# LIMITING DISTRIBUTIONS IN SIMPLE RANDOM SAMPLING FROM A FINITE POPULATION

by

JAROSLAV HÁJEK[1]

## 1. Introduction and summary

Sampling from a finite population may be considered as a random experiment whose outcomes are subsets $s$ of the set $S = \{1, 2, \ldots, N\}$; $s$ is called a sample and $S$ is called a population. Denote an $s$ consisting of $k$ elements by $s_k$ and the probability of $s_k$ by $\mathbf{P}(s_k)$.

Simple random sampling of sample size $n$ is defined by the following probabilities $\mathbf{P}(s_k)$:

$$(1.1)\qquad \mathbf{P}(s_k) = \begin{cases} \binom{N}{n}^{-1} & \text{if } k = n \quad \text{[Simple random sampling of size } n\text{]} \\ 0 & \text{otherwise}. \end{cases}$$

In this paper we shall make use of so called Poisson sampling defined as follows:

$$(1.2)\qquad \mathbf{P}(s_k) = \left(\frac{n}{N}\right)^k \left(1 - \frac{n}{N}\right)^{N-k} \quad \text{[Poisson sampling of mean size } n\text{]}.$$

Let us have a sequence $y_1, \ldots, y_N$ of real numbers, and put

$$(1.3)\qquad \xi = \sum_{i \in s_n} y_i$$

where $s_n$ is a simple random sample and $\sum_{i \in s_n}$ extends over all $i$ contained in the sample $s_n$. Obviously $\xi = \xi(s_n)$ is a random variable with finite mean value

$$(1.4)\qquad \mathbf{E}\,\xi = \frac{n}{N} \sum_{i=1}^{N} y_i$$

and variance

$$(1.5)\qquad \mathbf{D}\,\xi = \frac{n}{N}\,\frac{N-n}{N-1} \sum_{i=1}^{N} (y_i - \overline{Y})^2 \qquad \left[\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i\right].$$

---

[1] Mathematical Institute of the Czechoslovak Academy of Sciences, Prague.

Let us consider an infinite sequence of simple random sampling experiments, the $\nu$-th of which has the size $n_\nu$ and refers to a population of size $N_\nu$ and with values $y_{\nu 1}, \ldots, y_{\nu N_\nu}$. Let $\xi_\nu$ be the random variable defined by (1.3) corresponding to the $\nu$-th sampling experiment.

Now we ask about conditions concerning $\{y_{\nu i}, n_\nu, N_\nu\}$ under which

$$\lim_{\nu\to\infty} \mathbf{P}\{\xi_\nu - \mathbf{E}\,\xi_\nu < x\sqrt{\mathbf{D}\,\xi_\nu}\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt \tag{1.6}$$

or

$$\lim_{\nu\to\infty} \mathbf{P}\{\xi_\nu = k\} = e^{-\lambda}\frac{\lambda^k}{k!} \tag{1.7}$$

or, generally, the distribution function of $\xi_\nu$ converges to a limiting distribution law. We naturally suppose that $n_\nu \to \infty$ and $N_\nu - n_\nu \to \infty$.

In the following sections we shall give a complete solution of this problem, i. e. we shall indicate necessary and sufficient conditions. As for convergence to the normal distribution law the necessary and sufficient condition coincides with that one derived by ERDŐS and RÉNYI in the paper [1], as could be expected.

## 2. The fundamental lemma

We shall show that the above problem can be completely reduced to the same problem concerning sums of independent random variables.

First observe that Poisson sampling may be interpreted as simple random sampling of size $k$, where $k$ is a binomial random variable attaining the value $k$ with probability $\binom{N}{k}\left(\frac{n}{N}\right)^k\left(1-\frac{n}{N}\right)^{N-k}$. Actually, it suffices to consult (1.1) and (1.2) and notice that

$$\left(\frac{n}{N}\right)^k\left(1-\frac{n}{N}\right)^{N-k} = \left[\binom{N}{k}\left(\frac{n}{N}\right)^k\left(1-\frac{n}{N}\right)^{N-k}\right]\binom{N}{k}^{-1}$$

Clearly $\mathbf{E}\,k = n$ and

$$\mathbf{E}(k-n)^2 = n\left(1-\frac{n}{N}\right). \tag{2.1}$$

Now, it is easy to define an experiment producing simultaneously a simple random sample $s_n$ and a Poisson sample $s_k$ such that $s_n \subset s_k$ if $n \leq k$, and $s_n \supset s_k$ if $n \geq k$. This may be done as follows:

1° First we realize the binomial random variable $k$ attaining the value $k$ with probability

$$\binom{N}{k}\left(\frac{n}{N}\right)^k\left(1-\frac{n}{N}\right)^{N-k}, \qquad 0 \leq k \leq N.$$

2° a) When $k = n$, then we select a simple random sample $s_n = s_k$ which is a simultaneous realization of both simple random sampling and Poisson sampling. b) When $k > n$, then we select a simple random sample $s_k$, which is a realization of Poisson sampling; thereafter we select a simple random sample $s_n$ of size $n$ from $s_k$ ($s_k$ represents here a population), $s_n$ being a realization of simple random sampling. c) When $k < n$, then we select a simple random sample $s_n$, which is a realization of simple random sampling; thereafter we select a simple random sample $s_k$ from $s_n$ ($s_n$ represents here a population), $s_k$ being a realization of Poisson sampling.

