer bound:

$$\mu_{1n} = \frac{1}{n}\sum_{j=1}^{n} T_j + \frac{c_1(F)}{\sqrt{n}} z_{1-\alpha} + o(n^{-\frac{1}{2}}) \qquad \text{a.s.}$$

where

$$c_1(F) := \int_0^1 F^{-1}(t)\frac{2t-1}{2\sqrt{t(1-t)}}\,dt = \int_0^1 \sqrt{t(1-t)}\,dF^{-1}(t).$$

Since $c_1(F) \geq \sigma(F) := \sqrt{\mathrm{Var}\ T_1}$, the asymptotic expansion implies that

$$(2) \quad \lim_{n\to\infty} \mathbb{P}(\mu_{1n} \geq \mu) = \Phi\left(\frac{c_1(F)}{\sigma(F)} z_{1-\alpha}\right) \begin{cases} \geq \Phi(z_{1-\alpha}) = 1-\alpha & \text{for } \alpha \in \left[0, \frac{1}{2}\right], \\ \leq \Phi(z_{1-\alpha}) = 1-\alpha & \text{for } \alpha \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

In other words, the Stringer bound is not asymptotically conservative for $\alpha \in [\frac{1}{2}, 1]$, which implies that also in a finite sampling situation the Stringer bound does not necessarily have the right confidence level. Pap and Van Zuijlen [7] gave recursive relations for obtaining the exact distribution of the Stringer bound in case if the underlying distribution of the taintings is a uniform distribution on the interval $[0,1]$, or a distribution with positive mass at zero and conditionally uniform on $(0,1]$. Based on these recurrence relations a concrete example has been found where the Stringer bound is not conservative. Moreover, (2) implies that

$$\lim_{r\to\infty} \Phi(r z_{1-\alpha}) = \begin{cases} 1 & \text{for } \alpha \in \left(0, \frac{1}{2}\right), \\ \frac{1}{2} & \text{for } \alpha = \frac{1}{2}, \\ 0 & \text{for } \alpha \in \left(\frac{1}{2}, 1\right), \end{cases}$$

hence if the ratio $r = c_1(F)/\sigma(F)$ is a large number then the Stringer bound is highly asymptotically conservative for $\alpha \in (0, \frac{1}{2})$ and it is highly asymptotically non-conservative for $\alpha \in (\frac{1}{2}, 1)$. Pap and Van Zuijlen [8] showed that the ratio $r = c_1(F)/\sigma(F)$ can be an arbitrary large number. Of course, both the highly asymptotically conservative and highly asymptotically non-conservative cases are not desirable. Note that Pap and Van Zuijlen [8] contains several limit theorems for the Stringer bound, and from the general results in Pap and Van Zuijlen [9] a Berry-Esseen inequality has been derived.

In Pap and Van Zuijlen [8] a so called modified Stringer bound $\widetilde{\mu}_n$ has been proposed in order to remove the conservatism, respectively, anticonservatism. This $\widetilde{\mu}_n$ is defined as follows

$$\widetilde{\mu}_n := \mu_{1n} - \frac{c_1(F_n) - \sigma(F_n)}{\sqrt{n}} z_{1-\alpha},$$

where F_n denotes the empirical distribution function based on $T_1, T_2, \ldots, T_n$:

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,t]}(T_i),$$

with

$$\mathbf{1}_{[0,t]}(T_i) := \begin{cases} 1 & \text{if } T_i \le t \\ 0 & \text{otherwise.} \end{cases}$$

It has been shown that this modified Stringer bound $\widetilde{\mu}_n$ is asymptotically correct for all α, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\widetilde{\mu}_n \ge \mu) = 1 - \alpha, \quad \text{for } \alpha \in (0,1).$$

Next, let us define

$$M_n := \sum_{i=1}^{n} \mathbf{1}_{(0,1]}(T_i), \qquad \varrho := \mathbb{P}(T_i > 0).$$

Note that M_n is the number of nonzero $T_1, T_2, \ldots, T_n$,

$$\frac{M_n}{n} \xrightarrow{\text{a.s.}} \varrho \quad (n \to \infty)$$

by the strong law of large numbers, and

$$\mu_{1n} = \sum_{j=0}^{M_n} c_n(j) T_{n-j+1:n}.$$

We will consider the following four modifications of the Stringer bound (with $T_0 := 0$):

$$\mu_{2n} := \sum_{j=0}^{M_n} c_n(n - M_n + j) T_{n-j+1:n}, \qquad \mu_{3n} := \sum_{j=0}^{n} c_n(j) T_{j:n},$$

$$\mu_{4n} := \sum_{j=0}^{M_n} c_n(j) T_{n-M_n+j:n}, \qquad \mu_{5n} := \sum_{j=0}^{n} c_n(j) T_j,$$

and note that $\mu_{1n} = \mu_{2n}$ and $\mu_{3n} = \mu_{4n}$ in case if $M_n = n$, i.e., if the distribution of the taintings has no mass in zero.

Remark 1. In Van Batenburg et al. [6] μ_{4n} and μ_{5n} are introduced and called ITO (Increased Tainting Order) and RTO (Random Tainting Order), respectively. For completeness we have added the two others μ_{2n} and μ_{3n}, which are quite natural in this setting.

For these modifications we can prove the following asymptotic expansions.

