ON THE ASYMPTOTIC DISTRIBUTION OF THE MEAN OF DISTINCT UNITS IN SAMPLING FROM A FINITE POPULATION[1]

by

P. K. PATHAK[3] and J. SETHURAMAN[2]

1. Introduction

For each integer k, $S_{N_k} = \{1, 2, \ldots, N_k\}$ is a finite population of N_k units with characteristics $\{Y_{k,1}, Y_{k,2}, \ldots, Y_{k,N_k}\}$. A simple random sample, S_{n_k}, of size n_k is drawn with replacement from S_{n_k}. The set of distinct units in S_{n_k} is denoted by s_{m_k} and it contains m_k units. Let

$$\bar{y}_k = \frac{1}{m_k} \sum_{i \in s_{m_k}} Y_{k,i} .$$

This note is concerned with the asymptotic distribution of $\bar{y}_k$. In section 3, we show this to be normal by the following device. The conditional distribution of s_{m_k} when m_k is fixed is that of a simple random sample without replacement from S_{N_k}. The mean of a simple random sample without replacement and m_k are known to be asymptotically normal. These facts and the theorem SETHURAMAN [3] quoted in Section 2 (Lemma 3), establish the result.

2. Preliminaries

Let
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i \in S_{N_k}} Y_{k,i}$$

$$\sigma_k^2 = \frac{1}{m_k} \cdot \frac{N_k - m_k}{N_k} \cdot \frac{1}{N_k - 1} \cdot \sum_{i \in S_{N_k}} (Y_{k,i} - \overline{Y}_k)^2 .$$

We now define two conditions:

Condition A.

For each $\tau > 0$

$$\frac{\sum_{i \in S(\tau)} (Y_{k,i} - \overline{Y}_k)^2}{\sum_{i \in S_{N_k}} (Y_{k,i} - \overline{Y}_k)^2} \to 0$$

[1] This work was partially supported by the National Science Foundation, Grant G-18976.

[2] Michigan State University.

[3] University of Illinois.

where

$$S(\tau) = \left\{ i : |Y_{k,i} - \overline{Y}_k| \geqq \tau \sqrt{\sum_{i \in S_{N_k}} (Y_{k,i} - \overline{Y}_k)^2} \right\}.$$

Condition B.

$$N_k e^{-\frac{n_k}{N_k}} \left(1 - \left(1 + \frac{n_k}{N_k}\right) e^{-\frac{n_k}{N_k}}\right) \to \infty .$$

When $\frac{n_k}{N_k} \to \alpha$, finite or infinite, condition (B) is equivalent to the following:

$$\text{either } \frac{n_k}{N_k} \to \alpha,\ 0 < \alpha < \infty$$

$$\text{or} \quad \frac{n_k}{N_k} \to 0 \text{ and } \frac{n_k^2}{N_k} \to \infty$$

$$\text{or} \quad \frac{n_k}{N_k} \to \infty \text{ and } N_k e^{-\frac{n_k}{N_k}} \to \infty .$$

The following lemma is due to A. RÉNYI [2].

Lemma 1 (RÉNYI).

If condition B is satisfied then

$$\mathbf{P}\{\eta_k \leqq y\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp\,(-t^2/2)\, dt .$$

where

(1) $$\eta_k = \frac{m_k - N_k\left(1 - e^{-\frac{n_k}{N_k}}\right)}{\sqrt{N_k e^{-\frac{r_k}{N_k}} \left(1 - \left(1 + \frac{n_k}{N_k}\right) e^{-\frac{n_k}{N_k}}\right)}} .$$

For each θ, let $\{m_k(\theta)\}$ be a sequence of integers such that the following condition holds.

Condition C.

$$m_k(\theta) \text{ and } N_k - m_k(\theta) \to \infty \text{ uniformly in } \theta.$$

Let $s_{m_k(\theta)}$ be a simple random sample of size $m_k(\theta)$ drawn from S_{N_k} without replacement. Let

$$\overline{y}_k(\theta) = \frac{1}{m_k(\theta)} \sum_{i \in s_{m_k(\theta)}} Y_{k,i}$$

and

$$\sigma_k^2(\theta) = \frac{1}{m_k(\theta)} \cdot \frac{N_k - m_k(\theta)}{N_k - 1} \cdot \frac{1}{N_k} \cdot \sum_{i \in S_{N_k}} (Y_{k,i} - \overline{Y}_k)^2 ,$$

Let

$$\xi_k(\theta) = \frac{\overline{y}_k(\theta) - \overline{Y}_k}{\sigma_k(\theta)} .$$

Lemma 2 (HÁJEK).
Under conditions (A) and (C)

$$\mathbf{P}\{\xi_k(\theta) \leq x\} \to \frac{1}{\sqrt{2x}} \int_{-\infty}^{x} \exp(-t^2/2)\, dt$$

uniformly in θ.

