A CONVERGENCE THEOREM FOR DISCRETE PROBABILITY DISTRIBUTIONS

By MASASHI OKAMOTO

(Received Nov. 24, 1959)

1. Summary.

A convergence theorem is given which is useful in proving that a sequence of real-valued, discrete probability distributions converges to a continuous distribution. It is a modification of Scheffé's "useful" convergence theorem [1]. As examples the hypergeometric distribution and the conditional distribution in the contingency table are proved to be asymptotically normal.

2. Theorem.

At first the theorem is stated in the one-dimensional form and later generalized to the multi-dimensional case. Let $\{X_n\}_{(n=1,2,\cdots)}$ be a sequence of integral-valued, discrete random variables taking the value r ($r=0, \pm 1, \pm 2, \cdots$) with the probability $p_n(r)$, $\sum_{r=-\infty}^{\infty} p_n(r)=1$.

Theorem. *If there exist a sequence* $\{\mu_n\}$ *of real numbers and a pdf* $f(x)$ (*with regard to the Lebesgue measure*) *such that*

$$(1) \qquad \lim_{n\to\infty} \sqrt{n}\, p_n([\mu_n+\sqrt{n}\,x])=f(x) \qquad \textit{for any real } x\,,$$

brackets denoting Gauss symbol, then $(X_n-\mu_n)/\sqrt{n}$ *converges in law as* $n\to\infty$ *to the distribution with pdf* $f(x)$.

PROOF. Put for each n

$$f_n(x)=\sqrt{n}\, p_n([\mu_n+\sqrt{n}\,x]) \qquad -\infty<x<\infty\,.$$

Then it is easily seen that $f_n(x)$ is a pdf (w.r.t. then Lebesgue measure). Since the assumption (1) implies $\lim_{n\to\infty} f_n(x)=f(x)$ for any real x, we obtain from Scheffé's theorem [1] that

$$\lim_{n\to\infty}\int_{-\infty}^{a} f_n(x)dx=\int_{-\infty}^{a} f(x)dx \qquad \text{for any real } a\,.$$

For a given $\varepsilon>0$, if n is sufficiently large such as $\sqrt{n}\,\varepsilon\geqq 1$, then we have

$$\int_{-\infty}^{a} f_n(x)dx \leqq \sum_{r=-\infty}^{[\mu_n \sqrt{n}\, a]} p_n(r) \leqq \int_{-\infty}^{a+\varepsilon} f_n(x)dx \ .$$

Since

$$\begin{aligned}\sum_{r=-\infty}^{[\mu_n+\sqrt{n}\, a]} p_n(r) &= P_r(X_n \leqq [\mu_n + \sqrt{n}\, a]) \\ &= P_r(X_n \leqq \mu_n + \sqrt{n}\, a) = P_r((X_n - \mu_n)/\sqrt{n} \leqq a) \ ,\end{aligned}$$

it holds that

$$\lim_{n\to\infty} P_r\left(\frac{X_n-\mu_n}{\sqrt{n}} \leqq a\right) \int_{-\infty}^{a} f(x)dx \qquad \text{for any real a ,}$$

which proves the theorem.

Now the theorem can be generalized to the multi-dimensional case. Let $\{\boldsymbol{x}_n\}_{(n=1,2,\cdots)}$ be a sequence of k-dimensional random variables of the lattice type: there exist points $\boldsymbol{x}_{0n}$ $(n=1, 2, \cdots)$ in R^k and a system $\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k)$ of linearly independent vectors in R^k such that $\boldsymbol{x}_n$ takes only values of the form $\boldsymbol{x}_{0n} + \sum_{i=1}^{k} r_i \boldsymbol{a}_i$, r_i being integers, with the probability $p_n(\boldsymbol{x}_{0n} + \sum_{i=1}^{k} r_i \boldsymbol{a}_i)$. Let v be the volume of a unit lattice generated by $\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k\}$ and let the symbol $[\ ]_n$ be defined as follows: if a vector $\boldsymbol{x}$ is represented as $\boldsymbol{x} = \boldsymbol{x}_{0n} + \sum_{i=1}^{k} t_i \boldsymbol{a}_i$, t_i being real numbers, then $[\boldsymbol{x}_n] = \boldsymbol{x}_{0n} + \sum_{i=1}^{k} [t_i] \boldsymbol{a}_i$, $[\ ]$ denoting the ordinary Gauss symbol.