Put

$$\eta = \sum_{i \in s_n} (y_i - \overline{Y}) = \xi - n\overline{Y} \tag{2.2}$$

and

$$\eta^* = \sum_{i \in s_k} (y_i - \overline{Y}) \tag{2.3}$$

where $(s_n, s_k)$ are joined samples from the above experiments, $s_n$ representing a simple random sample and $s_k$ a Poisson sample. The number of summands equals constantly $n$ in (2.2) and is a binomial random variable in (2.3). Clearly,

(2.4)

$$\eta - \eta^* = \begin{cases} 0 & \text{if} \quad k = n \\ \sum\limits_{i \in s_n - s_k} (y_i - \overline{Y}) & \text{if} \quad k < n \\ \sum\limits_{i \in s_k - s_n} (y_i - \overline{Y}) & \text{if} \quad k > n \end{cases}$$

since either $s_k$ is a subset of $s_n$ or, conversely $s_n$ is a subset of $s_k$.

**Lemma 2.1.** *The following inequality holds true :*

$$\frac{\mathbf{E}(\eta - \eta^*)^2}{\mathbf{D}\,\eta^*} \leq \sqrt{\frac{1}{n} + \frac{1}{N-n}}. \tag{2.5}$$

**Proof.** If $k$ is fixed, then $s_n - s_k$ or $s_k - s_n$ represents a simple random sample of size $|k - n|$. Consequently, in view of (1.5), we have

$$\mathbf{E}\{(\eta - \eta^*)^2 | k\} = \mathbf{D}\{\eta - \eta^* | k\} = \frac{|k-n|}{N} \frac{N - |k-n|}{N-1} \sum_{i=1}^{N} (y_i - \overline{Y})^2 \leq \tag{2.6}$$

$$\leq |k-n| \frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{Y})^2 .$$

The inequality (2.6) together with (2.1) obtains that

$$\mathbf{E}\{(\eta - \eta^*)^2\} \leq \mathbf{E}\,|k-n| \frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{Y})^2 \leq \tag{2.7}$$

$$\leq \sqrt{\mathbf{E}(k-n)^2}\, \frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{Y})^2 = \sqrt{n\left(1 - \frac{n}{N}\right)} \frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{Y})^2 .$$

In a similar way we could derive

$$\mathbf{D}\,\eta^* = n\left(1 - \frac{n}{N}\right)\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2. \tag{2.8}$$

We shall prefer however, to prove (2.8) by the following consideration: Any sampling experiment consists of $N$ dichotomous experiments, the $i$-th of which has the following two possible outcomes: including the element $i$ in the sample $s$ and not including the element $i$ in the sample $s$. If all these experiments are mutually independent and the probability of including the element $i$ equals constantly $\frac{n}{N}$, $1 \leq i \leq N$, we easily see that one gets Poisson sampling, i. e. that each sample $s_k$ has the probability (1.2). This fact implies that $\eta^*$ may be judged as a sum of $N$ independent random variables,

$$\eta^* = \sum_{i=1}^{N}\zeta_i \tag{2.9}$$

where

$$\zeta_i = \begin{cases} y_i - \bar{Y} & \text{with probability } \frac{n}{N} \quad (\text{if } i \in s_k) \\ 0 & \text{with probability } 1 - \frac{n}{N} (\text{if } i \in S - s_k). \end{cases} \tag{2.10}$$

Clearly

$$\mathbf{D}\,\zeta_i = (y_i - \bar{Y})^2\frac{n}{N}\left(1 - \frac{n}{N}\right) \qquad (1 \leq i \leq N) \tag{2.11}$$

which proves (2.8).

Combining (2.7) and (2.8) we obviously obtain the inequality (2.5) which was to be proved.

Let us consider a sequence of experiments of the above kind (i. e. producing joined simple random and Poisson samples) and denote by $\eta_\nu$ and $\eta_\nu^*$ the random variables (2.2) and (2.3) referring to the $\nu$-th experiment. From Lemma 2.1 it follows that

$$\lim_{\nu\to\infty}\frac{\mathbf{E}(\eta_\nu - \eta_\nu^*)^2}{\mathbf{D}\,\eta_\nu^*} = 0 \qquad \text{if} \begin{cases} n_\nu \to \infty \\ N_\nu - n_\nu \to \infty. \end{cases} \tag{2.12}$$

**Remark 2.1.** The relation (2.12) implies that, provided $n_\nu \to \infty$ and $N_\nu - n_\nu \to \infty$, the limiting variances and distributions of random variables $A_\nu + B_\nu\,\eta_\nu$ and $A_\nu + B_\nu\,\eta_\nu^*$ exist under the same conditions, and if exist, are the same. The random variable $\eta_\nu^*$, however, is a sum of independent addends (2.10), so that when studying the limiting distributions of $A_\nu + B_\nu\,\eta_\nu^*$, we may simple apply the well-known theory of summation of independent random variables. See [3].

