International Journal of Statistics and Probability; Vol. 2, No. 4; 2013
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Probability Inequalities for the Sum of Random Variables When Sampling Without Replacement

Jeremy Becnel[1], Kent Riggs[1] & Dean Young[2]

[1] Department of Mathematics and Statistics, Stephen F. Austin State University, Nacogdoches, Texas, USA

[2] Department of Statistical Systems, Baylor University, Waco, TX, USA

Correspondence: Jeremy Becnel, Department of Mathematics and Statistics, Stephen F. Austin State University, Nacogdoches, Texas, USA. E-mail: becneljj@sfasu.edu

Received: August 11, 2013 Accepted: September 24, 2013 Online Published: October 11, 2013

doi:10.5539/ijsp.v2n4p75 URL: http://dx.doi.org/10.5539/ijsp.v2n4p75

Abstract

Exponential-type upper bounds are formulated for the probability that the maximum of the partial sample sums of discrete random variables having finite equispaced support exceeds or differs from the population mean by a specified positive constant. The new inequalities extend the work of Serfling (1974). An example of the results are given to demonstrate their efficacy.

Keywords: probability bounds, discrete probability

1. Introduction

Serfling (1974) has obtained upper bounds for the probability that the sum of observations sampled without replacement from a finite population exceeds its expected value by a specified quantity. Serfling (1974) also noted that his bound is crude due to the incorporation of the coarse variance upper bound $\sigma^2 < (b-a)^2/4$ and has suggested that "it would be desirable to obtain a sharpening of this result involving the quantity σ^2 in place of the quantity $(b-a)^2/4$." While this problem remains unsolved, we attempt to at least partially fulfill Serfling's suggestion. In order to do this we tighten his inequality bound by restricting ourselves to a particular class of discrete distributions. The general problem which we address may be stated as follows.

Consider a finite population of size N whose members are not necessarily distinct. Let the set $\Omega_N = \{x_1, x_2, \ldots, x_N\}$ be the set representation of this population. Denote by $X_1, X_2, \ldots, X_n$ the values of a sample of size n drawn without replacement from Ω_N. Define the statistics

$$a = \min_{1\le i\le N} x_i, \qquad b = \max_{1\le i\le N} x_i, \qquad \mu = \sum_{i=1}^{N} \frac{x_i}{N}, \qquad \sigma^2 = \sum_{i=1}^{N} (x_i-\mu)^2/N, \tag{1.1}$$

and let the sampling fractions be $f_n = (n-1)/(N-1)$ and $g_n = (n-1)/N$. We are concerned with the behavior of the sum

$$S_k = \sum_{i=1}^{k} X_i \tag{1.2}$$

for $1 \le k \le n$. In particular, we derive a new parameter-free upper bounds on the probabilities

$$P_n(\varepsilon) = P[S_n - n\mu \ge n\varepsilon] \tag{1.3}$$

and

$$R_n(\varepsilon) = P\left[\max_{n\le k\le N} \frac{S_k - k\mu}{k} \ge \varepsilon\right] \tag{1.4}$$

where $\varepsilon > 0$.

The most familiar upper bound for (1.3) is the *Bienayme-Chebyshev inequality*, which is of the form

$$P_n(\varepsilon) \le \frac{(1-f_n)\sigma^2}{n\varepsilon^2}. \tag{1.5}$$

Serfling (1974) has derived an alternative upper bound for (1.3), which may be expressed as

$$P_n(\varepsilon) \leq \exp\left[\frac{-2n\varepsilon^2}{(1-g_n)(b-a)^2}\right] \tag{1.6}$$

and an alternative bound for a two-sided version of (1.4), which is given as

$$R_n^*(\varepsilon) = P\left[\max_{n\leq k\leq N}\left|\frac{S_k - k\mu}{k}\right| \geq \varepsilon\right] \leq \frac{E(S_n - n\mu)^{2r}}{(n\varepsilon)^{2r}} \tag{1.7}$$

where r is a positive integer.

We then compare our new under bound with these under the following scenario, which is similar to an example presented by Savage (1961). Suppose one wishes to study the average height of a finite population of people. Assume that all individuals in the population are between 60 and 78 inches and their heights are measured to the nearest inch. The question we wish to answer is, "What is the probability that the average height of a sample of 100 from a population of 4,000 individuals is within two inches of the population mean height?"

