The Annals of Statistics
1974, Vol. 2, No. 1, 39-48

PROBABILITY INEQUALITIES FOR THE SUM IN SAMPLING WITHOUT REPLACEMENT[1]

BY R. J. SERFLING

The Florida State University

Upper bounds are established for the probability that, in sampling without replacement from a finite population, the sample sum exceeds its expected value by a specified amount. These are obtained as corollaries of two main results. Firstly, a useful upper bound is derived for the moment generating function of the sum, leading to an exponential probability inequality and related moment inequalities. Secondly, maximal inequalities are obtained, extending Kolmogorov's inequality and the Hájek-Rényi inequality.

Compared to sampling with replacement, the results incorporate sharpenings reflecting the influence of the sampling fraction, n/N, where n denotes the sample size and N the population size. We go somewhat beyond previous work by Hoeffding (1963) and Sen (1970). As in the latter reference, martingale techniques are exploited.

Applications to simple linear rank statistics are noted, dealing with the two-sample Wilcoxon statistic as an example. Finally, the question of sharpness of the exponential bounds is considered.

1. Introduction and key results. Consider sampling without replacement from a finite list of values $x_1, \cdots, x_N$ (not necessarily distinct), for example the weights of the individuals in some population, or the scores associated with some rank statistic. Denote by $X_1, \cdots, X_n$ the values of a sample of size n "drawn without replacement," i.e., $(X_1, \cdots, X_n) = (X_{I_1}, \cdots, X_{I_n})$, where

$$P[(I_1, \cdots, I_n) = (i_1, \cdots, i_n)] = 1/[N(N-1) \cdots (N-n+1)] \tag{1.1}$$

for each n-tuple $(i_1, \cdots, i_n)$ of distinct values from the set $\{1, \cdots, N\}$. Of fundamental interest in applications are the properties of the sum $S_n = \sum_1^n X_i$. It is desirable to specify its behavior as a function of the population parameters

$$a = \min_{1 \leq i \leq N} x_i, \qquad b = \max_{1 \leq i \leq N} x_i, \tag{1.2}$$
$$\mu = N^{-1} \sum_1^N x_i, \qquad \sigma^2 = N^{-1} \sum_1^N (x_i - \mu)^2$$

as well as of the sample size n and the "sampling fraction" $f_n = (n-1)/(N-1)$. In some instances the notation $f_n^* = (n-1)/N$ will be useful.

Of direct interest in various contexts are the probabilities

$$P_n(t) = P[S_n - n\mu \geqq nt],$$
$$Q_n(t) = P[\max_{1 \leq k \leq n} |S_k - k\mu| \geqq nt],$$

Received September 1972; revised February 1973.

[1] Research supported by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-67-A-0235-0006.

AMS 1970 *subject classifications.* Primary 62; Secondary 70.

Key words and phrases. Probability inequalities, sampling without replacement, Kolmogorov inequality, Hájek-Rényi inequality, moment generating function, rank statistics.

$$R_n(t) = P\left[\max_{1 \leqq k \leqq n} \left|\frac{S_k - k\mu}{k}\right| \geqq t\right],$$

and

$$R_n^*(t) = P\left[\max_{n \leqq k \leqq N} \left|\frac{S_k - k\mu}{k}\right| \geqq t\right],$$

for values of $t > 0$. The roles of $P_n(t)$ are manifold and widely known. The quantity $Q_n(t)$ is basic, for example, in showing weak convergence of certain stochastic processes associated with rank statistics of Kolmogorov–Smirnov type, as in Hájek and Šidák (1967), pages 184–186. The quantities $R_n(t)$ and $R_n^*(t)$ are clearly relevant to the study of the strong law of large numbers for the sample mean, S_n/n. Further discussion of applications appears in Sections 3 and 4.

The chief aim of this paper is to give useful exact upper bounds for $P_n(t)$, $Q_n(t)$, $R_n(t)$ and $R_n^*(t)$. From somewhat more general results developed in Section 2, we have the following consequences.