Theorem (generalized). *If there exist a sequence* $\{\mu_n\}$ *of points in* R^k *and a pdf* $f(\boldsymbol{x})$ *(with regard to the Lebesgue measure in* R^k*) such that*

$$\text{(1')} \qquad \lim_{n\to\infty} v^{-1} n^{k/2} p_n([\mu_n + \sqrt{n}\, \boldsymbol{x}]_n) = f(\boldsymbol{x}) \quad \textit{for any } \boldsymbol{x} \textit{ in } R^k \ ,$$

then $(\boldsymbol{x}_n - \mu_n)/\sqrt{n}$ *converges in law as* $n\to\infty$ *to the distribution with pdf* $f(\boldsymbol{x})$.

The proof is quite analogous to the preceding one and is omitted.

REMARK. The theorem can be generalized further by permitting the lattice form to depend on n, where it is required only to a assume that the unit lattices for $n=1, 2, \cdots$ are bounded. On the other hand it might happen that our theorem is a known result, considering the simplicity of its proof. We could not, however, find in literature an explicit statement of relevant conditions.

3. Examples.

Example 1 (Hypergeometric distribution)

Let the random variable X_N be defind by

$$P_r(X_N=r)=\binom{D}{r}\binom{N-D}{n-r}\Big/\binom{N}{n}, \qquad 0, n+D-N\leqq r\leqq N, D ,$$

where N, D and n are positive integers and $D, n\leqq N$.

We consider the limiting condition $N, D, n\to\infty$ under the restriction that

$$D=Np+o(\sqrt{N}) \text{ and } n=Np'+o(\sqrt{N}) ,$$

where p and p' are constants, with $0<p<1$ and $0<p'<1$. Putting

$$\mu_N=Npp' ,$$

$$f(x)=(2\pi pp'qq')^{-1/2}\exp[-x^2/(2pp'qq')] ,$$

where $q=1-p$ and $q'=1-p'$, we can verify the assumption (1) of the theorem by use of the Stirling formula and therefore X_N is asymptotically normal with mean Npp' and variance $Npp'qq'$.

Example 2 (Contingency table)

Let the random variable $n_{ij}(i=1, \cdots, r ; j=1, \cdots, s)$ follow the multinomial distribution $(n, \{p'_{ij}\})$, that is

$$P(n_{ij})=\frac{n\,!}{\Pi_i\Pi_j n_{ij}\,!}\Pi_i\Pi_j(p'_{ij})^{n_{ij}} .$$

Put $p'_{i\cdot}=\sum_{j=1}^{s}p'_{ij}$, $p'_{\cdot j}=\sum_{i=1}^{r}p'_{ij}$, $n_{i\cdot}=\sum_{j=1}^{s}n_{ij}$ and $n_{\cdot j}=\sum_{i=1}^{r}n_{ij}$. If it holds that $p'_{ij}=p'_{i\cdot}p'_{\cdot j}$ for any i and j or if row-and column-classifications are independent, then the conditional distribution of $\{n_{ij}\}$ given the marginal totals $\{n_{i\cdot}, n_{\cdot j}\}$ is well-known to be

$$(2) \qquad P(n_{ij}\,|\,n_{i\cdot}, n_{\cdot j})=\frac{\Pi_i n_{i\cdot}\,!\,\Pi_j n_{\cdot j}\,!}{n\,!\,\Pi_i\Pi_j n_{ij}\,!}$$

where independent variables are those $(r-1)(s-1)$ among n_{ij}'s which are linearly independent of $n_{i\cdot}$ and $n_{\cdot j}$. Suppose now that they are n_{ij} $(i=1, \cdots, r-1 ; j=1, \cdots, s-1)$.

We shall consider the asymptotic behavior of the distribution (2) when $n, n_{i\cdot}, n_{\cdot j}\to\infty$ under the restriction that

(3) $\quad n_{i\cdot}=np_{i\cdot}+o(\sqrt{n}),\ n_{\cdot j}=np_{\cdot j}+o(\sqrt{n}), \quad (i=1, \cdots, r\,; j=1, \cdots, s)$,

where $p_{i\cdot}$ and $p_{\cdot j}$ are positive constants with $\sum_i p_{i\cdot}=\sum_j p_{\cdot j}=1$. Put

(4) $$\mu_{ijn}=np_{i\cdot}p_{\cdot j} \qquad (i=1, \cdots, r\,; j=1, \cdots, s)$$

and

(5) $$f(x_{ij})=(2\pi)^{-(r-1)(s-1)/2}\Big(\prod_{i=1}^{r} p_{i\cdot}\Big)^{-(s-1)/2}\Big(\prod_{j=1}^{s} p_{\cdot j}\Big)^{-(r-1)/2} \times \exp\Big(-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{x_{ij}^2}{p_{i\cdot}p_{\cdot j}}\Big),$$

where independent variables are x_{ij} $(i=1, \cdots, r-1\,; j=1, \cdots, s-1)$ in accordance with n_{ij} in (2) and other x_{ij}'s are defined by $x_{i\cdot}=\sum_{j=1}^{s} x_{ij}=0$, $x_{\cdot j}=\sum_{i=1}^{r} x_{ij}=0$ for any i and j. Then it can be shown after some manipulation that the assumption (1′) of the generalized theorem is satisfied.