### 3. Convergence to the normal distribution

We shall prove that the condition derived by ERDŐS-RÉNYI in [1] is not only sufficient but also necessary provided that $n_\nu \to \infty$ and $N_\nu - n_\nu \to \infty$.

**Theorem 3.1.** *Let $S_{\nu\tau}$ be the subset of elements of $S_\nu = \{1, \ldots, N_\nu\}$ on which the inequality*

$$(3.1) \qquad |y_{\nu i} - \overline{Y}_\nu| > \tau \sqrt{\mathbf{D}\xi_\nu}$$

*holds; let $\mathbf{D}\xi_\nu$ denote the variance* (1.5) *referring to the $\nu$-th experiment. Suppose that $n_\nu \to \infty$ and $N_\nu - n_\nu \to \infty$.*

*Then the random variable $\xi_\nu$ defined by* (1.3) *has asymptotically normal distribution with parameters* ($\mathbf{E}\,\xi_\nu$, $\mathbf{D}\,\xi_\nu$) *if and only if*

$$(3.2) \qquad \lim_{\nu\to\infty} \frac{\sum\limits_{i\in S_{\nu\tau}} (y_{\nu i} - \overline{Y}_\nu)^2}{\sum\limits_{i\in S_\nu} (y_{\nu i} - \overline{Y}_\nu)^2} = 0 \qquad \text{for any } \tau > 0$$

*where $\sum\limits_{i\in S_\nu}$ denotes the same summation as $\sum\limits_{i=1}^{N_\nu}$.*

**Proof.** In view of Remark 2.1, it suffices to establish sufficient and necessary conditions for asymptotical normality of the random variable $\eta_\nu^*$ defined by (2.9), namely with parameters (0, $\mathbf{D}\,\eta_\nu^*$). Notice that $\mathbf{E}\,\eta_\nu^* = 0$, since

$$(3.3) \qquad \mathbf{E}\,\eta^* = \sum_{i=1}^{N} \mathbf{E}\,\zeta_i = \sum_{i=1}^{N} \frac{n}{N}(y_i - \overline{Y}) = 0\,.$$

First suppose that the random variables $\zeta_{\nu i} = \zeta_i$ defined by (2.10) are infinitesimal, i. e. that

$$(3.4) \qquad \lim_{\nu\to\infty} \frac{\max\limits_{1\le i\le N} \mathbf{D}\,\zeta_{\nu i}}{\sum\limits_{i=1}^{N_\nu} \mathbf{D}\,\zeta_{\nu i}} = 0\,.$$

In view of (2.11), (3.4) is equivalent to

$$(3.5) \qquad \lim_{\nu\to\infty} \frac{\max\limits_{1\le i\le N_\nu} (y_{\nu i} - \overline{Y}_\nu)^2}{\sum\limits_{i=1}^{N_\nu} (y_{\nu i} - \overline{Y}_\nu)^2} = 0\,.$$

The condition (3.5) is clearly much weaker than the condition (3.2); it is usually called the Noether condition.

Provided that (3.4) holds, the necessary and sufficient condition for asymptotical normality of $\eta_\nu^*$ with parameters (0, $\mathbf{D}\,\eta_\nu^*$) is given by the Lindeberg condition. Since the random variables $\zeta_{\nu i} - \mathbf{E}\,\zeta_{\nu i}$ take on values $(y_{\nu i} - \overline{Y}_\nu)\left(1 - \frac{n_\nu}{N_\nu}\right)$ and $-(y_{\nu i} - \overline{Y}_\nu)\frac{n_\nu}{N_\nu}$ with respective probabilities $\frac{n_\nu}{N_\nu}$ and $1 - \frac{n_\nu}{N_\nu}$ we have

$$\frac{1}{\mathbf{D}\,\eta_\nu^*}\sum_{i=1}^{N_\nu}\int\limits_{|x|>\tau\sqrt{\mathbf{D}\eta_\nu^*}} x^2\, d\,\mathbf{P}\{\zeta_{\nu i}-\mathbf{E}\,\zeta_{\nu i}<x\} =$$

$$(3.6)\qquad = \frac{\frac{n_\nu}{N_\nu}\left(1-\frac{n_\nu}{N_\nu}\right)^2 \sum_{i\in C_{\nu\tau}}(y_{\nu i}-\overline{Y}_\nu)^2 + \left(1-\frac{n_\nu}{N_\nu}\right)\left(\frac{n_\nu}{N_\nu}\right)^2 \sum_{i\in B_{\nu\tau}}(y_{\nu i}-\overline{Y}_\nu)^2}{\frac{n_\nu}{N_\nu}\left(1-\frac{n_\nu}{N_\nu}\right)\sum_{i\in S_\nu}(y_{\nu i}-\overline{Y}_\nu)^2} =$$