COROLLARY 1.1. *For* $t > 0$,

$$P_n(t) \leqq \exp[-2nt^2/(1 - f_n^*)(b - a)^2]. \tag{1.3}$$

The bound (1.3) for $P_n(t)$ is simple, converges exponentially to zero with increase in nt^2, and requires only the parameter $(b - a)$ to be given. If f_n^* is replaced by 0, (1.3) reduces to a probability inequality derived by Hoeffding (1963) for *independent* X_i's and shown by him to hold also in the context of sampling without replacement. We thus achieve in (1.3) a sharpening with increase in f_n^*. (The anticipation of such an improvement, analogous to the effect of f_n upon the variance of S_n, motivated this investigation.) In Section 3 the bound (1.3) is utilized to derive simple but effective moment inequalities of all orders for $(S_n - n\mu)$.

We now turn to bounds emphasizing the parameter σ^2 and we bring $Q_n(t)$, $R_n(t)$ and $R_n^*(t)$ into consideration. Of course, for $P_n(t)$ useful bounds are due to Chebychev (see [8] for discussion). Namely,

$$P_n(t) \leqq \frac{1}{1 + \dfrac{nt^2}{(1 - f_n)\sigma^2}} \leqq \frac{(1 - f_n)\sigma^2}{nt^2}. \tag{1.4}$$

Because (1.3) is sufficiently sensitive to the sampling fraction f_n^*, it enjoys a favorable comparison with the Chebychev result. In the same vein as (1.4), we have

COROLLARY 1.2. *For* $t > 0$,

$$Q_n(t) \leqq \frac{N\sigma^2}{(N - 1)n^2t^2}\left[\sum_{k=1}^{n} \frac{N - k}{N - k + 1}\right], \tag{1.5a}$$

$$Q_n(t) \leqq \frac{N\sigma^2}{(N - 1)n^2t^2}\left[\sum_{k=1}^{n-1} \frac{k}{k + 1} + \frac{n(N - n)}{N}\right], \tag{1.5b}$$

$$R_n(t) \leqq \frac{\sigma^2}{t^2}, \tag{1.6}$$

and

$$R_n^*(t) \leqq \frac{(1 - f_n)\sigma^2}{nt^2} . \tag{1.7}$$

The relations (1.5a, b) are complementary extensions of Kolmogorov's inequality to the context of sampling without replacement. Inequality (1.5b) has been derived by Sen (1970). Note that (1.7) implies the right-most inequality in (1.4).

The bound (1.5b) for $Q_n(t)$ was noted by Sen (1970) to be inadequate for the purposes of Hájek and Šidák (1967), who achieved their goal by proving

$$Q_n(t) \leqq [(1 - f_n)nt]^{-4} E(S_n - n\mu)^4 \tag{1.8}$$

and utilizing the exact expression for $E(S_n - n\mu)^4$ given by Isserlis (1931). In the spirit of (1.8), we state

COROLLARY 1.3. *For any positive integer r, and $t > 0$,*

$$Q_n(t) \leqq [(1 - f_n)nt]^{-2r} E(S_n - n\mu)^{2r} , \tag{1.9}$$

$$R_n(t) \leqq t^{-2r} E(X_1 - \mu)^{2r} , \tag{1.10}$$

and

$$R_n^*(t) \leqq [nt]^{-2r} E(S_n - n\mu)^{2r} . \tag{1.11}$$

For $r = 1$, (1.10) and (1.11) reduce to (1.6) and (1.7), respectively, in view of

$$E(S_n - n\mu)^2 = (1 - f_n)n\sigma^2 , \tag{1.12}$$

but (1.9) reduces to

$$Q_n(t) \leqq \frac{\sigma^2}{(1 - f_n)nt^2} , \tag{1.13}$$

which differs substantially from (1.5a, b), but likewise is inadequate for the purposes of [4].