Proof. J. Hájek [1] has shown that the above convergence holds for any fixed θ. A few simple modifications of this proof, which will not be recounted here, establish the required uniformity in convergence.

Definition. A sequence of random variables $\{\xi_k(\theta)\}$ is said to converge to $\{\xi(\theta)\}$ in the UC^* sense relative to θ in a bounded interval on the real line if

$$\mathbf{E}[g(\xi_k(\theta))] \to \mathbf{E}[g(\xi(\theta))] \text{ uniformly in } \theta$$

and

$\mathbf{E}[g(\xi(\theta))]$ is continuous in θ for every bounded continuous function $g(x)$.

We shall make use of the following lemma due to J. SETHURAMAN [3], Theorem 3 in the next section.

Lemma 3 (SETHURAMAN).

Let (ξ_k, η_k), $k = 0, 1, \ldots$ be a sequence of random variables. Let the conditional distribution of ξ_k given that $\eta_k = \eta$ converge in the UC^* sense to the conditional distribution of ξ_0 given that $\eta_0 = \eta$ relative to η in any bounded interval. Let the distribution of η_k converge to the distribution of η_0 in law. Then the joint distribution of (ξ_k, η_k) converges to the joint distribution of (ξ_0, η_0) in law.

3. Main theorem

Define
$$\xi_k = \frac{\bar{y}_k - \bar{Y}_k}{\sigma_k}$$
and let η_k be as in (1).

Theorem. *Under conditions* (A) *and* (B)

$$\mathbf{P}\{\xi_k \leq x, \eta_k \leq y\} \to \frac{1}{2\pi} \int_{-\infty}^{x} \int_{-\infty}^{y} \exp\left(-\frac{1}{2}(t^2 + u^2)\right) dt\, du\,. \tag{2}$$

In particular,

$$\mathbf{P}\{\xi_k \leq x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)\, dt\,. \tag{3}$$

Proof. The conditional distribution of s_{m_k} given that $\eta_k = \eta$, that is, given

$$m_k = m_k(\eta) = N_k(1 - e^{-\frac{n_k}{N_k}}) + \eta \sqrt{N_k e^{-\frac{n_k}{N_k}} \left(1 - \left(1 + \frac{n_k}{N_k}\right) e^{-\frac{n_k}{N_k}}\right)}$$

is the conditional distribution of a simple random sample of size $m_k(\eta)$ drawn from S_{N_k} without replacement. Further ξ_k is the normalized mean of this sample.

Under condition (B), $m_k(\eta)$ and $N_k - m_k(\eta) \to \infty$ uniformly in any bounded interval of η. This fact together with condition (B) yield the following from Lemma 2. The conditional distribution of ξ_k given that $\eta_k = \eta$ converges in the UC^* sense to the standard normal distribution relative to η in any bounded interval. Lemma 1 states that the distribution of η_k converges to the standard normal distribution in law. The relation (2) now follows from Lemma 3. Relation (3) is immediate from relation (1).

4. Remarks

It would be interesting to find out whether results like Lemmas 1 and 2 can be obtained when sampling with unequal probabilities. Some results on the lines of Lemma 1 are now under investigation and the results shall be published elsewhere.

(Received October 1, 1963)

REFERENCES

[1] HÁJEK, J.: "Limiting distributions in simple random sampling from a finite population". *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960) 361—374.

[2] RÉNYI, A.: "Three new proofs and a generalization of a theorem of Irving Weiss", *Publ. Math. Inst. Hung. Acad. Sci.* **6** (1962) 203—214.

[3] SETHURAMAN, J.: "Some limit theorems for joint distributions". *Sankhya, Series A* (1961) 379—386.

ОБ АСИМПТОТИЧЕСКОМ РАСПРЕДЕЛЕНИИ СРЕДНЕГО РАЗЛИЧНЫХ ЗНАЧЕНИЙ В ВЫБОРКЕ ИЗ КОНЕЧНОЙ ПОПУЛАЦИИ

P. K. PATHAK и J. SETHURAMAN

Резюме

Воспользуется теорема SETHURAMAN-а [3], чтобы получить асимптотическую нормальность различных значений в простой случайной выборке с возвращением из конечной популяции.