$$= \frac{\left(1-\frac{n_\nu}{N_\nu}\right)\sum_{i\in C_{\nu\tau}}(y_{\nu i}-\overline{Y}_\nu)^2 + \frac{n_\nu}{N_\nu}\sum_{i\in B_{\nu\tau}}(y_{\nu i}-\overline{Y}_\nu)^2}{\sum_{i\in S_\nu}(y_{\nu i}-\overline{Y}_\nu)^2}$$

where $C_{\nu\tau}$ and $B_{\nu\tau}$ are subsets of elements of $S_\nu$ on which

$$(3.7)\qquad C_{\nu\tau}:\ |y_{\nu i}-\overline{Y}_\nu|\left(1-\frac{n_\nu}{N_\nu}\right) > \tau\sqrt{\frac{n_\nu}{N_\nu}\left(1-\frac{n_\nu}{N_\nu}\right)\sum_{i\in S_\nu}(y_{\nu i}-\overline{Y}_\nu)^2}$$

and

$$(3.8)\qquad B_{\nu\tau}:\ |y_{\nu i}-\overline{Y}_\nu|\frac{n_\nu}{N_\nu} > \tau\sqrt{\frac{n_\nu}{N_\nu}\left(1-\frac{n_\nu}{N_\nu}\right)\sum_{i\in S_\nu}(y_{\nu i}-\overline{Y}_\nu)^2}$$

respectively. In view of (1.5), we see that

$$(3.9)\qquad C_{\nu\tau} = S_{\nu,\tau\sqrt{\frac{N_\nu(N_\nu-1)}{(N_\nu-n_\nu)^2}}} \quad\text{and}\quad B_{\nu\tau} = S_{\nu,\tau\sqrt{\frac{N_\nu(N_\nu-1)}{n_\nu^2}}}\,.$$

From (3.9) it follows that (3.2) is equivalent to the condition that the first member of (3.6) converges to 0, i. e. to the fulfilment of the Lindeberg condition for $\eta_\nu^*$.

Thus it remains to prove that $\eta_\nu^*$ cannot have a limiting normal distribution with parameters $(\mathbf{E}\,\eta_\nu^*, \mathbf{D}\,\eta_\nu^*)$, if (3.5) does not hold.

We may suppose without any loss of generality that $n_\nu \leqq \frac{1}{2}N_\nu$ and

$$(3.10)\qquad |y_{\nu 1}-\overline{Y}_\nu| \geqq |y_{\nu 2}-\overline{Y}_\nu| \geqq \ldots \geqq |y_{\nu N_\nu}-\overline{Y}_\nu|\,.$$

If (3.5) is not satisfied, then there exist an $\varepsilon \neq 0$ such that

$$(3.11)\qquad \lim_{\nu\to\infty}\frac{y_{\nu 1}-\overline{Y}_\nu}{\sqrt{\sum_{i\in S_\nu}(y_{\nu i}-\overline{Y}_\nu)^2}} = \varepsilon \neq 0$$

for some subsequence of indices $\nu$. Taking a new subsequence from this subsequence, we may assume that

$$(3.12)\qquad \lim_\nu \frac{n_\nu}{N_\nu} = c \leqq \frac{1}{2}\,.$$

For simplicity let us introduce no new symbols for denoting the subsequences.

Now the relations (3.11) and (3.12) mean that the distribution function of the random variable

$$\frac{\zeta_{\nu 1}}{\sqrt{\mathbf{D}\,\eta_\nu^*}} = \begin{cases} \dfrac{y_{\nu 1} - \overline{Y}_\nu}{\sqrt{\dfrac{n_\nu}{N_\nu}\left(1 - \dfrac{n_\nu}{N_\nu}\right) \sum\limits_{i \in S_\nu} (y_{\nu i} - \overline{Y}_\nu)^2}} & \text{with probability } \dfrac{n_\nu}{N_\nu} \\ 0 & \text{with probability } 1 - \dfrac{n_\nu}{N_\nu} \end{cases} \tag{3.13}$$

converges to a distribution function which has a jump $1 - c$ at the point 0 and, if $c > 0$, a jump $c$ at the point $\dfrac{\varepsilon}{\sqrt{c(1-c)}}$. Let us discuss each of the cases $c = 0$ and $c > 0$ separately.

If $c = 0$ the variance of (3.13) does not converge to the variance of the limiting distribution. Actually, (3.13) has the limiting variance $\varepsilon^2$ while the limiting distribution is degenerated to the single point 0 so that it has the variance 0. Hence if there existed a limiting distribution of the statistic

$$\frac{\eta_\nu^*}{\sqrt{\mathbf{D}\,\eta_\nu^*}} = \frac{\zeta_{\nu 1}}{\sqrt{\mathbf{D}\,\eta_\nu^*}} + \frac{\sum\limits_{i=2}^{N_\nu} \zeta_{\nu i}}{\sqrt{\mathbf{D}\,\eta_\nu^*}} \tag{3.13}$$

it would have a variance smaller than 1. Consequently, $\eta_\nu^*$ cannot have asymptotically normal distribution with parameters $(0, \mathbf{D}\,\eta_\nu^*)$.