The general results presented in Section 2 are obtained by exploiting the forward martingale structure of the sequence $(S_k - k\mu)/(N - k)$, $1 \leqq k < N$, and the reverse martingale structure of the sequence $(S_k - k\mu)/k$, $1 \leqq k \leqq N$. Upper bounds are established for the probabilities

$$U_n(t) = P\left[\max_{1\leqq k \leqq n} \left(\frac{S_k - k\mu}{N - k}\right) \geqq \frac{nt}{N - n}\right],$$

$$V_{i,j}(t; c_1, \cdots, c_j) = P\left[\max_{i \leqq k \leqq j} c_k \left|\frac{S_k - k\mu}{N - k}\right| \geqq t\right],$$

where $1 \leqq i \leqq j < N$ and $c_i \geqq c_{i+1} \geqq \cdots \geqq c_j \geqq 0$, and

$$W_{i,j}(t; d_i, \cdots, d_j) = P\left[\max_{i \leqq k \leqq j} d_k \left|\frac{S_k - k\mu}{k}\right| \geqq t\right],$$

where $1 \leqq i \leqq j \leqq N$ and $0 \leqq d_i \leqq d_{i+1} \leqq \cdots \leqq d_j$. From these results we obtain Corollaries 1.1, 1.2 and 1.3 via the elementary relations

$$P_n(t) \leqq U_n(t), \tag{1.14}$$

$$Q_n(t) = V_{1,n}(nt; N-1, N-2, \cdots, N-n) = W_{1,n}(nt; 1, 2, \cdots, n), \tag{1.15}$$

$$Q_n(t) \leqq V_{1,n}\left(\frac{nt}{N-1}; 1, \cdots, 1\right), \tag{1.16}$$

$$R_n(t) = V_{1,n}\left(t; \frac{N-1}{1}, \frac{N-2}{2}, \cdots, \frac{N-n}{n}\right) = W_{1,n}(t; 1, \cdots, 1), \tag{1.17}$$

and

$$\begin{aligned} R_n{}^*(t) &= V_{n,N-1}\left(t; \frac{N-n}{n}, \frac{N-n-1}{n+1}, \cdots, \frac{1}{N-1}\right) \\ &= W_{n,N}(t; 1, \cdots, 1). \end{aligned} \tag{1.18}$$

The specific results of Section 2 are characterized as follows. Theorem 2.1 provides a class of probability inequalities for $U_n(t)$. Theorem 2.2 gives a bound on the moment generating function of $(S_n - n\mu)$. The two results together yield exponential probability inequalities for $U_n(t)$ and hence, by (1.14), for $P_n(t)$. Finally, Theorem 2.2 gives a generalized version of the Hájek–Rényi inequality in the context of sampling without replacement.

As mentioned already, Sections 3 and 4 deal with applications. Concluding this paper, sharpness considerations and open questions are discussed in Section 5.

2. General results. We first present four lemmas, each of which places an upper bound on a quantity of interest.

LEMMA 2.1. *For integers* $1 \leqq k \leqq m$,

$$\textstyle\sum_{j=k+1}^{m} j^{-2} \leqq (m-k)/k(m+1). \tag{2.1}$$

PROOF. It is easily seen that the left hand side of (2.1) is less than $(k+\frac{1}{2})^{-1} - (m+\frac{1}{2})^{-1}$, which in turn is $\leqq (m-k)/k(m+1)$. □

The next two lemmas involve the function

$$f(x, y) = \frac{x}{x+y} e^{-y} + \frac{y}{x+y} e^{x}, \qquad x > 0, y > 0.$$

LEMMA 2.2.

$$f(x, y) \leqq \exp[\tfrac{1}{8}(x+y)^2]. \tag{2.2}$$

PROOF. In the proof of Theorem 2 of Hoeffding (1963), it is shown that

$$qe^{-zp} + pe^{zq} \leqq \exp[\tfrac{1}{8}z^2]$$

for $0 < p < 1$, $q = 1 - p$ and $z > 0$. Putting $p = y/(x+y)$ and $z = (x+y)$, we get (2.2). □

The next lemma is due to Bennett (1962), page 42.