The main vehicle we utilize for the sharpening of inequalities (1.6) and (1.7) is the additional assumption that the random variables given by

$$(X_i|X_1,\ldots,X_{i-1}) \qquad \text{for } i = 1,\ldots,n$$

are discrete random variables with probability functions having finite, equispaced support, and whose variance is bounded above by the discrete uniform variance. The remainder of the paper is as follows. In Section 2 we give some mathematical preliminaries while in Section 2.1 we derive the main inequality results. Finally, in Section 3 we present the before mentioned application of the newly-derived probability bounds.

2. The Set $V_{a,b,J}$

From this point forward we work almost exclusively with an equispaced set of J points in the interval $[a, b]$ beginning at a and ending at b. We denote the set as

$$\Omega_{a,b,J} = \{a, a+c, a+2c, \ldots, a+(J-1)c\}$$

where $b = a + (J-1)c$ and $c = \frac{b-a}{J-1}$. We refer to J as the *support size*. As before $X_1, X_2, \ldots, X_n$ denote the values of a sample of size n drawn without replacement from $\Omega_{a,b,J}$ and $S_k = \sum_{i=1}^{k} X_i$.

To sharpen the inequalities (1.6) and (1.7) we work with probability distributions that have a variance bounded by the variance of a discrete uniform probability distribution. This leads us to the following definition.

Definition 2.1 Let $V_{a,b,J}$ be the set of probability functions f with support on $\Omega_{a,b,J}$ such that for a random variable X having probability function f the variance is bounded as per

$$\mathrm{Var}_f(X) \leq \frac{J+1}{J-1}\frac{(b-a)^2}{12}.$$

Remark 2.2 For $f \in V_{a,b,J}$, the above bound on $\mathrm{Var}_f(X)$ is simply the variance of a discrete uniform probability function on $\Omega_{a,b,J}$.

Remark 2.3 The definition for $V_{a,b,J}$, although appearing somewhat restrictive, still allows for a broad and rich range of distributions. In particular, it applies to a broad range of discrete unimodal distributions.

2.1 Probability Inequalities for $V_{a,b,J}$

In this section we derive a new maximal probability inequality for sums of discrete unimodal random variables sampled without replacement from a set of probability functions belonging to $V_{a,b,J}$. We shall need the following lemmas, theorems, and corollaries to develop the new maximal probability inequality. We now develop two lemmas which are used in the proof of the main theorem.

Lemma 2.4 *Let X be a random variable with probability function in $V_{a,b,J}$ and let $E(X) = \mu$. Then for any $\lambda \geq 0$ we have*

$$E[e^{\lambda(X-\mu)}] \leq \exp[\alpha(e^{\lambda d} - \lambda d - 1)],$$

where $d = b - a$ and $\alpha = \frac{J+1}{J-1}\frac{1}{12}$.

Proof. Let $Z = X - \mu$ and notice that

$$E[e^{\lambda Z}] = 1 + \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} E[Z^j] = 1 + \sum_{j=2}^{\infty} \frac{\lambda^j}{j!} E[Z^j], \tag{2.1}$$

since $E[Z] = 0$. Now let f be the probability function for X. Then for any $j \geq 2$ we have

$$\begin{aligned} E[Z^j] &= \sum_{x=a}^{b} (x-\mu)^j f(x) \\ &\leq \sum_{x=a}^{b} d^{j-2}(x-\mu)^2 f(x) \quad \text{since } x - \mu \leq d \\ &= d^{j-2} \sum_{x=a}^{b} (x-\mu)^2 f(x) \\ &= d^{j-2} \mathrm{Var}_f(X) \\ &\leq d^{j-2} \frac{J+1}{J-1} \frac{d^2}{12} \quad \text{since } f \in V_{a,b,f} \\ &\leq \frac{J+1}{J-1} \frac{d^j}{12}. \end{aligned}$$

Substituting the result $E[Z^j] \leq \frac{J+1}{J-1} \frac{d^j}{12}$ into (2.1), we get

$$\begin{aligned} E[e^{\lambda Z}] &\leq 1 + \sum_{j=2}^{\infty} \frac{\lambda^j}{j!} \frac{J+1}{J-1} \frac{d^j}{12} = 1 + \frac{J+1}{J-1} \frac{1}{12} \sum_{j=2}^{\infty} \frac{\lambda^j d^j}{j!} \\ &= 1 + \alpha \sum_{j=2}^{\infty} \frac{\lambda^j d^j}{j!} \\ &= 1 + \alpha[e^{\lambda d} - \lambda d - 1] \\ &\leq \exp[\alpha(e^{\lambda d} - \lambda d - 1)]. \end{aligned}$$