It now remains to show that $f(x_{ij})$ defined in (5) is really a *pdf* with $(r-1)(s-1)$ dimensions. This is seen as follows: Let the random variable X_{ij} $(i=1, \cdots, r\,; j=1, \cdots, s)$ follow independently the normal distribution $N(0, p_{i\cdot}p_{\cdot j})$. Their joint *pdf* is then written as

$$p(x_{ij})=(2\pi)^{-rs/2}(\Pi_i p_{i\cdot})^{-s/2}(\Pi_j p_{\cdot j})^{-r/2}\exp\Big(-\frac{1}{2}\sum_i\sum_j\frac{x_{ij}^2}{p_{i\cdot}p_{\cdot j}}\Big).$$

Put $X_{i\cdot}=\sum_j X_{ij}$, $X_{\cdot j}=\sum_i X_{ij}$, $X_{\cdot\cdot}=\sum_i\sum_j X_{ij}$, $Y_i=X_{i\cdot}-p_{i\cdot}X_{\cdot\cdot}$ and $Z_j=X_{\cdot j}-p_{\cdot j}X_{\cdot\cdot}$. Since $(Y_1, \cdots, Y_{r-1})$, $(Z_1, \cdots, Z_{s-1})$, and $X_{\cdot\cdot}$ are mutually independent, their joint *pdf* is

$$p(y_i, z_j, x_{\cdot\cdot})=(2\pi)^{-(r+s-1)/2}(\Pi_i p_{i\cdot})^{-1/2}(\Pi_j p_{\cdot j})^{-1/2}\exp\Big(-\frac{1}{2}Q\Big),$$

where

$$Q=\sum_i\frac{y_i^2}{p_{i\cdot}}+\sum_j\frac{z_j^2}{p_{\cdot j}}+x_{\cdot\cdot}^2=\sum_i\frac{x_{i\cdot}^2}{p_{i\cdot}}+\sum_j\frac{x_{\cdot j}^2}{p_{\cdot j}}-x_{\cdot\cdot}^2 .$$

Hence by division the conditional distribution of X_{ij}, given $Y_i=y_i$, $Z_j=z_j$ and $X_{\cdot\cdot}=x_{\cdot\cdot}$ or in other word given $X_{i\cdot}=x_{i\cdot}$ and $X_{\cdot j}=x_{\cdot j}$ is represented by the normal *pdf*

$$p(x_{ij}\,|\,x_{i\cdot}, x_{\cdot j})=(2\pi)^{-(r-1)(s-1)/2}(\Pi_i p_i)^{-(s-1)/2}(\Pi_j p_j)^{-(r-1)/2} \times\exp\Big[-\frac{1}{2}\Big(\sum_i\sum_j\frac{x_{ij}^2}{p_{i\cdot}p_{\cdot j}}-\sum_i\frac{x_{i\cdot}^2}{p_{i\cdot}}-\sum_j\frac{x_{\cdot j}^2}{p_{\cdot j}}+x_{\cdot\cdot}^2\Big)\Big].$$

When $x_{i\cdot}=x_{\cdot j}=x_{\cdot\cdot}=0$ this conincides with (5).

Consequently the conditional distribution (2) is asymptotically normal with mean (4) and standardized *pdf* (5). This fact is stated by Rao [2] in a heuristic manner, though not rigorously. This together with Mann-Wald-Chernoff's theorem ([4], theorem 2) implies that given $\{n_{i\cdot}, n_{\cdot j}\}$, the statistic

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{s}\left(n_{ij}-\frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2 \Big/ \frac{n_{i\cdot}n_{\cdot j}}{n}$$

follows conditionally asymptotically χ^2 distribution with $(r-1)(s-1)$ degrees of freedom.