Theorem 1. *We have a.s.*

$$\mu_{2n} = \frac{1}{n}\sum_{i=1}^{n} T_i + \frac{c_2(F)}{\sqrt{n}} z_{1-\alpha} + o(n^{-\frac{1}{2}}),$$

$$\mu_{3n} = \frac{1}{n}\sum_{i=1}^{n} T_i - \frac{c_1(F)}{\sqrt{n}} z_{1-\alpha} + o(n^{-\frac{1}{2}}),$$

$$\mu_{4n} = \frac{1}{n}\sum_{i=1}^{n} T_i - \frac{c_2(F)}{\sqrt{n}} z_{1-\alpha} + o(n^{-\frac{1}{2}}),$$

and

$$\mu_{5n} \xrightarrow{\mathbb{P}} \mu,$$

where $c_1(F)$ *and* $z_{1-\alpha}$ *are defined as earlier and*

$$\begin{aligned} c_2(F) &:= \int_0^{\varrho} F^{-1}(t+1-\varrho)\frac{2t-1}{2\sqrt{t(1-t)}}\, dt \\ &= \int_0^{\varrho} \sqrt{t(1-t)}\, dF^{-1}(t+1-\varrho) - \sqrt{\varrho(1-\varrho)}. \end{aligned}$$

Remark 2. It follows from Theorem 1 that for instance

$$\lim_{n\to\infty} \mathbb{P}(\mu_{2n} \geq \mu) = \Phi\left(\frac{c_2(F)}{\sigma(F)} z_{1-\alpha}\right).$$

However, in this case we do not know whether

$$c_2(F) \geq \sigma(F).$$

In fact we have

$$c_2(F) \leq c_1(F),$$

since $c_2(F)$ can be written in the form

$$c_2(F) = \int_{1-\varrho}^{1} \frac{2(t+\varrho)-3}{2\sqrt{(t-1+\varrho)(2-t-\varrho)}} F^{-1}(t)\,dt,$$

and the integrand is monotone increasing on the interval $[1-\varrho, 1]$, hence

$$c_2(F) \le \int_{1-\varrho}^{1} \frac{2t-1}{2\sqrt{t(1-t)}} F^{-1}(t)\,dt = \int_{0}^{1} \frac{2t-1}{2\sqrt{t(1-t)}} F^{-1}(t)\,dt = c_1(F).$$

But the following examples shows that $c_2(F)$ can be smaller but also greater than $\sigma(F)$, depending on F. Hence $\lim\limits_{n\to\infty} \mathbb{P}(\mu_{2n} \ge \mu)$ can be on both sides of $1-\alpha$.

Example 1. Let X be a random variable with

$$\begin{cases} \mathbb{P}(X=0) = 1-\varrho, \\ \mathbb{P}(X=x) = \varrho, \end{cases}$$

where $x \in (0,1)$. We have

$$\mu = \mathbb{E}X = \varrho x, \qquad \mathbb{E}X^2 = \varrho x^2,$$

so that

$$\sigma^2(X) = \operatorname{Var}(X) = \varrho(1-\varrho)x^2.$$

Moreover,

$$\begin{aligned} c_2(X) &= \int_0^{\varrho} \frac{2u-1}{2\sqrt{u(1-u)}} F^{-1}(u+1-\rho)\,du \\ &= \int_0^{\varrho} \frac{2u-1}{2\sqrt{u(1-u)}} x\,du = \Big[-x\sqrt{u(1-u)}\Big]_{u=0}^{u=\varrho} \\ &= -\,x\sqrt{\varrho(1-\varrho)} < 0 < \sigma(X) = x\sqrt{\varrho(1-\varrho)}. \end{aligned}$$

Hence in this example we have

$$c_2(X) < 0 < \sigma(X).$$

Example 2. The purpose of this example (which is an extension of Example 1) is to demonstrate that $c_2(X)$ can also be positive. Let X be a random variable with

$$\begin{cases} \mathbb{P}(X = 0) = 1 - \varrho, \\ \mathbb{P}(X = x) = \varrho - q, \\ \mathbb{P}(X = 1) = q, \end{cases}$$

where $x \in (0, 1)$ and $0 < q < \varrho < 1$. We have

$$\mathbb{E}X = \varrho x + q(1 - x), \qquad \mathbb{E}X^2 = \varrho x^2 + q(1 - x^2)$$
$$\sigma^2(F) = \varrho(1 - \varrho)x^2 + q(1 - q)(1 - x^2) - 2q(\varrho - q)x(1 - x)$$

and

$$\int_0^{\varrho} \sqrt{u(1 - u)}\, dF^{-1}(u + 1 - \rho) = (1 - x)\sqrt{(\varrho - q)(1 - \varrho + q)}$$

so that

$$c_2(X) = \sqrt{(\varrho - q)(1 - \varrho + q)} - \sqrt{\varrho(1 - \varrho)} - x\sqrt{(\varrho - q)(1 - \varrho + q)}.$$

Taking for example $q = \frac{1}{4}$ and $\varrho = \frac{1}{2}$ we obtain

$$c_2(X) = \frac{1 - x}{2} - \frac{\sqrt{3}}{4} \begin{cases} > 0 & \text{for } x < 1 - \dfrac{1}{2}\sqrt{3}, \\ < 0 & \text{for } x > 1 - \dfrac{1}{2}\sqrt{3}. \end{cases}$$

Clearly $c_2(F)$ can be even larger than $\sigma(F)$ since if F is not concentrated on only two points and $\varrho = 1$ then $c_2(F) = c_1(F) > \sigma(F)$.