□

Corollary 2.5 *Let X be a random variable with probability function in $V_{a,b,J}$ and let $E(X) = \mu$. Then for any $\lambda \geq 0$ we have*

$$E[e^{\lambda(X-\mu)}] \leq \exp[\beta \lambda^2 r(\lambda, d)],$$

where $d = b - a$, $\beta = \frac{J+1}{J-1} \frac{(b-a)^2}{12}$, and $r(\lambda, d) = \sum_{j=2}^{\infty} \frac{\lambda^{j-2} d^{j-2}}{j!}$.

Proof. By Lemma 2.4, we have

$$E[e^{\lambda(X-\mu)}] \leq \exp\left[\frac{1}{12} \frac{J+1}{J-1} (e^{\lambda d} - \lambda d - 1)\right].$$

Thus,

$$\begin{aligned} \exp\left[\frac{1}{12} \frac{J+1}{J-1} (e^{\lambda d} - \lambda d - 1)\right] &= \exp\left[\frac{1}{12} \frac{J+1}{J-1} \sum_{j=2}^{\infty} \frac{\lambda^j d^j}{j!}\right] \\ &= \exp\left[\frac{(b-a)^2}{12} \frac{J+1}{J-1} \lambda^2 \sum_{j=2}^{\infty} \frac{\lambda^{j-2} d^{j-2}}{j!}\right] \\ &= \exp[\beta \lambda^2 r(\lambda, d)]. \end{aligned}$$

□

The following lemma uses an argument similar to Theorem 2.2 in Sefling's paper (1974).

Lemma 2.6 *Let $\Omega_N = \{x_1, x_2, \ldots, x_N\}$ where each x_k is in $\Omega_{a,b,J}$. Also let $\mu = \sum_{k=1}^{N} \frac{x_k}{N}$. If the probability function of X_1 and the conditional probability functions of $(X_k|X_1, X_2, \ldots, X_{k-1})$, $k = 2, \ldots, n$ are in $V_{a,b,J}$, then, for any $\lambda \geq 0$ and any $n \in \{1, 2, \ldots, N\}$ we have*

$$E[e^{\lambda(S_n - n\mu)}] \leq \exp\left[n\frac{1-g_n}{12}\frac{J+1}{J-1}(e^{\lambda d - \lambda d - 1})\right],$$

where $S_n = \sum_{k=1}^{n} X_k$.

Proof. For $\lambda = 0$, the result is obvious. Given $\lambda > 0$ let

$$\lambda_k = \frac{N-n}{N-k}\lambda \quad \text{for } 1 \leq k \leq n.$$

Notice λ_k is increasing in k up to λ as k goes from 1 to n.

Because $(X_k - \mu|X_{k-1}, \ldots X_1) \in V_{a,b,J}$ and $X_1 \in V_{a,b,J}$ letting μ_k denote the conditional expectation of $(X_k - \mu|X_{k-1}, \ldots, X_1)$ we have by Corollary 2.5 that

$$E[\exp(\lambda_k(X_k - \mu - \mu_k))|X_{k-1}, \ldots, X_1] \leq \exp[\beta\lambda_k r(\lambda_k, d)] \leq \exp[\beta\lambda_k r(\lambda, d)] \tag{2.2}$$

because $\lambda_k \leq \lambda$.

Using the conditional independence of the random variables S_{k-1} and $X_k - \mu - \mu_k$ given $X_{k-1}, \ldots, X_1$, we can apply Corollary 3.6 so that

$$E[\exp(\lambda_k(S_k - k\mu))] = E[\exp(\lambda_{k-1}(S_{k-1} - (k-1)\mu))]E[\exp(\lambda_k(X_k - \mu - \mu_k)) \,||X_{k-1}, \ldots, X_1]. \tag{2.3}$$

Combining (2.3) with (2.2) we have that

$$E[\exp(\lambda_k(S_k - k\mu))] \leq E[\exp(\lambda_{k-1}(S_{k-1} - (k-1)\mu))]\exp[\beta\lambda_k^2 r(\lambda, d)]. \tag{2.4}$$