If $c > 0$, the distribution of (3.12) converges to a distribution concentrated in the points 0 and $\dfrac{\varepsilon}{\sqrt{c(1-c)}}$. If (3.13) had asymptotically normal distribution, this distribution could be decomposed in a convolution of two distributions one of which is not normal. This is, however, not possible, in view of the well-known theorem by H. CRAMÉR.

The theorem is completely proved.

**Remark 3.1.** In paper [2] there is proved that, provided we have a fixed double sequence $\{N_\nu, y_{\nu i}\}$, the Lindeberg condition (3.2) is fulfilled for *any* sequence $\{n_\nu\}$, such that $n_\nu \to \infty$ and $N_\nu - n_\nu \to \infty$, if and only if the relation

$$\lim_{\nu \to \infty} \frac{\sum\limits_{i \in s_{r_\nu}} (y_{\nu i} - \overline{Y}_\nu)^2}{\sum\limits_{i \in S_\nu} (y_{\nu i} - \overline{Y}_\nu)^2} = 0 \tag{3.14}$$

holds for any sequence $\{s_{r_\nu}\}$ such that $s_{r_\nu} \subset S_\nu$ and

$$\lim_{\nu \to \infty} \frac{r_\nu}{N_\nu} = 0 \tag{3.15}$$

where $r_\nu$ denotes the number of elements in $s_{r_\nu}$.

**Remark 3.2.** According to a theorem by CRAMÉR, ([5] p. 105.) a vector $(\xi_{\nu 1}, \ldots, \xi_{\nu m})$ has a $m$-dimensional normal limit distribution with parameters $\{\mathsf{E}\, \xi_{\nu j}, \operatorname{Cov}(\xi_{\nu j}, \xi_{\nu h}), h, j = 1, \ldots, m\}$ if any linear combination $\sum_{j=1}^{m} \lambda_j \xi_{\nu j}$ has a one-dimensional normal limit distribution with respective parameters. Let $\xi_{\nu j}$ be given by (1.3) where $y_i = y_{\nu j i}$, where $\nu$ labels the experiment and $j$ the variable. Suppose that the sequences $\{y_{\nu j i}, n_\nu, N_\nu\}$, $j = 1, \ldots, m$, fulfil the condition (3.2) and that the multiple correlation coefficients $\varrho_{\nu j}$ between $\xi_{\nu j'}$ and $\{\xi_{\nu j'}\, j' \neq j\}$ are uniformly bounded from 1, i. e. that

$$\lim_{\nu\to\infty} \sup \varrho^2_{\nu j} < 1 \qquad j = 1, \ldots, m. \tag{3.16}$$

Then any sequence $\left\{\sum_{j=1}^{m} \lambda_j y_{\nu j i}, n_\nu, N_\nu\right\}$, where $\lambda_j$ are arbitrary constants, fulfils the condition (3.2) and hence the random vector $(\xi_{\nu 1}, \ldots, \xi_{\nu m})$ has asymptotically normal $m$-dimendional distribution with respective parameters. Actually, we have

$$\mathsf{D}\left(\sum_{j=1}^{m} \lambda_j \xi_{\nu j}\right) \geqq \max_{1 \leqq j \leqq m} (1 - \varrho^2_{\nu j})\, \lambda_j^2\, \mathsf{D}\, \xi_{\nu j} \tag{3.17}$$

and

$$\left|\sum_{j=1}^{m} \lambda_j (y_{\nu j i} - \overline{Y}_{\nu j})\right| \leqq m \max_{1 \leqq j \leqq m} |\lambda_j|\, |y_{\nu j i} - Y_{\nu j}|\,. \tag{3.18}$$

The rest follows by easy computations.

**Remark 3.3.** If $\frac{n_\nu}{N_\nu}$ is bounded from 0 and 1, i. e.

$$0 < \varepsilon < \frac{n_\nu}{N_\nu} < 1 - \varepsilon \qquad (\nu \geqq \nu_0) \tag{3.19}$$

then the Lindeberg condition (3.2) reduces to the Noether condition (3.5). Really, if (3.5) and (3.19) are satisfied, then the subset $S_{\nu\tau}$ is empty for all sufficiently large $\nu$ so that (3.2) clearly holds. If (3.5) is not satisfied, (3.2) does not hold in any case.

