A NOTE ON THE DEGREE OF NORMAL APPROXIMATION TO THE DISTRIBUTION FUNCTION OF THE MEAN OF SAMPLES FROM FINITE POPULATIONS

YASUSHI TAGA

Summary

From the standpoint of asymptotic equivalence, a sufficient condition is obtained, under which the degree of normal approximation to the distribution function of the mean of samples from finite populations can be ascertained.

1. Introduction

W. G. Madow [1] has shown that the mean of a sample from a finite population is asymptotically normally distributed under some conditions on sampling ratio and moments. However, the degree of such normal approximation has not been given. In this note, a sufficient condition for ascertaining the degree of the normal approximation is presented by making use of the asymptotic equivalence measure defined in [2].

2. Normal approximation

LEMMA 1. *Let $X^{(N)}=(X_1^{(N)}, \cdots, X_p^{(N)})$, $Y^{(N)}=(Y_1^{(N)}, \cdots, Y_p^{(N)})$ be discrete vector random variables in the p-dimensional Euclidian space R_p, and $f^{(N)}(x_1, \cdots, x_p)$, $g^{(N)}(y_1, \cdots, y_p)$ be probability functions corresponding to $X^{(N)}$ and $Y^{(N)}$ $(N=1, 2, \cdots)$. Let $T(z_1, \cdots, z_p)$ be a one-valued function of $(z_1, \cdots, z_p)$. If $X^{(N)}$ and $Y^{(N)}$ are asymptotically equivalent, i.e.,*

$$(2.1) \quad \rho(f^{(N)}, g^{(N)})=\sum_{R_p} |f^{(N)}(x_1, \cdots, x_p)-g^{(N)}(x_1, \cdots, x_p)| \to 0 \text{ as } N\to\infty ,$$

then $T(X_1^{(N)}, \cdots, X_p^{(N)})$ and $T(Y_1^{(N)}, \cdots, Y_p^{(N)})$ are also asymptotically equivalent.

PROOF. By the assumptions, both $T(X_1^{(N)}, \cdots, X_p^{(N)})$ and $T(Y_1^{(N)}, \cdots, Y_p^{(N)})$ may be thought as discrete random variables, and we have

$$P\{T(X_1^{(N)}, \cdots, X_p^{(N)})=z\}=\sum_{E(z)} f^{(N)}(x_1, \cdots, x_p)$$

and

$$P\{T(Y_1^{(N)}, \cdots, Y_p^{(N)})=z\}=\sum_{E(z)} g^{(N)}(x_1, \cdots, x_p)$$

where $E(z)=\{(x_1, \cdots, x_p);\ T(x_1, \cdots, x_p)=z\}$. Hence the assertion of this lemma is clear.

Remark. In the above lemma, $\rho(f^{(N)}, g^{(N)})$ may be taken as a measure of asymptotic equivalence.

LEMMA 2. *Let $X^{(N)}=(X_1^{(N)}, \cdots, X_n^{(N)})$ be a sample of size n taken equally probably without replacement from a finite population, π_N, of size N, and $Y^{(N)}=(Y_1^{(N)}, \cdots, Y_n^{(N)})$ be a sample of size n taken equally probably with replacement from π_N. Then $(X_1^{(N)}, \cdots, X_n^{(N)})$ and $(Y_1^{(N)}, \cdots, Y_n^{(N)})$ are asymptotically equivalent if $\prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right)\to 1$ as $N\to\infty$.*

PROOF. Let $f^{(N)}(x_1, \cdots, x_n)$ and $g^{(N)}(y_1, \cdots, y_n)$ be probability functions corresponding to $X^{(N)}$ and $Y^{(N)}$, respectively. Then it is clear that

$$f^{(N)}(x_1, \cdots, x_n)=\begin{cases} \dfrac{1}{N(N-1)\cdots(N-n+1)} & \text{if } x_i\neq x_j \text{ for any } i\neq j, \\ 0 & \text{if } x_i=x_j \text{ for some } i\neq j, \end{cases} \tag{2.2}$$

$$g^{(N)}(x_1, \cdots, x_n)=\frac{1}{N^n}. \tag{2.3}$$

Hence, from (2.2) and (2.3), we obtain

$$\begin{aligned} \rho(f^{(N)}, g^{(N)}) &= \sum_{R_n} |f^{(N)}(x_1, \cdots, x_n)-g^{(N)}(x_1, \cdots, x_n)| \\ &= N(N-1)\cdots(N-n+1)\left\{\frac{1}{N(N-1)\cdots(N-n+1)}-\frac{1}{N^n}\right\} \\ &\quad +\left\{N^n-N(N-1)\cdots(N-n+1)\right\}\frac{1}{N^n} \\ &= 2\left\{1-\prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right)\right\}. \end{aligned} \tag{2.4}$$

By the assumption that $\prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right)\to 1$ as $N\to\infty$, we can conclude from (2.4) that $\rho(f^{(N)}, g^{(N)})\to 0$ as $N\to\infty$.

Remark. In order that $\prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right)\to 1$, it is sufficient that $\frac{n^2}{N}\to 0$.

In fact, from inequalities $1 \geqq \prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right) > \left(1-\frac{n}{N}\right)^n \geqq 1-\frac{n^2}{N}$, it is easily seen that $\prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right) \to 1$ if $\frac{n^2}{N} \to 0$, and that $\rho(f^{(N)}, g^{(N)}) < \frac{2n^2}{N}$.