LEMMA 2.3. *Let Z be a random variable satisfying $P[Z \leqq B] = 1$ for a finite constant B and having mean m and variance v. Then, for $h > 0$,*

$$E[e^{h(Z-m)}] \leqq f(h(B-m), hv/(B-m)). \tag{2.3}$$

The preceding two lemmas will be utilized through

LEMMA 2.4. *Let Z be a random variable satisfying $P[A \leqq Z \leqq B] = 1$ for finite constants $A \leqq B$ and having mean m. Then, for $h > 0$,*

$$E[e^{h(Z-m)}] \leqq \exp[\tfrac{1}{8}h^2(B-A)^2]\,. \tag{2.4}$$

PROOF. Let Z have variance v. By (2.2) and (2.3), for $h > 0$,

$$E[e^{h(Z-m)}] \leqq \exp\{\tfrac{1}{8}h^2[(B-m) + v/(B-m)]^2\}\,. \tag{2.5}$$

Now, as pointed out by Hoeffding (1963), $v = E(Z-m)^2 = E(Z-m)(Z-A) \leqq (B-m)E(Z-A) = (B-m)(m-A)$. Thus the right hand side of (2.4) exceeds that of (2.5).

Our final preliminary is to take note of the martingale structures inherent in the scheme of sampling without replacement. Define

$$T_k = \frac{S_k - k\mu}{k}\,, \qquad T_k^* = \frac{S_k - k\mu}{N-k}$$

for $1 \leqq k \leqq N$. It is easily checked that

$$E[T_k \mid T_{k+1}, \cdots, T_{N-1}] = T_{k+1}\,, \qquad 1 \leqq k \leqq N-2\,, \tag{2.6}$$

and

$$E[T_k^* \mid T_{k-1}^*, \cdots, T_1^*] = T_{k-1}^*\,, \qquad 2 \leqq k \leqq N-1\,, \tag{2.7}$$

i.e., the sequence $T_1, T_2, \cdots, T_{N-1}$ is a *reverse* martingale and the sequence $T_1^*, T_2^*, \cdots, T_{N-1}^*$ a *forward* martingale.

THEOREM 2.1. *Let $u(x)$ be convex and nonnegative on $-\infty < x < \infty$ and non-decreasing and positive on $0 < x < \infty$. Then, for $t > 0$,*

$$U_n(t) \leqq \frac{E\left[u\left(\frac{S_n - n\mu}{N-n}\right)\right]}{u\left(\frac{nt}{N-n}\right)}\,. \tag{2.8}$$

PROOF. Since $u(x)$ is non-decreasing on $0 < x < \infty$,

$$P[\max\{T_1^*, \cdots, T_n^*\} \geqq x] \leqq P[\max\{u(T_1^*), \cdots, u(T_n^*)\} \geqq u(x)] \tag{2.9}$$

for $x > 0$. Since u is convex and $\{T_k^*\}$ is a (forward) martingale, the sequence $\{u(T_k^*)\}$ is a submartingale (cf. Feller (1966), page 215). Since $\{u(T_k^*)\}$ is thus a nonnegative submartingale and $u(nt/(N-n)) > 0$, we obtain (2.8) from (2.9) and Kolmogorov's inequality ([3], page 235). □

REMARK. By a similar argument follows

$$P\left[\max_{1\leqq k\leqq n}\left|\frac{S_k - k\mu}{N-k}\right| \geqq \frac{nt}{N-n}\right] \leqq \frac{E\left[u\left(\left|\frac{S_n - n\mu}{N-n}\right|\right)\right]}{u\left(\frac{nt}{N-n}\right)}\,. \tag{2.10}$$

Such a result is of some utility, but Theorem 2.3 appears to yield more fruit, so we have not emphasized (2.10).