Remark 3. Let us consider the proposal to replace the Stringer bound μ_{1n} by μ_{2n} from another point of view. Note that one can rewrite μ_{1n} and μ_{2n} as follows:

$$\mu_{1n} = \sum_{j=0}^{n} c_n(j) T_{n-j+1:n} = \sum_{j=1}^{n+1} p_n(n - j + 1)(T_{j:n} - T_{j-1:n})$$

$$\mu_{2n} = \sum_{j=0}^{M_n} c_n(n - M_n + j)T_{n-j+1:n}$$
$$= \sum_{j=n-M_n+1}^{n+1} p_n(2n - M_n - j + 1)(T_{j:n} - T_{j-1:n}) - p_n(n - M_n - 1).$$

Theorem 2 in De Jager, Pap and Van Zuijlen [4] states that the coefficients in the Stringer bound μ_{1n} are minimal in a certain sense. Namely, if

$$\mathbb{P}(\mu_{2n} \geq \mu) \geq 1 - \alpha$$

is satisfied for all $n \in \mathbb{N}$, $\alpha \in (0,1)$ and for all distributions concentrated on the set $\{0,1\}$, then the coefficients of μ_{2n} have to be not less than the coefficients of μ_{1n}, that is,

$$p_n(2n - M_n - j + 1) \geq p_n(n - j + 1) \quad \text{for } j = 1, 2, \ldots, n+1.$$

Note, that these inequalities hold, since $p_n(j)$ is increasing in j.

Remark 4. To compare the alternative μ_{4n} for μ_{1n}, let $U_1, U_2, \ldots, U_n$ be i.i.d. random variables, uniformly distributed on the interval $[0,1]$, with order statistics

$$U_{1:n} \leq U_{2:n} \leq \cdots \leq U_{n:n}.$$

We define

$$\widetilde{U}_i := 1 - U_i \quad \text{for } i = 1, 2, \ldots, n$$

so that

$$\widetilde{U}_1, \widetilde{U}_2, \ldots, \widetilde{U}_n$$

are also i.i.d. random variables, uniformly distributed on the interval $[0,1]$, with order statistics

$$\widetilde{U}_{1:n} \leq \widetilde{U}_{2:n} \leq \cdots \leq \widetilde{U}_{n:n}.$$

With probability one we have

$$1 - \widetilde{U}_{i:n} = U_{n-i+1:n} \quad \text{for } i = 1, 2, \ldots, n.$$

Note that $\mu = \mathbb{E}U_i = \mathbb{E}\widetilde{U}_i = \frac{1}{2}$ and note that in this case we have $M_n = n$ with probability one. Writing down the bounds μ_{1n} and μ_{4n} for the sample $U_1, U_2, \ldots, U_n$ respectively. $\widetilde{U}_1, \widetilde{U}_2, \ldots, \widetilde{U}_n$ we have

$$\mu_{1n} = \sum_{j=0}^{n} c_n(j) U_{n-j+1:n} \qquad \mu_{4n} = \sum_{j=0}^{n} c_n(j) \widetilde{U}_{j:n}.$$

Since μ_{1n} is an asymptotic $(1-\alpha)$-*upper* confidence bound for $\mu = \frac{1}{2}$, it follows from symmetry reasons that $1 - \mu_{1n}$ is an asymptotic $(1-\alpha)$-*lower* conficence bound for $\mu = \frac{1}{2}$. Moreover, with probability one, we have

$$\begin{aligned} 1 - \mu_{1n} &= 1 - \sum_{j=0}^{n} c_n(j) U_{n-j+1:n} = 1 - \sum_{j=0}^{n} c_n(j)(1 - \widetilde{U}_{j:n}) \\ &= \mu_{4n} - p_n(0) = \mu_{4n} + O\left(\frac{1}{n}\right). \end{aligned}$$

We conclude that in case of a uniform $[0,1]$ underlying distribution μ_{4n} is even an asymptotic $(1-\alpha)$-lower bound for the mean, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\mu_{4n} \leq \mu) \geq 1 - \alpha.$$

The following theorem gives an analytic, but asymptotic support to the conjecture on page 19 in Lucassen, Moors and Van Batenburg [6] concerning the monotonicity of the function $j \mapsto c_n(j)$.

Theorem 2. *For $j = 1, 2, \ldots$ we have*

$$c_n(j) - c_n(j+1) = \frac{z_{1-\alpha}(n+1)^{\frac{1}{2}}}{4(j(n-j+1))^{\frac{3}{2}}} + O\left(\frac{n}{j^2(n-j+1)^2}\right) \quad \text{as } n \to \infty.$$

Hence if $0 < \gamma < \delta < 1$ then there exists $n_0(\alpha, \gamma, \delta)$ such that

$$c_n(j) > c_n(j+1) \qquad \text{for} \quad \alpha \in \left(0, \frac{1}{2}\right) \text{ and } n \geq n_0(\alpha, \gamma, \delta),\ \frac{j}{n} \in (\gamma, \delta)$$

and

$$c_n(j) < c_n(j+1) \qquad \text{for } \alpha \in \left(\frac{1}{2}, 1\right) \quad \text{and } n \geq n_0(\alpha, \gamma, \delta),\ \frac{j}{n} \in (\gamma, \delta).$$