## 4. Convergence to the Poisson distribution

Using the same method as in Section 3, the following theorem will be proved:

**Theorem 4.1.** *Suppose that* $n_\nu \to \infty$, $n_\nu \leqq \frac{1}{2} N_\nu$, *and*

$$\lim_{\nu\to\infty} \mathsf{E}\, \xi_\nu = \lim_{\nu\to\infty} \mathsf{D}\, \xi_\nu = \lambda > 0\,. \tag{4.1}$$

*Then the relation* (1.7) *is fulfilled if and only if, first*

$$\lim_{\nu\to\infty} \frac{n_\nu}{N_\nu} = 0 \tag{4.2}$$

*and, second,*

$$\lim_{\nu\to\infty} \frac{n_\nu}{N_\nu} \sum_{|y_{\nu i}-1|>\tau} y_{\nu i}^2 = 0 \qquad \text{for any } \tau > 0. \tag{4.3}$$

**Proof.** First assume that the infinitesimality condition (3.5) holds. If (4.2) does not hold, then, in view of Remark 3.3, the limiting distribution may be only normal. Consequently, the condition (4.2) is necessary. Now $\eta^*_\nu$ has limiting Poisson distribution if and only if

$$\lim_{\nu\to\infty} \sum_{i=1}^{N_\nu} \int\limits_{|x-1|>\tau} x^2 \, d\,\mathbf{P}\{\zeta_{\nu i} - \mathbf{E}\,\zeta_{\nu i} < x\} = 0\,. \tag{4.4}$$

We may write, as in (3.6),

$$\begin{aligned}\sum_{i=1}^{N_\nu} \int\limits_{|x-1|>\tau} x^2 \, d\,\mathbf{P}\{\zeta_{\nu i} - \mathbf{E}\,\zeta_{\nu i} < x\} &= \frac{n_\nu}{N_\nu}\left(1-\frac{n_\nu}{N_\nu}\right)^2 \sum_{i\in C'_{\nu\tau}} (y_{\nu i} - \overline{Y}_\nu)^2 + \\ &+ \left(1-\frac{n_\nu}{N_\nu}\right)\left(\frac{n_\nu}{N_\nu}\right)^2 \sum_{i\in B'_{\nu\tau}} (y_{\nu i} - \overline{Y}_\nu)^2\end{aligned} \tag{4.5}$$

where $C'_{\nu\tau}$ and $B'_{\nu\tau}$ are subsets of elements of $S_\nu$ on which

$$C'_{\nu\tau}: \quad \left| (y_{\nu i} - \overline{Y}_\nu)\left(1-\frac{n_\nu}{N_\nu}\right) - 1 \right| > \tau \tag{4.6}$$

and

$$B'_{\nu\tau}: \quad \left| -(y_{\nu i} - \overline{Y}_\nu)\frac{n_\nu}{N_\nu} - 1 \right| > \tau\,. \tag{4.7}$$

In view of (4.1), it holds that

$$\lim_{\nu\to\infty} \frac{n_\nu}{N_\nu} \sum_{i=1}^{N_\nu} y_{\nu i} = \lim_{\nu\to\infty} n_\nu \overline{Y}_\nu = \lambda\,, \qquad \text{i.e. } \lim_{\nu\to\infty} \overline{Y}_\nu = 0 \tag{4.8}$$

and

$$\lim_{\nu\to\infty} \left(1-\frac{n_\nu}{N_\nu}\right) \frac{n_\nu}{N_\nu} \sum_{i=1}^{N_\nu} (y_{\nu 1} - \overline{Y}_\nu)^2 = \lambda\,. \tag{4.9}$$

Concequently, in accordence with (4.2),

$$\lim_{\nu\to\infty} \frac{\displaystyle\sum_{i=1}^{N_\nu} \int\limits_{|x-1|>\tau} x^2 \, d\,\mathbf{P}\{\zeta_{\nu i} - \mathbf{E}\,\zeta_{\nu i} < x\}}{\displaystyle\frac{n_\nu}{N_\nu} \sum_{|y_{\nu i}-1|>\tau} y_{\nu i}^2} = 1\,,$$

which proves the equivalency of conditions (4.3) and (4.4).

As for the case when the condition (3.5) is not fulfilled, we could prove, as in the proof of Theorem 3.1 that the limiting distribution cannot preserve variance.

**Remark 4.1.** If the sampling were done with replacement (i. e. as $n$ independent drawings of one element) we would get just the conditions (4.2) and (4.3) for asymptotic Poisson distribution of the sum of selected values. This coincidence is clearly caused by the fact that the difference between with and without replacement sampling becomes negligible if $\frac{n_\nu}{N_\nu} \to 0$.

### 5. Other cases

Developing the basic idea further, we get

**Theorem. 5.1.** *Suppose that*

$$\lim_{\nu \to \infty} \mathbf{E}\, \xi_\nu = \mu \tag{5.1}$$

*and*

$$\lim_{\nu \to \infty} \mathbf{D}\, \xi_\nu = \sigma^2 \tag{5.2}$$

*and consider an infinitely divisible law — distinct from normal law — with mean value* $\mu$, *variance* $\sigma^2$ *and cumulant — generating function*

$$i\,\mu\,t + \int (e^{itu} - 1 - itu) \frac{1}{u^2}\, d\,K(u)\,. \tag{5.3}$$

*Then the distribution of* $\xi_\nu$ *converges to the law given by* (5.3) *if and only if*

$$\lim_{\nu \to \infty} \frac{n_\nu}{N_\nu} = 0 \tag{5.4}$$

*and*

$$\lim_{\nu \to \infty} \frac{n_\nu}{N_\nu} \sum_{y_{\nu i} < u} y_{\nu i}^2 = K(u) \tag{5.5}$$

*in all continuity points of* $K(u)$.