In order to make use of (2.8) in conjunction with the function $u(x) = \exp(hx)$, where $h > 0$, we prove

THEOREM 2.2. *For* $h > 0$,

$$E\{\exp[h(S_n - n\mu)]\} \leqq \exp[\tfrac{1}{8}h^2(1 - f_n^*)n(b-a)^2]\,. \tag{2.11}$$

PROOF. Let $h > 0$ and write

$$h_k = \frac{N-n}{N-k}h\,, \qquad 1 \leqq k \leqq n\,. \tag{2.12}$$

Denote by μ_k the conditional expectation of $(X_k - \mu)$, given $X_1, \cdots, X_{k-1}$. As per (2.7),

$$\mu_k = -\frac{S_{k-1} - (k-1)\mu}{N-k+1}\,, \qquad 2 \leqq k \leqq n\,. \tag{2.13}$$

We thus have

$$\begin{aligned} h_k(S_k - k\mu) &= h_k\left(\frac{N-k}{N-k+1}\right)[S_{k-1} - (k-1)\mu] + h_k(X_k - \mu - \mu_k) \\ &= h_{k-1}[S_{k-1} - (k-1)\mu] + h_k(X_k - \mu - \mu_k)\,. \end{aligned} \tag{2.14}$$

Since $a - \mu \leqq X_k - \mu \leqq b - \mu$, we have by Lemma 2.4 that

$$E\{\exp[h_k(X_k - \mu - \mu_k)] \mid X_1, \cdots, X_{k-1}\} \leqq \exp[\tfrac{1}{8}h_k^2(b-a)^2]\,, \qquad 2 \leqq k \leqq n\,. \tag{2.15}$$

Therefore, using (2.14) and (2.15), for $2 \leqq k \leqq n$,

$$E\{\exp[h_k(S_k - k\mu)]\} \leqq \exp[\tfrac{1}{8}h_k^2(b-a)^2]E\{\exp[h_{k-1}[S_{k-1} - (k-1)\mu]]\}\,. \tag{2.16}$$

Also, again by Lemma 2.4, we have

$$E\{\exp[h_1(X_1 - \mu)]\} \leqq \exp[\tfrac{1}{8}h_1^2(b-a)^2]\,. \tag{2.17}$$

It follows from (2.16), taken for $2 \leqq k \leqq n$, and (2.17) that

$$E\{\exp[h(S_n - n\mu)]\} \leqq \exp[\tfrac{1}{8}h^2\Delta_n(b-a)^2]\,, \tag{2.18}$$

where

$$\Delta_n = \textstyle\sum_{k=1}^{n}\left(\dfrac{N-n}{N-k}\right)^2 = 1 + (N-n)^2 \sum_{k=N-n+1}^{N-1} k^{-2}\,. \tag{2.19}$$

Applying Lemma 2.1, we see that

$$\Delta_n \leqq 1 + (N-n)^2(n-1)/N(N-n) = n\left(1 - \frac{n-1}{N}\right) = n(1 - f_n^*)\,.$$

Thus (2.11) follows. □

Together, Theorems 2.1 and 2.2 yield Corollary 1.1, as will now be shown.

PROOF OF COROLLARY 1.1. Let $h > 0$. By (2.8) with $u(x) = \exp(hx)$ and (2.11), we have

$$U_n(t) \leqq \exp\left[-\frac{hnt}{N-n} + \frac{1}{8}\left(\frac{h}{N-n}\right)^2 n(1 - f_n^*)(b-a)^2\right]. \tag{2.20}$$

The right hand side of (2.20) is minimized when $h = 4t(N-n)/(1-f_n^*)(b-a)^2$. With the use of relation (1.14), this gives (1.3). □

We might also consider applications of Theorem 1.1, or likewise (2.10), in connection with the function $u(x) = x^{2r}$ for a positive integer r. However, superior results flow from the following theorem, our final main result.