Recursively applying (2.4) we get

$$E[\exp(\lambda_n(S_n - n\mu))] \leq \exp[\lambda^2\Delta_n\beta r(\lambda, d)], \tag{2.5}$$

where $\Delta_n = \sum_{k=1}^{n}\left[\frac{N-n}{N-k}\right]^2 = 1 + (N-n)^2\sum_{k=N-n+1}^{N-1}\frac{1}{k^2}$. Next, using that $\Delta_n \leq n(1-g_n)$ from (2.5) we get

$$E[\exp(\lambda_n(S_n - n\mu))] \leq \exp[\lambda^2 n(1-g_n)\beta r(\lambda, d)].$$

Replacing $r(\lambda, d)$ with $\frac{e^{\lambda d} - \lambda d - 1}{\lambda^2 d^2}$ we have

$$\begin{aligned} E[\exp(\lambda_n(S_n - n\mu))] &\leq \exp[\lambda^2 n(1-g_n)\beta\tfrac{e^{\lambda d} - \lambda d - 1}{\lambda^2 d^2}] \\ &= \exp\left[n\frac{1-g_n}{12}\frac{J+1}{J-1}(e^{\lambda - \lambda d - 1})\right]. \end{aligned}$$

□

Theorem 2.7 *Let X_k, S_k, and μ be as before, then for any $\varepsilon, \lambda > 0$ we have*

$$E[\exp(\lambda(S_n - n\mu - n\varepsilon))] \leq \exp\left\{-n\left[\left(\frac{d(1-g_n)(J+1) + 12(J-1)\varepsilon}{12(J-1)d}\right)\ln\left(\frac{12(J-1)\varepsilon}{(1-g_n)(J+1)d} + 1\right) - \frac{\varepsilon}{d}\right]\right\}$$

Proof. First note that by the Lemma 2.6 we have

$$\begin{aligned} E[\exp(\lambda(S_n - n\mu - n\varepsilon))] &= e^{-\lambda n\varepsilon}E[e^{\lambda(S_n - n\mu)}]. \\ &\leq e^{-\lambda n\varepsilon}\exp[\alpha(e^{\lambda d} - \lambda d - 1)] \quad \text{where} \quad \alpha = n\frac{1-g_n}{12}\frac{J+1}{J-1} \\ &= \exp[\alpha(e^{\lambda d} - \lambda d - 1) - \lambda n\varepsilon]. \end{aligned} \tag{2.6}$$

In terms of λ, (2.6) is minimized when $g(\lambda) = \alpha(e^{\lambda d} - \lambda d - 1) - \lambda n\varepsilon$ is minimized, which occurs at

$$\lambda^* = \frac{1}{d}\ln\left[\frac{n\varepsilon}{\alpha d} + 1\right]. \tag{2.7}$$

Substituting the value in (2.7) into (2.6) we get

$$\begin{aligned}
&\exp[\alpha(e^{\lambda^* d} - \lambda^* d - 1) - \lambda^* n\varepsilon] \\
&= \exp\left\{\alpha\left[\frac{n\varepsilon}{\alpha d} + 1 - \ln\left(\frac{n\varepsilon}{\alpha d} + 1\right) - 1\right] - \frac{n\varepsilon}{d}\ln\left(\frac{n\varepsilon}{\alpha d} + 1\right)\right\} \\
&= \exp\left\{-\left(\alpha + \frac{n\varepsilon}{d}\right)\ln\left(\frac{n\varepsilon}{\alpha d} + 1\right) + \frac{n\varepsilon}{d}\right\}
\end{aligned}$$

Substituting for α we get

$$\begin{aligned}
&= \exp\left\{-\left[n\frac{1-g_n}{12}\frac{J+1}{J-1} + \frac{n\varepsilon}{d}\right]\ln\left[\frac{12}{1-g_n}\frac{J-1}{J+1}\frac{\varepsilon}{d} + 1\right] + \frac{n\varepsilon}{d}\right\} \\
&= \exp\left\{-n\left[\left(\tfrac{d(1-g_n)(J+1)+12(J-1)\varepsilon}{12(J-1)d}\right)\ln\left(\tfrac{12(J-1)\varepsilon}{(1-g_n)(J+1)d} + 1\right) - \tfrac{\varepsilon}{d}\right]\right\}.
\end{aligned}$$

□

2.2 Main Result

We now give the main result of the paper in the following theorem.