**Proof.** The same as of Theorem 4.1.

### 6. Conclusions

If $\frac{n_\nu}{N_\nu}$ does not converge to 0, normal limiting distribution is possible, namely under conditions established in Theorem 3.1. We can also get a limiting distribution formed as a convolution of a normal distribution and some two-points distributions.

If $\frac{n_\nu}{N_\nu}$ converges to 0, the variance preserving limiting distribution may be only infinitely divisible. The conditions for this are the same as if the sampling were carried out with replacement.

(Received May 7, 1960.)

REFERENCES

[1] Erdős, P. and Rényi, A.: „On a central limit theorem for samples from a finite population." *Publ. Math. Inst. Hung. Acad. Sci.* **4** (1959) 49—61.
[2] Hájek, J.: „Some exsensions of the Wald—Wolfowitz—Noether theorem." (In print.)
[3] Гнеденко, Б. В. и Колмогоров, А. Н.: *Предельные распределения для сумм независимых случайных величин.* Гостехиздат, Москва, 1949.
[4] Madow, W. G.: „On the limiting distributions of estimates based on samples from finite universes." *Ann. Math. Stat.* **19** (1948) 535—545.
[5] Cramér, H.: *Random variables and probability distributions.* Cambridge University Press, London, 1937.

# ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ПРИ ПРОСТОЙ СЛУЧАЙНОЙ ВЫБОРКЕ ИЗ КОНЕЧНОЙ СОВОКУПНОСТИ

J. HÁJEK

## Резюме

В статье вероятностная выборка из конечной совокупности рассматривается как случайный опыт, при котором из множества $S = \{1, 2, \ldots, N\}$ выбирается подмножество $s$, $s \subset S$. Множество $S$ и случайное подмножество $s$ называем соответственно основной совокупностей и выборочной совокупностей. Обозначим $s$ составляющее из $k$ элементов через $s_k$ и вероятность $s_k$ через $\mathbf{P}(s_k)$.

При простой случайной выборке (без возвращения), объема $n$, мы имеем

$$\mathbf{P}(s_k) = \begin{cases} \binom{N}{n}^{-1} & \text{для} \quad k = n \\ 0 & \text{для} \quad k \neq n. \end{cases} \tag{1.1}$$

В статьи показывается, что задача о предельных распределениях при предположении (1.1) сводится к той же задаче при предложении

$$\mathbf{P}(s_k) = \left(\frac{n}{N}\right)^k \left(1 - \frac{n}{N}\right)^{N-k} \quad \text{для} \quad 0 \leqq k \leqq N. \tag{1.2}$$

Вероятностную выборку с вероятностями (1.2) называем пуассоновской выборкой. Эту выборку возможно понимать, во первых, как простую случайную выборку объема $k$, при чем $k$ есть случайная величина с биномиальным законом распределения с вероятностями $\binom{N}{k}\left(\frac{n}{N}\right)^k \left(1 - \frac{n}{N}\right)^{N-k}$, или во вторых, как $N$ независимых опытов таких, что при $i$-том опыте элемент $i$ включен во выборочную совокупность с вероятностью $\frac{n}{N}$ и не включен с вероятностью $1 - \frac{n}{N}$.

Пусть $y_1, \ldots, y_N$ — вещественные числа; положим

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

и далее

(2.2) $$\eta = \sum_{i \in s_n} (y_i - \overline{Y})$$

и

(2.3) $$\eta^* = \sum_{i \in s_k} (y_i - Y),$$

где $s_n$ и $s_k$ представляют соответственно результат простой случайной вы, борки и пуассоновской выборки. В пар. 2 показывается, что выборки $(s_n, s_k)$ возможно осуществить современно таким образом, что $s_n \subset s_k$, если $n \leqq k$, и $s_n \supset s_k$, если $n \geqq k$, и

(2.5) $$\frac{\mathbf{E}(\eta - \eta^*)^2}{\mathbf{D}\,\eta^*} \leqq \sqrt{\frac{1}{n} + \frac{1}{N-n}},$$

где $\mathbf{E}(\cdot)$ обозначает среднюю и $\mathbf{D}(\cdot)$ дисперсию.