Remark 5. We note that tedious calculations leads to the longer expansion

$$c_n(j) - c_n(j+1) = \frac{z_{1-\alpha}(n+1)^{\frac{1}{2}}}{4(j(n-j+1))^{\frac{3}{2}}} + \frac{3z_{1-\alpha}(n+1)^{\frac{1}{2}}(n-2j+1)}{4(j(n-j+1))^{\frac{5}{2}}}$$
$$+ \frac{z_{1-\alpha}(43 - z_{1-\alpha}^2)(3(n+1)^2 - 8j(n+1) + 8j^2)}{144(n+1)^{\frac{1}{2}}(j(n-j+1))^{\frac{5}{2}}}$$
$$+ \frac{z_{1-\alpha}(3n-4j+3)}{2(n+1)^{\frac{1}{2}} j^{\frac{5}{2}} (n-j+1)^{\frac{3}{2}}}$$
$$+ \frac{z_{1-\alpha}(1 - z_{1-\alpha}^2)}{8(n+1)^{\frac{1}{2}}(j(n-j+1))^{\frac{3}{2}}} + O\left(\frac{n^2}{j^3(n-j+1)^3}\right).$$

Remark 6. We have for $j = 0, 1, \ldots, n+1$

$$p_n(n-j; 1-\alpha) = 1 - p_n(j-1; \alpha),$$

(see Pap and Van Zuijlen [8]) which implies

$$c_n(n-j; 1-\alpha) = c_n(j; \alpha),$$

hence

$$c_n(n-j; 1-\alpha) - c_n(n-j-1; 1-\alpha) = c_n(j; \alpha) - c_n(j+1; \alpha).$$

Particularly, in case $\alpha = \frac{1}{2}$ we obtain the symmetry

$$c_n\left(n-j; \frac{1}{2}\right) - c_n\left(n-j-1; \frac{1}{2}\right) = c_n\left(j; \frac{1}{2}\right) - c_n\left(j+1; \frac{1}{2}\right),$$

so that monotonicity of the function $j \mapsto c_n(j; \frac{1}{2})$ cannot take place.

3. "My dollar right or wrong" method

Let $U_1, U_2, \ldots, U_n$ be i.i.d. random variables, uniformly distributed on the interval $[0,1]$. We define the random variables $X_1, X_2, \ldots, X_n$ as follows

$$X_i := \mathbf{1}_{[0,T_i]}(U_i) = \mathbf{1}_{[U_i,1]}(T_i) = \begin{cases} 1 & \text{if } U_i \le T_i \le 1, \\ 0 & \text{elsewhere.} \end{cases}$$

It is clear that the X_i's are i.i.d. Bernoulli random variables with the same mean as the T_i's since

$$\mathbb{E}(X_i) = \mathbb{E}\mathbb{E}(X_i \mid T_i) = \mathbb{E}\mathbb{E}\left(\mathbf{1}_{[0,T_i]}(U_i) \mid T_i\right) = \mathbb{E}(T_i) = \mu.$$

Note that the T_i's represent the taintings, whereas the X_i's represent the "good" or "false" dollars, obtained from the taintings, when the "my dollar right or wrong" method is used.

The Stringer bound based on the Bernoulli random variables $X_1, X_2, \ldots, X_n$, denoted by ν_n, is

$$\nu_n := \sum_{j=1}^{n} c_n(j) X_{n-j+1:n} = p_n(X),$$

where $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics of the X_i's and where $X := \sum_{i=1}^{n} X_i$. As has been mentioned in Section 2,

$$\lim_{n\to\infty} \mathbb{P}(\mu_{1n} \geq \mu) = \Phi\left(\frac{c_1(F)}{\sigma(F)} z_{1-\alpha}\right),$$

and $c_1(F) \geq \sigma(F)$ implies that for $\alpha \in (0, \frac{1}{2}]$ we also have

$$\lim_{n\to\infty} \mathbb{P}(\mu_{1n} \geq \mu) \geq \Phi(\Phi^{-1}(1-\alpha)) = 1-\alpha.$$

Since X_1 has a 2-point distribution on $[0,1]$ we have $c_1(X_1) = \sigma(X_1)$; consequently

$$\lim_{n\to\infty} \mathbb{P}(\nu_n \geq \mu) = \Phi(z_{1-\alpha}) = \Phi(\Phi^{-1}(1-\alpha)) = 1-\alpha,$$

and we conclude that for $\alpha \in (0, \frac{1}{2}]$

$$\lim_{n\to\infty} \mathbb{P}(\nu_n \geq \mu) = 1-\alpha \leq \lim_{n\to\infty} \mathbb{P}(\mu_{1n} \geq \mu).$$

Hence, indeed, we see that the "my dollar right or wrong" method is asymptotically preferable to the tainting method, since asymptotically the coverage probability of ν_n is closer to $1-\alpha$ than that of μ_{1n}.

Moreover, the following important 'finite n result' holds:

$$\mathbb{P}(\nu_n \geq \mu) \geq 1-\alpha \qquad \text{for all } n = 1, 2, \ldots,$$

since X_1 has a 2-point distribution (see Theorem 1 in De Jager, Pap and Van Zuijlen [4]), and the closeness of $\mathbb{P}(\nu_n \geq \mu)$ to $1-\alpha$ is described by the following statement.