THEOREM 2.3. *For any positive integer r, and $t > 0$,*

$$(2.21)\qquad V_{i,j}(t; c_i, \cdots, c_j) \leqq t^{-2r}\left[\sum_{k=i}^{j-1}(c_k^{2r} - c_{k+1}^{2r})E\left(\frac{S_k - k\mu}{N-k}\right)^{2r} + c_j^{2r}E\left(\frac{S_j - j\mu}{N-j}\right)^{2r}\right]$$

for $1 \leqq i \leqq j < N$ and $c_i \geqq c_{i+1} \geqq \cdots \geqq c_j \geqq 0$, and

$$(2.22)\qquad W_{i,j}(t; c_i, \cdots, c_j) \leqq t^{-2r}\left[d_i^{2r}E\left(\frac{S_i - i\mu}{i}\right)^{2r} + \sum_{k=i+1}^{j}(d_k^{2r} - d_{k-1}^{2r})E\left(\frac{S_k - k\mu}{k}\right)^{2r}\right]$$

for $1 \leqq i \leqq j \leqq N$ and $0 \leqq d_i \leqq d_{i+1} \leqq d_j$.

PROOF. As noted earlier, the sequence $\{T_k^*\}$ is a martingale and thus $\{(T_k^*)^{2r}\}$ is a submartingale ([3], page 215). A direct application of Theorem 1 of Chow (1960) yields (2.21). By a similar argument applied to the reverse martingale $\{T_k\}$, we obtain (2.22). □

In the case $r = 1$, formula (2.22) was given by Sen (1970) as an extension of the Hájek–Rényi (1955) inequality to the situation of sampling without replacement. We note that (2.21) offers an alternative extension.

The implications of (2.21) and (2.22) for the case $r = 1$ assume relatively simple forms in terms of σ^2. With the use of (1.12) we find, under the restrictions of the theorem,

$$(2.23\text{a})\qquad V_{i,j}(t; c_i, \cdots, c_j) \leqq \frac{\sigma^2}{(N-1)t^2}\left[\sum_{k=i}^{j-1}(c_k^2 - c_{k+1}^2)\frac{k}{N-k} + c_j^2\frac{j}{N-j}\right]$$

$$(2.23\text{b})\qquad = \frac{N\sigma^2}{(N-1)t^2}\left[c_i^2\frac{i}{N(N-i)} + \sum_{k=i+1}^{j}\frac{c_k^2}{(N-k)(N-k+1)}\right]$$

and

$$(2.24\text{a})\qquad W_{i,j}(t; c_i, \cdots, c_j) \leqq \frac{\sigma^2}{(N-1)t^2}\left[d_i^2\frac{(N-i)}{i} + \sum_{k=i+1}^{j}(d_k^2 - d_{k-1}^2)\frac{N-k}{k}\right]$$

$$(2.24\text{b})\qquad = \frac{N\sigma^2}{(N-1)t^2}\left[\sum_{k=i}^{j-1}\frac{d_k^2}{k(k+1)} + d_j^2\frac{N-j}{j}\right].$$

PROOF OF COROLLARIES 1.2 AND 1.3. Using (1.15) with (2.23b) and (2.24b), we obtain (1.5a) and (1.5b). Using (1.17) and (1.18) with (2.24a), we obtain (1.6) and (1.7). Using (1.16) with (2.21) we obtain (1.9). Using (1.17) and (1.18) with (2.22), we obtain (1.10) and (1.11). □

3. Moment inequalities for the sum. The exponential probability inequality of Corollary 1.1 yields simple but powerful moment inequalities of all orders:

THEOREM 3.1. *For* $\nu > 0$,

$$E[|S_n - n\mu|^\nu] \leqq \frac{\Gamma(\frac{1}{2}\nu + 1)}{2^{\frac{1}{2}\nu+1}} [(1 - f_n^*)n(b - a)^2]^{\frac{1}{2}\nu} . \tag{3.1}$$