THEOREM. *Let $(X_1^{(N)}, \cdots, X_n^{(N)})$ be a sample of size n taken equally probably without replacement from a population, π_N, of size N with mean μ_N, variance σ_N^2 and the third order absolute moment β_{3N}. Further, suppose that $\frac{\beta_{3N}}{\sigma_N^3}$ is bounded for all N and that N and n tend to infinity in such a way that $\frac{n^2}{N} \to 0$. Then the random variable $\frac{1}{\sqrt{n}\,\sigma_N}\sum_{j=1}^{n}(X_j^{(N)}-\mu_N)$ is asymptotically normally distributed with mean 0 and variance 1, and the degree of approximation is of order of $O\left(\frac{1}{\sqrt{n}}\right)$ in the sense of distribution function if $N \geqq bn^{5/2}$, where b is a positive constant.*

PROOF. Let $(Y_1^{(N)}, \cdots, Y_n^{(N)})$ be a sample taken equally probably with replacement from π_N. Then $(X_1^{(N)}, \cdots, X_n^{(N)})$ and $(Y_1^{(N)}, \cdots, Y_n^{(N)})$ are asymptotically equivalent by the remark of lemma 2 if $\frac{n^2}{N} \to 0$ as $N \to \infty$, and $\rho(f^{(N)}, g^{(N)}) < \frac{2n^2}{N}$. Hence, by lemma 1, $\tilde{X}_n^{(N)} = \frac{1}{\sqrt{n}\,\sigma_N}\sum_{j=1}^{n}(X_j^{(N)}-\mu_N)$ and $\tilde{Y}_n^{(N)} = \frac{1}{\sqrt{n}\,\sigma_N}\sum_{j=1}^{n}(Y_j^{(N)}-\mu_N)$ are also asymptotically equivalent with measure $\rho(f^{(N)}, g^{(N)}) < \frac{2n^2}{N}$. Since $(Y_1^{(N)}, \cdots, Y_n^{(N)})$ can be taken as a sequence of independent random variables from π_N, as is well known, $\tilde{Y}_n^{(N)} = \frac{1}{\sqrt{n}\,\sigma_N}\sum_{j=1}^{n}(Y_j^{(N)}-\mu_N)$ is asymptotically normally distributed with mean 0 and variance 1 and the order of normal approximation is $n^{-1/2}$ (see [3]). Let $F_n^{(N)}(x)$ and $G_n^{(N)}(y)$ be distribution functions corresponding to $\tilde{X}_n^{(N)}$ and $\tilde{Y}_n^{(N)}$ respectively. Then, from the above arguments, we can see that

$$|F_n^{(N)}(x) - G_n^{(N)}(x)| \leqq \rho(f^{(N)}, g^{(N)}) < \frac{2n^2}{N} \tag{2.5}$$

and

$$|G_n^{(N)}(x) - \Phi(x)| \leqq c\frac{\beta_{3N}}{\sigma_N^3\sqrt{n}}, \tag{2.6}$$

where $\Phi(x)$ denotes the unit normal distribution function and c is a constant, say, 2.03 (see [4]). From (2.5) and (2.6) we obtain

$$|F_n^{(N)}(x)-\Phi(x)|<\frac{2n^2}{N}+c\frac{\beta_{3N}}{\sigma_N^3\sqrt{n}} . \tag{2.7}$$

Since $\frac{\beta_{3N}}{\sigma_N^3}$ is bounded for all N and $N\geqq bn^{5/2}$ by the assumptions, we can rewrite (2.7) as

$$|F_n^{(N)}(x)-\phi(x)|<\frac{d}{\sqrt{n}}, \quad d=2b+c\frac{\beta_{3N}}{\sigma_N^3} . \tag{2.8}$$

Remark.

1) The sample mean $\frac{1}{n}\sum_{j=1}^{n} X_j^{(N)}=\frac{\sigma_N}{\sqrt{n}}\tilde{X}_n^{(N)}+\mu_N$ is asymptotically normally distributed with mean μ_N and variance σ_N^2/n, and its distribution function is $F_n^{(N)}\left(\frac{x-\mu_N}{\sigma_N/\sqrt{n}}\right)$ which is different from $\Phi\left(\frac{x-\mu_N}{\sigma_N/\sqrt{n}}\right)$ at most by $\frac{d}{\sqrt{n}}$.

2) In the above discussion, n may or may not depend on N.

3) The condition that $\frac{n^2}{N}\to 0$ as $N\to\infty$ may be replaced by the condition that $\prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right)\to 0$ as $N\to\infty$, which is more general than the former, but is difficult to check.

4) The above evaluation of the degree of approximation may not be so accurate in comparison with the true difference $|F_n^{(N)}(x)-\Phi(x)|$.

5) In practical applications, population size N is generally much larger than sample size n, and σ_N and β_{3N} can be estimated from the sample. Therefore, we can evaluate the difference $|F_n^{(N)}(x)-\Phi(x)|$ by (2.7).

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] W. G. Madow, "On the limiting distributions of estimates based on samples from finite universes," *Ann. Math. Stat.*, 14 (1948), 535-545.

[2] S. Ikeda, "Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence," *Ann. Inst. Stat. Math.*, 15 (1963), 87-116.

[3] A. C. Berrey, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.*, 49 (1941), 122-131.

[4] K. Takano, "A note on the paper of A. C. Berry," *Res. Memoir of Inst. Stat. Math.*, 6 (1950), 408-415.