Theorem 2.8 *For any $\varepsilon > 0$ and $\lambda > 0$ we have*

$$\begin{aligned}
R_n(\varepsilon) &:= P\left(\max_{n\le k\le N}\frac{S_k - k\mu}{k} \ge \varepsilon\right) \\
&\le \exp\left\{-n\left[\left(\tfrac{d(1-g_n)(J+1)+12(J-1)\varepsilon}{12(J-1)d}\right)\ln\left(\tfrac{12(J-1)\varepsilon}{(1-g_n)(J+1)d} + 1\right) - \tfrac{\varepsilon}{d}\right]\right\}
\end{aligned}$$

Proof. By Proposition 3.4, we see that

$$\begin{aligned}
R_n(\varepsilon) &= P\left(\max_{n\le k\le N}\frac{S_k - k\mu}{k} \ge \varepsilon\right) \\
&\le e^{-\lambda\varepsilon}E\left[\exp\left(\lambda\tfrac{S_n - n\mu}{\mu}\right)\right] \\
&= E\left[\exp\left(\tfrac{\lambda}{n}(S_n - n\mu - n\varepsilon)\right)\right] \\
&\le \exp\left\{-n\left[\left(\tfrac{d(1-g_n)(J+1)+12(J-1)\varepsilon}{12(J-1)d}\right)\ln\left(\tfrac{12(J-1)\varepsilon}{(1-g_n)(J+1)d} + 1\right) - \tfrac{\varepsilon}{d}\right]\right\},
\end{aligned}$$

because Theorem 2.7 holds for any $\lambda > 0$ (here $\frac{\lambda}{n}$). □

As per (1.3) we have

$$P_n(\varepsilon) = P(S_n - n\mu \ge n\varepsilon).$$

Noting that $P_n(\varepsilon) \le R_n(\varepsilon)$ and applying Theorem 2.8 we get the following corollary.

Corollary 2.9 *For any $\varepsilon > 0$ we have*

$$P_n(\varepsilon) \le \exp\left\{-n\left[\left(\tfrac{d(1-g_n)(J+1)+12(J-1)\varepsilon}{12(J-1)d}\right)\ln\left(\tfrac{12(J-1)\varepsilon}{(1-g_n)(J+1)d} + 1\right) - \tfrac{\varepsilon}{d}\right]\right\}.$$

From (1.6) we have

$$R_n^*(\varepsilon) = P\left[\max_{n\le k\le N}\left|\frac{S_k - k\mu}{k}\right| \ge \varepsilon\right].$$

Observe that

$$R_n^*(\varepsilon) = P\left[\max_{n\le k\le N}\frac{S_k - k\mu}{k} \ge \varepsilon\right] + P\left[\max_{n\le k\le N} -\frac{S_k - k\mu}{k} \ge \varepsilon\right].$$

Applying Theorem 2.8 to the above we get the following corollary.

Corollary 2.10 *For any $\varepsilon > 0$,*

$$R_n^*(\varepsilon) \le 2\exp\left\{-n\left[\left(\tfrac{d(1-g_n)(J+1)+12(J-1)\varepsilon}{12(J-1)d}\right)\ln\left(\tfrac{12(J-1)\varepsilon}{(1-g_n)(J+1)d} + 1\right) - \tfrac{\varepsilon}{d}\right]\right\}.$$

3. An Application

Corollaries 2.9 and 2.10 give parameter-free maximal inequalities which are shaper than Serfling's (1974) inequalities given in (1.6) and (1.7), respectively. Of course this fact is not surprising because we are utilizing the additional assumption that $X_1, X_2, \ldots, X_n$ belong to the set of probability distributions whose variance is bounded by the variance of the discrete uniform distribution as described in Definition 2.1. However, the extent of the improvement can be substantial as demonstrated by the following example.

Consider the following scenario, which is similar to an example presented by Savage (1961). Suppose one wishes to study the average height of a finite population of people. Assume that all individuals in the population are between 60 and 78 inches and their heights are measured to the nearest inch. The question we wish to answer is, "What is the probability that the average height of a sample of 100 from a population of 4,000 individuals is within two inches of the population mean height?"