PROOF. By a well-known formula ([3], page 148),

$$E[|S_n - n\mu|^\nu] = \int_0^\infty P[|S_n - n\mu|^\nu > t]\, dt . \tag{3.2}$$

From Corollary 1.1 it follows easily that

$$P[|S_n - n\mu|^\nu > t] \leqq 2 \exp[-2t^{2/\nu}/n(1 - f_n^*)(b - a)^2] . \tag{3.3}$$

Inserting (3.3) in (3.2) and integrating, we obtain (3.1). □

4. Other applications. Briefly we augment the applications mentioned in Sections 1 and 3.

(i) *Confidence intervals for* μ. A bound on $P_n(t)$ may be utilized in the usual way to attach a (conservative) confidence coefficient to an interval of the form $(\bar{X}_n - L_1, \bar{X}_n + L_2)$, where $\bar{X}_n = S_n/n$, and L_1 and L_2 are positive constants. Or, utilizing a bound on $R_n(t)$, a somewhat more sophisticated confidence interval procedure can be developed.

(ii) *Optional stopping in sequential sampling.* A bound on $R_n^*(t)$ is relevant in establishing confidence coefficients in the case of sequential sampling terminated in a completely optional way.

(iii) *Large deviations of simple linear rank statistics.* As an example, let us consider the two-sample Wilcoxon statistic, which may be represented as the sum of ranks of the first sample among the combined observations, i.e., $W = \sum_1^n X_i$, where the first sample is of size n and the second of size $N - n$, and $X_1, \cdots, X_n$ are a sample without replacement from $\{1, 2, \cdots, N\}$. The large deviation index of this statistic, i.e., the value $I = I(\gamma, \lambda)$ for which

$$-\frac{\ln P[W - E(W) \geqq \gamma N^2]}{N} \to I \tag{4.1}$$

as $n \to \infty$, $N \to \infty$ such that $n/N \to \lambda$, $0 < \lambda < 1$, has been determined independently by Hoadley (1967) and Stone (1967). It is a complicated function of γ and λ. On the other hand, using (1.3) with $a = 1$ and $b = N$, we derive the simple inequality

$$-\frac{\ln P[W - E(W) \geqq \gamma N^2]}{N} \geqq \frac{2\gamma^2}{\lambda(1 - \lambda + 1/N)} , \tag{4.2}$$

where $\lambda = n/N$. Whereas (4.1) gives an approximation valid for large n and N, (4.2) asserts a relation holding exactly for all n, N, though not an approximation. As a numerical example, let $\lambda = .5$ and $\gamma = .05$ and assume $1/N$ negligible. Then the right hand side of (4.2) is .02, whereas the limiting value of (4.1) is .064. The latter value is obtained from Figure 1 of Stone (1967). (His ρ corresponds to $\gamma + \frac{1}{2}\lambda$ in my notation.) Considering the crudeness of (1.3), this appears quite satisfactory. Given a sharper version of Corollary 1.1 involving the parameter σ instead of $b - a$, as discussed in the next section, perhaps very good agreement between the approximation and the lower bound would occur in some examples.

5. Sharpness considerations and open questions. Let $C_0 = C_0(\sigma^2, b - \mu, \mu - a)$ be a constant depending only on σ^2, $(b - \mu)$, $(\mu - a)$ and such that

$$P_n(t) \leqq \exp[-C_0 n t^2/(1 - f_n^*)], \qquad \text{all } t > 0, \text{ all } n. \tag{5.1}$$

Let $C_1 = C_1(b - \mu, \mu - a)$ denote the inf of C_0 as σ^2 varies while $b - \mu$ and $\mu - a$ remain fixed, and let $C_2 = C_2(b - a)$ denote the inf of C_1 as $b - \mu$ and $\mu - a$ vary while $b - a$ remains fixed.