Theorem 3.

$$\mathbb{P}(\nu_n \geq \mu) = 1 - \alpha + O\left(\frac{1}{\sqrt{n}}\right) \qquad \text{as } n \to \infty.$$

4. Proofs

Proof of Theorem 1. We will use the method of the proof of Theorem 3 in Pap and Van Zuijlen [8]. Applying Theorem 2 in Pap and Van Zuijlen [8] for $m = 2$ we obtain

$$c_n(j) = \frac{1}{n+1} + \frac{z_{1-\alpha}}{(n+1)^{\frac{3}{2}}} \frac{n-2j+1}{2\sqrt{j(n-j+1)}} + O\left(\frac{1}{j(n-j+1)}\right).$$

We note that from the expansion for $m = 3$ given in Pap and Van Zuijlen [8] one term is missing; the true expansion is

$$c_n(j) = \frac{1}{n+1} + \frac{z_{1-\alpha}}{(n+1)^{\frac{3}{2}}} \frac{n-2j+1}{2\sqrt{j(n-j+1)}} + \frac{2(1-z_{1-\alpha})}{3(n+1)^2} + O\left(\frac{n^{\frac{1}{2}}}{(j(n-j+1))^{\frac{3}{2}}}\right).$$

But in fact, we need only the shorter expansion of the case $m = 2$, which implies

$$\left|\mu_{2n} - p_n(0) - \sum_{j=1}^{M_n} T_{n-j+1:n}\left(\frac{1}{n+1} + \frac{z_{1-\alpha}}{(n+1)^{\frac{3}{2}}} \frac{2(M_n-j)-n+1}{2\sqrt{(n-M_n+j)(M_n-j+1)}}\right)\right|$$
$$\leq c\sum_{j=1}^{M_n} \frac{T_{n-j+1:n}}{(n-M_n+j)(M_n-j+1)} \leq c\sum_{j=1}^{n} \frac{1}{j(n-j+1)}$$
$$= \frac{c}{n+1}\sum_{j=1}^{n}\left(\frac{1}{j} + \frac{1}{n-j+1}\right) = O\left(\frac{\log n}{n}\right).$$

Moreover

$$p_n(0) = 1 - \sqrt[n]{\alpha} = O(n^{-1}),$$

and we have

$$\begin{aligned}
&\sum_{j=1}^{M_n} \frac{2(M_n - j) - n + 1}{2\sqrt{(n - M_n + j)(M_n - j + 1)}} T_{n-j+1:n} \\
&\qquad = \sum_{k=n-M_n+1}^{n} \frac{2(M_n + k) - 3n - 1}{2\sqrt{(M_n - n + k)(2n - M_n - k + 1)}} T_{k:n} \\
&\qquad = n \int_{1-\frac{M_n - 1}{n}}^{1} \frac{2\left(t + \frac{M_n}{n}\right) - 3 - \frac{1}{n}}{2\sqrt{\left(2 - \frac{M_n}{n} - t + \frac{1}{n}\right)\left(\frac{M_n}{n} - 1 + t\right)}} F_n^{-1}(t)\, dt.
\end{aligned}$$

Hence we have the expansion

$$\begin{aligned}
\mu_{2n} = \frac{1}{n} \sum_{j=1}^{n} T_j \\
&+ \frac{n z_{1-\alpha}}{(n+1)^{\frac{3}{2}}} \int_{1-\frac{M_n - 1}{n}}^{1} \frac{2\left(t + \frac{M_n}{n}\right) - 3 - \frac{1}{n}}{2\sqrt{\left(2 - \frac{M_n}{n} - t + \frac{1}{n}\right)\left(\frac{M_n}{n} - 1 + t\right)}} F_n^{-1}(t)\, dt \\
&+ O\left(\frac{\log n}{n}\right).
\end{aligned}$$

As in Lemma 5 in Pap and Van Zuijlen [8] we obtain

$$\begin{aligned}
&\int_{1-\frac{M_n - 1}{n}}^{1} \frac{2\left(t + \frac{M_n}{n}\right) - 3 - \frac{1}{n}}{2\sqrt{\left(2 - \frac{M_n}{n} - t + \frac{1}{n}\right)\left(\frac{M_n}{n} - 1 + t\right)}} F_n^{-1}(t)\, dt \\
&\qquad \xrightarrow{\text{a.s.}} \int_{1-\varrho}^{1} \frac{2(t + \varrho) - 3}{2\sqrt{(2 - \varrho - t)(\varrho - 1 + t)}} F_n^{-1}(t)\, dt = c_2(F).
\end{aligned}$$

Consequently

$$\mu_{2n} = \frac{1}{n} \sum_{j=1}^{n} T_j + \frac{c_2(F)}{\sqrt{n}} z_{1-\alpha} + o(n^{-\frac{1}{2}}) \qquad \text{a.s.}$$

The expansion for μ_{3n} and μ_{4n} can be proved similarly.

Moreover, we clearly have

$$\mathbb{E}\mu_{5n} = \sum_{j=1}^{n} c_n(j)\mathbb{E}T_j = \mu \sum_{j=1}^{n} c_n(j) = \mu(1 - p_n(0)),$$

$$p_n(0) = 1 - \sqrt[n]{\alpha} = O(n^{-1}),$$

and

$$\operatorname{Var} \mu_{5n} = \sum_{j=1}^{n} (c_n(j))^2 \operatorname{Var} T_j = \sigma^2 \sum_{j=1}^{n} (c_n(j))^2.$$

Applying Theorem 2 in Pap and Van Zuijlen [8] for $m = 1$ we obtain

$$c_n(j) = \frac{1}{n+1} + O\left(\frac{1}{\sqrt{nj(n-j+1)}}\right),$$

hence

$$(c_n(j))^2 \le \frac{2}{(n+1)^2} + \frac{c}{nj(n-j+1)}$$

which implies

$$\sum_{j=1}^{n} (c_n(j))^2 \le \frac{2n}{(n+1)^2} + \frac{c}{n} \sum_{j=1}^{n} \frac{1}{j(n-j+1)} = O\left(\frac{1}{n}\right).$$