Теперь рассмотрим последовательность $\{y_{\nu i}, n_\nu, N_\nu\}_{\nu=1}^{\infty}$ основных совокупностей объема $N_\nu$ со значениями $y_{\nu i}$ $(i = 1, \ldots, N_\nu)$, простых случайных выборек объема $n_\nu$ и сопровождающих их пуассоновских выборек. Если предположим, что $n_\nu \to \infty$ и $N_\nu - n_\nu \to \infty$, и обозначим через $\eta_\nu$ и $\eta_\nu^*$ случайные величины (2.2) и (2.3) относящиеся к $\nu$-тому члену нашей последовательности, то из (2.5) следует, что

(2.12) $$\lim_{\nu \to \infty} \frac{\mathbf{E}(\eta_\nu - \eta_\nu^*)^2}{\mathbf{D}\,\eta^*} = 0.$$

Значит, предельные дисперсии и распределения случайных величин $A_\nu + B_\nu \eta_\nu$ и $A_\nu + B_\nu \eta_\nu^*$ ($A_\nu$, $B_\nu$ — любие постоянные) существуют при тех же условиях, и если они существуют, то они совпадают, друг с другом. Но случайная величина $\eta^*$ равняется сумме $N$ не зависимых случайных величин $\zeta_1, \ldots, \zeta_N$,

(2.9) $$\eta^* = \sum_{i=1}^{N} \zeta_i$$

определенных так, что

(2.10) $$\zeta_i = \begin{cases} y_i - \overline{Y} & \text{если} \quad i \in s_k \\ 0 & \text{если} \quad i \notin s_k. \end{cases}$$

Таким образом мы свели задачу о предельном распределении случайной величины $\eta$ к той же самой задаче о сумме независимых слагаемых $\eta^*$. В результате применения этого простого факта, получаются следующие теоремы:

**Теорема 3.1.** *Пусть $S_{\nu\tau}$ — подмножество множества $S_\nu = \{1, \ldots, N_\nu\}$*

*состоящее из элементов для которых имеем*

$$|y_{\nu i} - \overline{Y}_\nu| > \tau \sqrt{\mathbf{D}\,\eta_\nu}. \tag{3.1}$$

*Предположим что* $n_\nu \to \infty$ *и* $N_\nu - n_\nu \to \infty$.

*Потом случайная величина* (2.2) *имеет предельное нормальное распределение с параметрами* ($\mathbf{E}\,\eta_\nu$, $\mathbf{D}\,\eta_\nu$) *тогда и только тогда, если*

$$\lim_{\nu\to\infty} \frac{\sum\limits_{i\in S_{\nu\tau}} (y_{\nu i} - \overline{Y}_\nu)^2}{\sum\limits_{i=1}^{N_\nu} (y_{\nu i} - \overline{Y}_\nu)^2} \qquad \text{при всяком } \tau > 0. \tag{3.2}$$

Достаточность условия (3.2) была в первые доказана другим методом в работе [1].

**Теорема 4.1** *Предлположим, что* $n_\nu \to \infty$, $n_\nu \leqq \frac{1}{2} N_\nu$ *и*

$$\lim_{\nu\to\infty} \mathbf{E}\Big(\sum_{i\in s_{n_\nu}} y_{\nu i}\Big) = \lim_{\nu\to\infty} \mathbf{D}\Big(\sum_{i\in s_{n_\nu}} y_{\nu i}\Big) = \lambda > 0\,. \tag{4.1}$$

*Потом соотношение*

$$\lim_{\nu\to\infty} \mathbf{D}\Big\{\sum_{i\in s_{n_\nu}} y_{\nu i} = k\Big\} = e^{-\lambda}\frac{\lambda^k}{k!} \qquad (k = 0, 1, 2, \ldots) \tag{1.7}$$

*имеет место тогда и только тогда, если*

$$\lim_{\nu\to\infty} \frac{n_\nu}{N_\nu} = 0 \tag{4.2}$$

*и*

$$\lim_{\nu\to\infty} \frac{n_\nu}{N_\nu} \sum_{|y_{\nu i} - 1| > \tau} y_{\nu i}^2 = 0 \qquad \text{при всяком } \tau > 0. \tag{4.3}$$

**Теорема 5.1** *Предлположим что*

$$\lim_{\nu\to\infty} \mathbf{E}\Big(\sum_{i\in s_{n_\nu}} y_{\nu i}\Big) = \mu \tag{5.1}$$

*и*

$$\lim_{\nu\to\infty} \mathbf{D}\Big(\sum_{i\in s_{n_\nu}} y_{\nu i}\Big) = \sigma^2 \tag{5.2}$$

*и рассматрим бесконечно-делимый закон, отличный от нормального и имеющий среднюю* $\mu$, *дисперсию* $\sigma^2$ *и логарифм характеристической функции*

$$i\,\mu t + \int_{-\infty}^{\infty} (e^{itu} - 1 - itu)\,\frac{1}{u^2}\,d\,K(u)\,. \tag{5.3}$$

*Это распределение является предельным для случайной величины* $\sum_{i \in s_{n_\nu}} y_{\nu i}$ *тогда и только тогда, если*

$$\lim_{\nu \to \infty} \frac{n_\nu}{N_\nu} = 0 \tag{5.4}$$

*и*

$$\lim_{\nu \to \infty} \frac{n_\nu}{N_\nu} \sum_{y_{\nu i} < u} y_{\nu i}^2 = K(u) \tag{5.5}$$

*во всех точках непрерывности* $K(u)$.