LEMMA 5.1.

$$C_0 \leqq 1/2\sigma^2; \tag{5.2}$$

$$C_1 \leqq 1/2(b - \mu)(\mu - a); \tag{5.3}$$

$$C_2 \leqq 2/(b - a)^2. \tag{5.4}$$

PROOF. (5.3) and (5.4) follow from (5.2) with the use of

$$\sigma^2 \leqq (b - \mu)(\mu - a) \leqq (b - a)^2/4, \tag{5.5}$$

wherein the first inequality was seen in the proof of Lemma 2.4 and the second is well known. It remains to prove (5.2). Here a technique of Kemperman (1972) shall be used. For fixed σ^2, $(b - \mu)$, $(\mu - a)$, consider a sequence of populations and samples with $N = N_k \to \infty$, $n = n_k \to \infty$, $n_k/N_k \to \gamma (0 < \gamma < 1)$ and $\sigma_k^2 \to \sigma^2 > 0$, as $k \to \infty$. Fix t and put $t_k = t\sigma_k[1 - (n_k - 1)/N_k]^{\frac{1}{2}} n_k^{-\frac{1}{2}}$. Then (5.1) implies that $P_{n_k}(t) \leqq \exp(-C_0\sigma^2 t^2)$, i.e.,

$$t^{-2} \ln P_{n_k}(t_k) \leqq -C_0\sigma_k^2 \to -C_0\sigma^2, \qquad k \to \infty. \tag{5.6}$$

On the other hand, Hájek (1960) has proved a central limit theorem for sampling without replacement, which gives $P_{n_k}(t_k) \to 1 - \Phi(t) = (2\pi)^{-\frac{1}{2}} \int_t^\infty \exp(-\frac{1}{2}u^2)\, du$, $k \to \infty$. Now let $\varepsilon > 0$ be given. Since $\ln[1 - \Phi(t)] \sim (-\frac{1}{2}t^2)$, $t \to \infty$, choose and fix t large enough that $t^{-2} \ln[1 - \Phi(t)] > -\frac{1}{2} - \varepsilon$. Then, for this value of t and for k sufficiently large,

$$t^{-2} \ln P_{n_k}(t_k) > -\tfrac{1}{2} - \varepsilon. \tag{5.7}$$

Combining (5.6) and (5.7) we have (5.2). □

It follows from (5.4) that the constant in the bound of Corollary 1.1 is the best that can be asserted with knowledge only of the parameter $b - a$. It would be desirable to obtain a sharpening of this result involving the quantity σ^2 in place of the quantity $(b - a)^2/4$. Such a result would be sharper than Chebychev's inequality as well as more useful in applications like 4(iii). It also is of interest to obtain Corollary 1.1 with the usual sampling fraction f_n instead of f_n^*.

6. Acknowledgment. The author is very grateful to Wassily Hoeffding and J. H. B. Kemperman for suggesting corrections and improvements to an earlier version of this paper.

REFERENCES

[1] Bennett, George (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[2] Chow, Y. S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* **11** 107–111.

[3] Feller, William (1966). *An Introduction to Probability Theory and Its Applications*, **2**. Wiley, New York.

[4] Hájek, J. and Rényi, A. (1955). Generalization of an inequality of Kolmogorov. *Acta. Math. Acad. Sci. Hungar.* **6** 282–283.

[5] Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 361–374.

[6] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

[7] Hoadley, A. Bruce (1967). On the probability of large deviations of functions of several empirical cdf's. *Ann. Math. Statist.* **38** 360–381.

[8] Hoeffding, Wassily (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[9] Isserlis, L. (1931). On the moment distributions of moments in the case of samples drawn from a limited universe. *Proc. Roy. Soc. Ser. A* **132** 586–604.

[10] Kemperman, J. H. B. (1972). Moment problems for sampling without replacement. Unpublished.

[11] Sen, Pranab Kumar (1970). The Hájek-Rényi inequality for sampling from a finite population. *Sankhyā Ser. A* **32** 181–188.

[12] Stone, M. (1967). Extreme tail probabilities for the null distribution of the two-sample Wilcoxon statistic. *Biometrika* **54** 629–640.

Department of Statistics
Florida State University
Tallahassee, Florida 32306