One possible solution is to apply the Bienayme-Chebyshev inequality given in (1.5), where σ^2 is replaced by the maximum possible variance of distributions with support on $\Omega_{60,78,19}$, which for this problem is 81. This solution gives

$$P[|\bar{X} - \mu| \leq 2] \geq .8025. \tag{3.1}$$

A second possible solution to this example is to apply the probability inequality (1.6) given by Serfling (1974) and assumes only finite support of a discrete random variable. For our example, (1.6) yields

$$P[|\bar{X} - \mu| \leq 2] \geq .9205. \tag{3.2}$$

If we make the additional and, in this case, reasonable assumption that the probability functions of $X_1, X_2, \ldots X_n$ are from $V_{60,78,19}$, then we may apply Chebyshev's inequality with variance bound given in Definition 2.1. This method yields

$$P[|\bar{X} - \mu| \leq 2] \geq .9268. \tag{3.3}$$

Applying the newly-derived inequality given in Corollary 2.9, we get the following result

$$P[|\bar{X} - \mu| \leq 2] \geq .9936. \tag{3.4}$$

Clearly, inequality (3.4) not only yields a better bound than inequalities (3.1), (3.2), and (3.3), but, moreover, considerably increases the degree of improvement.

As an extension of the above example, let the amount of error, ε, between $\bar{X}$ and μ be arbitrary, but retain all other values in the example. Figure 1 demonstrates the sharpening of Serfling's (1974) inequality in (1.6) by the inequality given in Corollary 2.9 across values of ε from 0 to 3.

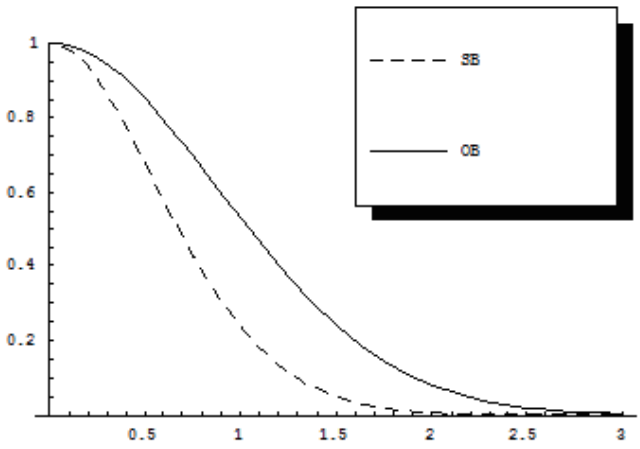

Figure 1. Comparison of our bound (OB) to Serfling's bound (SB) over value of ε from 0 to 3

Acknowledgements

The authors would like to thank the anynomous referee for the careful review and helpful comments pertaining to this article.

References

Dharmadhikari, S., & Joag-Dev, K. (1989). Upper bounds for the variances of certain random variables. *Communications in Statistics-Theory and Methods, 18*, 3235-3247.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications* (vol. 2). New York: Wiley.

Gray, H., & Odell, P. (1967). On least favorable density functions. *SIAM Review, 9*, 715-720. http://dx.doi.org/10.1137/1009112

Jacobson, H. (1969). The maximum variance of restricted unimodal distributions. *Annals of Statistics, 40*, 1746-1752.

Klaassen, C. (1985). On an inequality of chernoff. *Annals of Probabiliy, 13*, 966-974. http://dx.doi.org/10.1214/aop/1176992917

Moors, J., & Muilwijk, J. (1971). Note on a theorem of m. n. murthy and v. k. sethi. *Sankhyā, 33*, 385-388.

Rayner, J. (1975). Variance bounds. *Sankhyā, 37*, 135-138.

Savage, I. R. (1961). Probability inequalities of the tchebycheff type. *Journal of Research of the National Bureau of Standards-B, 65*, 211-222.

Seaman, J., Young, D., & Turner, D. (1987). On the variance of certain bounded random variables. *The Mathematical Scientists, 12*, 109-116.

Serfling, R. J. (1974). Probability inequalities for the sum in sampling without replacement. *The Annuals of Statistics, 2*, 39-48.

Appendix

Here we present some results that are mostly standard, but are complied here for ease of reference. We begin with a brief review of martingales including some results pertinent to this work.

1. Martingales

In the proofs of the results for this paper we need a few results about submartingales, in particular, reverse submartingales. We present the necessary results here for ease of reference. For a more indepth treatment of this subject see Feller (1966). For our purposes we will use the following definitions for *martingales*, *submartingales*, *reverse martingales*, and *reverse submartingales*.

Definition 1 Let $\{Z_k\}$ be a a sequence of random variables such that $E(|Z_k|) < \infty$. We say $\{Z_k\}$ is a *martingale* if $E[Z_k|Z_{k-1}, Z_{k-2}, \dots] = Z_{k-1}$ for all k. We say $\{Z_k\}$ is a *submartingale* if $E[Z_k|Z_{k-1}, Z_{k-2}, \dots] \geq Z_{k-1}$ for all k. We say $\{Z_k\}$ is a *reverse martingale* if $E[Z_k|Z_{k+1}, Z_{k+2}, \dots] = Z_{k+1}$ for all k. We say $\{Z_k\}$ is a *reverse submartingale* if $E[Z_k|Z_{k+1}, Z_{k+2}, \dots] \geq Z_{k+1}$ for all k.