Thus for all $\varepsilon > 0$ we have

$$\mathbb{P}(|\mu_{5n} - \mathbb{E}\mu_{5n}| \ge \varepsilon) \le \frac{\operatorname{Var} \mu_{5n}}{\varepsilon^2} \to 0 \qquad \text{as} \quad n \to \infty,$$

consequently we obtain $\mu_{5n} - \mathbb{E}\mu_{5n} \xrightarrow{\mathbb{P}} 0$ and finally $\mu_{5n} \xrightarrow{\mathbb{P}} \mu$. See also Schapendonk-Maas [12]. □

Proof of Theorem 2. Applying Theorem 1 in Pap and Van Zuijlen [8] for $m = 3$ we obtain the expansion

$$\frac{(n+1)^{\frac{3}{2}}}{(j(n-j+1))^{\frac{1}{2}}} \left(p_n(j-1) - \frac{j}{n+1}\right)$$

$$= z_{1-\alpha} + R_{1,j,n}(z_{1-\alpha}) + R_{2,j,n}(z_{1-\alpha}) + O\left(\frac{(n+1)^{\frac{3}{2}}}{(j(n-j+1))^{\frac{3}{2}}}\right),$$

where

$$R_{1,j,n}(z) = -S_{1,j,n}(z) = -\frac{n-2j+1}{3(n+1)^{\frac{1}{2}}(j(n-j+1))^{\frac{1}{2}}}(1-z^2),$$

$$R_{2,j,n}(z) = S_{1,j,n}(z)S'_{1,j,n}(z) - \frac{z}{2}(S_{1,j,n}(z))^2 - S_{2,j,n}(z)$$

$$= \frac{z}{j} + \frac{z(1-z^2)}{4(n+1)} - \frac{(n-2j+1)^2 z(43-z^2)}{36(n+1)j(n-j+1)}.$$

Hence

$$p_n(j-1) = \frac{j}{n+1} + \frac{(j(n-j+1))^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}} z_{1-\alpha} - \frac{n-2j+1}{3(n+1)^2}(1-z_{1-\alpha}^2)$$

$$- \frac{(n-2j+1)^2 z_{1-\alpha}(43-z_{1-\alpha}^2)}{36(n+1)^{\frac{5}{2}}(j(n-j+1))^{\frac{1}{2}}} - \frac{(n-j+1)^{\frac{1}{2}} z_{1-\alpha}}{(n+1)^{\frac{3}{2}} j^{\frac{1}{2}}}$$

$$+ \frac{(j(n-j+1))^{\frac{1}{2}} z_{1-\alpha}(1-z_{1-\alpha}^2)}{4(n+1)^{\frac{5}{2}}} + O\left(\frac{1}{j(n-j+1)}\right).$$

Using Taylor expansion of the functions $t \mapsto (t(1-t))^{\frac{1}{2}}$, $t \mapsto \frac{(1-2t)^2}{(t(1-t))^{\frac{1}{2}}}$ and $t \mapsto \left(\frac{1-t}{t}\right)^{\frac{1}{2}}$ at the point $t = \frac{j}{n+1}$ we obtain

$$((j+1)(n-j))^{\frac{1}{2}} - (j(n-j+1))^{\frac{1}{2}}$$

$$= \frac{n-2j+1}{2(j(n-j+1))^{\frac{1}{2}}} - \frac{(n+1)^2}{2(j(n-j+1))^{\frac{3}{2}}} + O\left(\frac{n^3}{(j(n-j+1))^{\frac{5}{2}}}\right),$$

$$\frac{(n-2j-1)^2}{((j+1)(n-j))^{\frac{1}{2}}} - \frac{(n-2j+1)^2}{(j(n-j+1))^{\frac{1}{2}}}$$

$$= -\frac{(n+1)^2(n-2j+1)}{2(j(n-j+1))^{\frac{3}{2}}} + O\left(\frac{n^4}{(j(n-j+1))^{\frac{5}{2}}}\right),$$

$$\left(\frac{n-j}{j+1}\right)^{\frac{1}{2}} - \left(\frac{n-j+1}{j}\right)^{\frac{1}{2}}$$
$$= -\frac{n+1}{2j^{\frac{3}{2}}(n-j+1)^{\frac{1}{2}}} + O\left(\frac{n^2}{j^{\frac{5}{2}}(n-j+1)^{\frac{3}{2}}}\right).$$

We remark that by using Theorem 1 in Pap and Van Zuijlen [8] for $m = 5$ we can even derive an expansion for $p_n(j-1)$ with remainder term

$$O\left(\frac{n}{j^2(n-j+1)^2}\right).$$

The differrences of the additional terms in the expansion of $p_n(j)$ and $p_n(j-1)$ turn out to be of order $O\left(\frac{n}{j^2(n-j+1)^2}\right)$ due to the smoothness of the functions involved. In this way we obtain the expansion for $c_n(j) = p_n(j) - p_n(j-1)$:

$$c_n(j) = \frac{1}{n+1} + \frac{z_{1-\alpha}}{(n+1)^{\frac{3}{2}}}\left(\frac{n-2j+1}{2(j(n-j+1))^{\frac{1}{2}}} - \frac{(n+1)^2}{2(j(n-j+1))^{\frac{3}{2}}}\right)$$
$$+ \frac{2(1-z_{1-\alpha})}{3(n+1)^2} + \frac{z_{1-\alpha}(43-z_{1-\alpha}^2)}{36(n+1)^{\frac{1}{2}}}\frac{n-2j+1}{2(j(n-j+1))^{\frac{3}{2}}}$$
$$+ \frac{z_{1-\alpha}}{(n+1)^{\frac{1}{2}}}\frac{1}{2j^{\frac{3}{2}}(n-j+1)^{\frac{1}{2}}}$$
$$+ \frac{z_{1-\alpha}(1-z_{1-\alpha}^2)}{4(n+1)^{\frac{5}{2}}}\frac{n-2j+1}{2(j(n-j+1))^{\frac{1}{2}}} + O\left(\frac{n}{j^2(n-j+1)^2}\right).$$

Using Taylor expansion of the functions $t \mapsto \frac{1-2t}{(t(1-t))^{\frac{1}{2}}}$, $t \mapsto (t(1-t))^{-\frac{3}{2}}$, $t \mapsto \frac{1-2t}{(t(1-t))^{\frac{3}{2}}}$ and $t \mapsto \frac{1}{t^{\frac{3}{2}}(1-t)^{\frac{1}{2}}}$ at the point $t = \frac{j}{n+1}$ we obtain

$$\frac{n-2j+1}{2(j(n-j+1))^{\frac{1}{2}}} - \frac{n-2j-1}{2((j+1)(n-j+1))^{\frac{1}{2}}}$$
$$= \frac{(n+1)^2}{4(j(n-j+1))^{\frac{3}{2}}} + O\left(\frac{n^{\frac{5}{2}}}{j^2(n-j+1)^2}\right),$$

$$\frac{1}{(j(n-j+1))^{\frac{3}{2}}} - \frac{1}{((j+1)(n-j))^{\frac{3}{2}}} = O\left(\frac{n^{\frac{1}{2}}}{j^2(n-j+1)^2}\right),$$

$$\frac{n-2j+1}{(j(n-j+1))^{\frac{3}{2}}} - \frac{n-2j+1}{((j+1)(n-j))^{\frac{3}{2}}} = O\left(\frac{n^{\frac{3}{2}}}{j^2(n-j+1)^2}\right),$$

$$\frac{1}{j^{\frac{3}{2}}(n-j+1)^{\frac{1}{2}}} - \frac{1}{(j+1)^{\frac{3}{2}}(n-j)^{\frac{1}{2}}} = O\left(\frac{n^{\frac{3}{2}}}{j^2(n-j+1)^2}\right).$$

Finally we conclude the expansion of $c_n(j)-c_n(j+1)$ given in the theorem. Thus there is a constant $c>0$ such that

$$\left| c_n(j) - c_n(j+1) - \frac{z_{1-\alpha}\sqrt{n+1}}{4(j(n-j+1))^{\frac{3}{2}}} \right| \le \frac{cn}{j^2(n-j+1)^2}.$$

Let now $\alpha \in (0, \frac{1}{2})$, which implies $z_{1-\alpha} > 0$. If $\frac{j}{n} \in (\gamma, \delta)$ with some $0 < \gamma < \delta < 1$ then we have

$$\frac{z_{1-\alpha}\sqrt{n+1}}{4(j(n-j+1))^{\frac{3}{2}}} \ge \frac{cn}{j^2(n-j+1)^2}$$

for sufficiently large n, since

$$\begin{aligned} \frac{\sqrt{(n+1)j(n-j+1)}}{n} &= \sqrt{(n+1)\frac{j}{n}\left(1-\frac{j}{n}+\frac{1}{n}\right)} \\ &\ge \sqrt{(n+1)\gamma\left(1-\delta+\frac{1}{n}\right)} \ge \frac{4c}{z_{1-\alpha}} \end{aligned}$$

for sufficiently large n. Consequently, we obtain the monocity statement for the function $j \mapsto c_n(j)$. The case $\alpha \in (\frac{1}{2}, 1)$ can be handled similarly. □

Proof of Theorem 3. Clearly

$$\mathbb{P}(X=j) = \binom{n}{j}\mu^j(1-\mu)^{n-j} \quad \text{for } j = 0, 1, \ldots, n.$$

Let n_μ be the index such that

$$p_n(n_\mu) \le \mu < p_n(n_\mu + 1).$$

First we show that $\frac{n_\mu}{n} \to \mu$ as $n \to \infty$. We have the expansion

$$p_n(j-1) = \frac{j}{n+1} + O\left(\frac{(j(n-j+1))^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}}\right)$$

(see Theorem 1 in Pap and Van Zuijlen [8]), hence there exists $c > 0$ such that

$$\mu \geq p_n(n_\mu) \geq \frac{n_\mu + 1}{n+1} - c\,\frac{((n_\mu+1)(n-n_\mu))^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}}$$

and

$$\mu < p_n(n_\mu+1) \leq \frac{n_\mu + 2}{n+1} + c\,\frac{((n_\mu+2)(n-n_\mu-1))^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}}.$$

Hence we can conclude

$$-c\,\frac{((n_\mu+2)(n-n_\mu-1))^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}} - \frac{1}{n+1} \leq \frac{n_\mu+1}{n+1} - \mu \leq c\,\frac{((n_\mu+1)(n-n_\mu))^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}},$$

which together with $0 \leq n_\mu \leq n$ imply $\frac{n_\mu}{n} \to \mu$ as $n \to \infty$.

