SAMPLING DISTINGUISHABLE ELEMENTS WITH REPLACEMENT

BY JAN LANKE

University of Lund

In sampling with replacement from a finite population, the sample mean is known to get a smaller variance if repetitions are deleted before forming the mean. The asymptotic behavior of the variance just mentioned is studied.

0. Introduction and summary. When estimating a population mean using simple random sampling with replacement, one had better average over the distinct units in the sample if this is possible, i.e. if the elements are distinguishable. This fact was noted by Basu (1958) and Raj and Khamis (1958). A simple expression for the variance of the estimator thus obtained was given by Pathak (1961), independently by Thionet (1967) and, still independently, by Korwar and Serfling (1970). Inequalities for the variance were given by Korwar and Serfling (1970) and, independently, by Valemois (1971). In this note we give inequalities from above and from below which for sufficiently large population sizes are better than the previous ones; in particular our inequalities are “asymptotically correct.”

1. Results. Let $u$ denote the number of different elements obtained when performing simple random sampling of size $n$ with replacement from a population of size $N$ and let $\bar{z}$ denote the arithmetic mean of the values of the $u$ different elements. Then (see [3], [5] or [2])

$$\begin{align*}
V(\bar{z}) &= S^2 \left( \frac{1}{u} \frac{1}{N} - \frac{1}{N} \right) = \frac{S^2}{N} \sum_{k=1}^{N-1} \left( \frac{k}{N} \right)^{u-1}
\end{align*}$$

where $S^2$ is the population variance (defined with the factor $N - 1$ in the denominator). The inequalities in [2] may be written

$$\begin{align*}
1 - \frac{361}{720} f + \frac{1}{12} f^2 - \frac{1}{12} \cdot \frac{f}{N} \leq V(\bar{z}) / S^2/n \leq 1 - \frac{1}{2} f + \frac{1}{12} f^2 - \frac{1}{12} \cdot \frac{f}{N}
\end{align*}$$

where $f = n/N$. If $n \to \infty$, $N \to \infty$, $n/N \to f_0$ they obviously give

$$\begin{align*}
1 - \frac{361}{720} f_0 + \frac{1}{12} f_0^2 \leq \lim \inf \frac{V(\bar{z})}{S^2/n} \leq \lim \sup \frac{V(\bar{z})}{S^2/n} \leq 1 - \frac{1}{2} f_0 + \frac{1}{12} f_0^2.
\end{align*}$$

In this note we will show that

$$\begin{align*}
\lim \frac{V(\bar{z})}{S^2/n} &= \frac{f_0}{e^{f_0} - 1}
\end{align*}$$

and furthermore that for finite $n$ and $N$ with $3 \leq n \leq N$ we have

Received June 3, 1971; revised December, 6, 1971.

1329
\[
\frac{f}{e^f - 1} - \frac{1}{10} \cdot \frac{f}{N} \leq \frac{V(z)}{S^2/n} \leq \frac{f}{e^f - 1} - \frac{1}{15} \cdot \frac{f}{N}
\]

where \( f = n/N \).

2. **Proofs.** Using the notation

\[
g_k(n, N) = (\max(0, 1 - k/N))^{n-1}
\]

(1) can be written

\[
\frac{V(z)}{S^2/n} = \frac{n}{N} \sum_{k=1}^{\infty} g_k(n, N).
\]

If \( n \to \infty, N \to \infty, n/N \to f_0 \), we shall eventually have \((n-1)/N > f_0/2\) whence

\[
0 \leq g_k(n, N) \leq e^{-(n-1)/2} \leq Ce^{-k/f_0^2}.
\]

Since

\[
g_k(n, N) \to e^{-kf_0}
\]

dominated convergence shows that provided \( f_0 > 0 \) we have

\[
\frac{n}{N} \sum_{k=1}^{\infty} g_k(n, N) \to f_0 \sum_{k=1}^{\infty} e^{-kf_0}
\]

and (3) is proved for \( f_0 > 0 \). If \( f_0 = 0 \), (3) follows from (2).

To prove (4), let \( B_n(t) \) denote the Bernoulli polynomial of order \( n \). Then we have

\[
\sum_{k=1}^{n-1} k^{n-1} = \frac{1}{n} (B_n(N) - B_n)
\]

where \( B_n = B_n(0) \) is a Bernoulli number. Thus (1) gives

\[
\frac{V(z)}{S^2/n} = N^{-n}(B_n(N) - B_n)
\]

and by the expansion

\[
B_n(t) = \sum_{k=0}^{n} \binom{n}{k} B_k t^{n-k}
\]

we get

\[
\frac{V(z)}{S^2/n} = N^{-n} \sum_{k=0}^{n-1} \binom{n}{k} B_k N^{n-k} = \sum_{k=0}^{n-1} \frac{1}{k!} B_k n^{(k)} N^{-k}.
\]

The desired inequality (4) can be written

\[
-\frac{1}{10} \cdot \frac{f}{N} \leq \Delta \leq -\frac{1}{15} \cdot \frac{f}{N}
\]

where

\[
\Delta = \frac{V(z)}{S^2/n} - \frac{f}{e^f - 1}.
\]
Now the expansion

\[ \frac{f}{e^f - 1} = \sum_{k=0}^{\infty} \frac{1}{k!} B_k f^k \]

is valid (since \(0 < f \leq 1\) whence in particular \(|f| < 2\pi\)) and thus (5) and (7) give

\[ \Delta = \sum_{k=0}^{n-1} \frac{1}{k!} B_k n^{(k)} N^{-k} - \sum_{k=0}^{\infty} \frac{1}{k!} B_k n^k N^{-k}. \]

Since

for \(k = 0, 1:\)
\[ n^{(k)} = n^k \]

for \(k = 2:\)
\[ n^{(k)} - n^k = -n \quad \text{and} \quad B_2 = \frac{1}{6} \]

for \(k = 3:\)
\[ B_3 = 0 \]

for \(4 \leq k < n:\)
\[ 0 \leq n^k - n^{(k)} \leq \binom{k}{3} n^{k-1} \quad \text{(induction on } k) \]

for \(k \geq n:\)
\[ n^k \leq \binom{k}{3} n^{k-1} \quad \text{if} \quad n \geq 3 \]

(8) gives

\[ \left| \Delta + \frac{1}{12} \cdot \frac{f}{N} \right| \leq \sum_{k=4}^{\infty} \frac{1}{k!} |B_k| \binom{k}{3} n^{k-1} N^{-k} \]
\[ \leq \frac{f^3}{N} \sum_{k=4}^{\infty} \frac{1}{k!} |B_k| \binom{k}{3}. \]

Now it is easily seen that the last sum is approximately 0.0089, either by numerical summation or by the fact that \(|B_4| = i^3 - B_3\) for \(k \geq 2\) whence the sum equals \(g''(i)/2 - 1/12\) where \(g(z) = z/(e^z - 1)\) and thus
\[ g''(i) = \frac{1}{2} \sin^{-2} \frac{1}{2} - \frac{1}{4} \cos \frac{1}{2} \sin^{-3} \frac{1}{2} \approx 0.1844. \]

But from
\[ \left| \Delta + \frac{1}{12} \cdot \frac{f}{N} \right| \leq 0.0089 \frac{f^3}{N} < \frac{1}{60} \cdot \frac{f^3}{N} \leq \frac{1}{60} \cdot \frac{f}{N} \]

(6) follows.

Acknowledgment. The author is indebted to the referee for removing one unnecessary condition for the validity of (3) and for pointing out one essential reference concerning (1).

REFERENCES