We now present several results which exploit these properties.

Proposition 2 *Let $\{Z_n\}_{n=1}^N$ be a sequence of non-negative reverse submartingales, i.e.*

$$E[Z_n|Z_{n+1}, \dots, Z_N] \geq Z_{n+1}.$$

Then for any $c \geq 0$ we have

$$cP\left[\max_{n\leq k\leq N} Z_k \geq c\right] \leq E[Z_n].$$

Proof. Let $F = \{\max_{n\leq k\leq N} Z_k \geq c\}$. Then F can be expressed as the disjoint union of

$$\begin{aligned}
F_N &= \{Z_N \geq c\} \\
F_{N-1} &= \{Z_N < c\} \cap \{Z_{N-1} \geq c\} \\
F_{N-2} &= \{Z_N < c\} \cap \{Z_{N-1} < c\} \cap \{Z_{N-2} \geq c\} \\
&\vdots \\
F_n &= \{Z_N < c\} \cap \cdots \cap \{Z_{n+1} < c\} \cap \{Z_n \geq c\}
\end{aligned}$$

Now observe for each $k = n, \dots, N$,

$$\begin{aligned}
E[Z_n\,;\,F_k] &= E(E[Z_n|Z_{n+1}, \dots, Z_N]\,;\,F_k) \\
&\geq E[Z_k\,;\,F_k] \quad \text{since } \{Z_n\} \text{ is a reverse submartingale} \\
&\geq cP(F_k) \quad \text{since } Z_k \geq c \text{ on } F_k.
\end{aligned}$$

Summing over all k we get

$$\sum_{k=n}^{N} E[Z_n\,;\,F_k] \geq \sum_{k=n}^{N} cP(F_k)$$

or, equivalently,

$$E[Z_n\,;\,F] \geq cP(F) \quad \text{since } F = \bigcup_{k=n}^{N} F_k.$$

Therefore,

$$E[Z_n] \geq E[Z_n\,;\,F] \geq cP(F) = cP\Big(\max_{n\leq k\leq N} Z_k \geq c\Big),$$

which yields the desired result. □

Remark 3 If $\{Z_k\}$ is a reverse martingale, then applying Jensen's inequality immediately gives us that in$\{e^{\lambda Z_k}\}$ is a reverse submartingale for any $\lambda > 0$.

Proposition 4 *Let $\{Z_k\}_{k=1}^N$ be a reverse martingale, i.e.*

$$E[Z_k\,|\,Z_{k+1}, Z_{k+2}, \dots] = Z_{k+1}.$$

For any $\varepsilon > 0$ and $\lambda > 0$ we have

$$P\left(\max_{n\le k\le N} Z_k \ge \varepsilon\right) \le \frac{E(e^{\lambda Z_n})}{e^{\lambda\varepsilon}}.$$

Proof. By Remark 3, $\{e^{\lambda Z_k}\}$ is a reverse submartingale. Now using $c = e^{\lambda\varepsilon}$ in Proposition 2 we have

$$P\left(\max_{n\le k\le N} e^{\lambda Z_k} \ge e^{\lambda\varepsilon}\right) \le \frac{E(e^{\lambda Z_n})}{e^{\lambda\varepsilon}}. \tag{A.1}$$

Combining (A.1) with the fact that

$$P\left(\max_{n\le k\le N} e^{\lambda Z_k} \ge e^{\lambda\varepsilon}\right) = P\left(\max_{n\le k\le N} Z_k \ge \varepsilon\right),$$

we get the desired result. □

2. Finite Population Drawn Without Replacement

Here, we present some results which will aid us in our quest for sharpening of inequalities (1.6) and (1.7). These results were first used without proof in Serfling's paper (1974). We give proofs here for the sake of completeness. We work under the same set-up as presented in Section 1. That is $\Omega_N = \{x_1, x_2, \ldots, x_N\}$ is a finite population of size N, the members of which are not necessarily distinct. Also $X_1, X_2, \ldots, X_n$ denote the values of a sample of size n drawn without replacement from Ω_N and S_k is the sum of the first k samples, as in (1.2). We also take μ as in (1.1).