We have

$$\mathbb{P}(\nu_n \geq \mu) = \mathbb{P}(p_n(X) \geq \mu) = \sum_{j=n_\mu+1}^{n} \binom{n}{j} \mu^j (1-\mu)^{n-j} = \int_0^\mu n \binom{n-1}{n_\mu} x^{n_\mu} (1-x)^{n-n_\mu-1}\,dx$$

and

$$1-\alpha = \sum_{j=n_\mu+1}^{n} \binom{n}{j} (p_n(n_\mu))^j (1-p_n(n_\mu))^{n-j} = \int_0^{p_n(n_\mu)} n \binom{n-1}{n_\mu} x^{n_\mu} (1-x)^{n-n_\mu-1}\,dx,$$

hence

$$\mathbb{P}(\nu_n \geq \mu) - (1-\alpha) = \int_{p_n(n_\mu)}^{\mu} n\binom{n-1}{n_\mu} x^{n_\mu}(1-x)^{n-n_\mu-1}\,dx$$
$$\leq n(\mu - p_n(n_\mu)) \sup_{0\leq x\leq 1} \binom{n-1}{n_\mu} x^{n_\mu}(1-x)^{n-n_\mu-1}$$
$$= n(\mu - p_n(n_\mu))\binom{n-1}{n_\mu} q_n^{n_\mu}(1-q_n)^{n-n_\mu-1},$$

where $q_n := \frac{n_\mu}{n-1} \to \mu$ as $n \to \infty$. We have

$$\binom{n-1}{n_\mu} q_n^{n_\mu}(1-q_n)^{n-n_\mu-1} = \mathbb{P}(\mathrm{Bin}(n-1, q_n) = n_\mu) = O\left(\frac{1}{\sqrt{n}}\right),$$

since by the Moivre–Laplace theorem

$$\left|\mathbb{P}(\mathrm{Bin}(n-1, q_n) = n_\mu) - \frac{1}{\sqrt{2\pi(n-1)q_n(1-q_n)}}\right| \to 0 \quad \text{as } n \to \infty.$$

Moreover, we have

$$\mu - p_n(n_\mu) < p_n(n_\mu + 1) - p_n(n_\mu),$$

and Theorem 2 in Pap and Van Zuijlen [8] implies for some $c_1 > 0$

$$|p_n(j) - p_n(j-1)| \leq \frac{c_1}{n}\left(\frac{n+1}{j(n-j+1)}\right)^{\frac{1}{2}} \leq \frac{\sqrt{2}c_1}{n},$$

hence $\mu - p_n(n_\mu) = O\left(\frac{1}{n}\right)$. □

References

[1] P. C. van Batenburg and J. Kriens, Applications of statistical methods and techniques to auditing and accounting, Invited paper for the 48th Conference of the International Statistical Institute, Cairo, 9–17 September, *Proceedings,* and *Katholieke Universiteit Brabant*, 1991, FEW 493.

[2] P. J. Bickel, Inference and auditing: the Stringer bound, *International Statistical Review* **60** (1992), 197–209.

[3] R. K. Burdick and J. H. Reneau, The impact of different error distribution on the performance of selected sampling estimators in accounting populations, Proceedings of the Business and Economical Statistics Section, 1978, 779–781.

[4] N. G. de Jager, G. Pap and M. C. A. van Zuijlen, Facts, phantasies and a new proposal concerning the Stringer bound, *Computers Math. Applic.* **33** no. 5 (1997), 37–54.

[5] R. A. Leitch, J. Neter, R. Plante and P. Sinha, Modified multinomial bound for larger number of errors in audits, *The Accounting Review* no. 2 (April 1982), 384–400.

[6] A. Lucassen, H. Moors and P. C. van Batenburg, Modifications of the Stringer bound: A simulation study on the performance of audit sampling evaluation methods, *Kwantitative methoden* **51** (1996), 17–25.

[7] G. Pap and M. C. A. van Zuijlen, On the Stringer bound in case of uniform taintings, *Computers Math. Applic.* **29** no. 10 (1995), 51–59.

[8] G. Pap and M. C. A. van Zuijlen, On the asymptotic behaviour of the Stringer bound, *Statistica Neerlandica* **50** nr. 3 (1996), 367–389.

[9] G. Pap and M. C. A. van Zuijlen, A Berry-Esseen bound for linear combinations of order statistics, *Statistics* **31** (1998), 151–186.

[10] R. Plante, J. Neter and R. A. Leitch, Comperative performance of multinomial, cell and Stringer bounds, *Auditing: a journal of practice and theory* (1985), 40–56.

[11] J. H. Reneau, C.A.V. bounds in dollar unit sampling some simulation results, *The Accounting Review* no. 3 (1987), 669–680.

[12] H. Schapendonk-Maas, Betrouwbaarheids bovengrenzen in de accountancy, *Studentwork*, *University Nijmegen*, 1998.

[13] H. Tamura, Statistical models and Analysis in auditing, *Statistical Science* **4** (February 1989), 2–33.

G. PAP
INSTITUTE OF MATHEMATICS AND INFORMATICS
LAJOS KOSSUTH UNIVERSITY
H–4010 DEBRECEN, P.O. BOX 12
HUNGARY

M. C. A. VAN ZUIJLEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN
TOERNOOIVELD 1
6525 ED NIJMEGEN
THE NETHERLANDS

(Received June 4, 1999; revised March 6, 2000)