Proposition 5 *Let $\mu_k \equiv (X_k - \mu | X_1, X_2, \ldots, X_{k-1})$. Then,*

$$\mu_k = -\frac{S_{k-1} - (k-1)\mu}{N - (k-1)}. \tag{A.2}$$

Proof. For $k = 1, 2, \ldots, N-1$, let

$$T_k \equiv \frac{S_k - k\mu}{k} \quad \text{and} \quad T_k^* \equiv \frac{S_k - k\mu}{N-k}.$$

One can easily check that T_k is a reverse martingale and T_k^* are martingales Serfling (1974). That is,

$$E[T_k | T_{k+1}, \ldots, T_{N-1}] = T_{k+1}, \qquad 1 \le k \le N-2$$

and

$$E[T_k^* | T_{k-1}^*, \ldots, T_1^*] = T_{k-1}^*, \qquad 2 \le k \le N-1.$$

To prove (A.2) note that

$$\begin{aligned}
T_k^* &= E[T_{k+1}^* \,|\, T_k^*, \ldots T_1^*] \\
&= E\left[\frac{S_{k+1} - (k+1)\mu}{N-k-1} \,|\, T_k^*, \ldots T_1^*\right] \\
&= \frac{N-k}{N-k-1} E\left[\frac{S_{k+1} - (k+1)\mu}{N-k} \,|\, T_k^*, \ldots T_1^*\right] \\
&= \frac{N-k}{N-k-1} E\left[\frac{X_{k+1} - \mu + S_k - k\mu}{N-k} \,|\, T_k^*, \ldots T_1^*\right] \\
&= \frac{N-k}{N-k-1}\left(E\left[\frac{X_{k+1} - \mu}{N-k} \,|\, T_k^*, \ldots T_1^*\right] + E\left[\frac{S_k - k\mu}{N-k} \,|\, T_k^*, \ldots T_1^*\right]\right) \\
&= \frac{N-k}{N-k-1}\left(\frac{1}{N-k} E\left[X_{k+1} - \mu \,|\, X_k, \ldots X_1\right] + E\left[T_k^* \,|\, T_k^*, \ldots T_1^*\right]\right) \\
&= \frac{N-k}{N-k-1}\left(\frac{1}{N-k}\mu_{k+1} + T_k^*\right)
\end{aligned}$$

Thus,

$$T_k^* = \frac{\mu_k}{N-k-1} + \frac{(N-k)T_k^*}{N-k-1}. \tag{A.3}$$

Multiplying both sides of (A.3) by $N-k-1$, we get

$$(N-k-1)T_k^* = \mu_{k+1} + (N-k)T_k^*$$

or $T_k^* = -\mu_{k+1}$. Therefore (A.2) holds. □

In the next corollary utilizes a nonobvious recursive relationship between the S_k's.

Corollary 6 *For a fixed integer any $\lambda > 0$, let $\lambda_k = \frac{N-n}{N-k}$ for $1 \le k \le n$. Then*

$$\lambda_k(S_k - k\mu) = \lambda_{k-1}[S_{k-1} - (k-1)\mu] + \lambda_k(X_k - \mu - \mu_k) \tag{A.4}$$

for $k = 2, 3, \ldots, N$.

Proof. Using (A.2), we have that

$$\begin{aligned}
&\lambda_{k-1}[S_{k-1} - (k-1)\mu] + \lambda_k(X_k - \mu - \mu_k) \\
&= \lambda\frac{N-n}{N-(k-1)}\left[S_{k-1} - (k-1)\mu\right] + \lambda\frac{N-n}{N-k}\left[X_k - \mu + \frac{S_{k-1} - (k-1)\mu}{N-k+1}\right] \\
&= \lambda(N-n)\left[\frac{S_{k-1} - (k-1)\mu}{N-(k-1)} + \frac{X_k - \mu}{N-k} + \left(\frac{1}{N-k}\right)\frac{S_{k-1} - (k-1)\mu}{N-k+1}\right] \\
&= \lambda(N-n)\left[\left(\frac{N-k+1}{N-k}\right)\left(\frac{S_{k-1} - (k-1)\mu}{N-(k-1)}\right) + \frac{X_k - \mu}{N-k}\right] \\
&= \lambda\frac{N-n}{N-k}\left[S_{k-1} - (k-1)\mu + X_k - \mu\right] \\
&= \lambda_k[S_k - k\mu],
\end{aligned}$$

which yields (A.4). □