Example 3 (Intraclass contingency table)

Let the random variable $\{n_{ii}, n_{ij}\}(i, j=1, \cdots, k\,; i<j)$ follow the multinomial distribution $(n, \{p_i'^2, 2p_i'p_j'\})$, where p_i' $(i=1, \cdots, k)$ are positive constants, $\sum_i p_i'=1$. The intraclass contingency table is referred to by the present author and G. Ishii [3] in contrast to the ordinary contingency table. Put $n_{ji}=n_{ij}$ $(i<j)$ and $n_i=2n_{ii}+\sum_{j\neq i} n_{ij}$. It is proved in [3] that the conditional distribution of $\{n_{ii}, n_{ij}\}$ given the "marginal totals" n_i $(i=1, \cdots, k)$ is

$$P(n_{ii}, n_{ij}\,|\,n_i)=\frac{2^{\Sigma\Sigma_{i<j}n_{ij}}n\,!\,\Pi_i n_i\,!}{(2n)\,!\,\Pi_i n_{ii}\,!\,\Pi\Pi_{i<j}n_{ij}\,!}\,, \tag{6}$$

where independent variables are those $k(k-1)/2$ among $\{n_{ii}, n_{ij}\}$ which are linearly independent of n_i. Suppose that they are n_{ij} $(i, j=1, \cdots, k\,; i<j)$. Then the distribution (6) is of the lattice type of which a set of generating vectors is given by

$$\left[\begin{array}{c} \cdot \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_{k-1} \\ \hline a_k \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_{k(k-1)/2} \end{array}\right] = \left[\begin{array}{cccccc|cccccc} 2 & & & & & 0 & & & & & & \\ & \cdot & & & & & & & & & & \\ & & \cdot & & & & & & 0 & & & \\ & & & \cdot & & & & & & & & \\ & & & & \cdot & & & & & & & \\ & & & & & 2 & & & & & & \\ \hline 1 & 1 & 0 & \cdots & 0 & 0 & -1 & & & & & \\ 1 & 0 & 1 & \cdots & 0 & 0 & & \cdot & & & 0 & \\ & \cdot & \cdot & \cdot & \cdot & & & & \cdot & & & \\ & \cdot & \cdot & \cdot & \cdot & & & & & \cdot & & \\ 1 & 0 & 0 & \cdots & 0 & 1 & & & -1 & & & \\ 0 & 1 & 0 & \cdots & 0 & 1 & & & & \cdot & & \\ & \cdot & \cdot & \cdot & \cdot & & & & & & \cdot & \\ & \cdot & \cdot & \cdot & \cdot & & & & & & & \cdot \\ 0 & 0 & 0 & \cdots & 1 & 1 & & & & & & -1 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\end{array}}\right\} k-1 \\ \left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\\1\\1\\1\end{array}}\right\} (k-1)(k-2)/2 \end{array}$$

where coordinates relate to n_{ij} arranged in the lexicographic order with regard to i and j $(i, j=1, \cdots, k\,;\, i<j)$, whence v, the volume of a unit lattice, is 2^{k-1}.

We shall consider the limiting condition $n, n_i \to \infty$ under the restriction $n_i = 2np_i + o(\sqrt{n})$, where $p_i (i=1, \cdots, k)$ are positive constants, $\sum_i p_i = 1$. Put

$$\mu_{iin} = np_i^2, \qquad \mu_{ijn} = 2np_ip_j\,, \tag{7}$$

and

$$f(x_{ii}, x_{ij}) = (2\pi)^{-k(k-1)/2}2^{-(k-1)(k+2)/2}(\Pi_i p_i)^{-k/2} \times \exp\left[-\frac{1}{2}\left(\sum_i \frac{x_{ii}^2}{p_i^2} + \sum\sum_{i<j} \frac{x_{ij}^2}{2p_ip_j}\right)\right], \tag{8}$$

where independent variable are x_{ij} $(i, j=1, \cdots, k\,;\, i<j)$, x_{ij} defined by $x_i = 2x_{ii} + \sum_{j\neq i} x_{ij} = 0$, $x_{ji} = x_{ij} (i<j)$.

Now it can be shown as in Example 2 that the equation (1′) of the generalized theorem is satisfied and that the function f defined in (8) is really a normal *pdf*. The conditional distribution (6) is therefore asymptotically normal with mean (7) and the standardized *pdf* (8), while the conditional distribution of the statistic $\chi^2 = \sum\sum_{i\leq j} (n_{ij} - E(n_{ij}))^2/E(n_{ij})$ is asymptotically a χ^2 distribution with $k(k-1)/2$ degrees of freedom.

OSAKA UNIVERSITY

REFERENCES

[1] Scheffé, H., "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.

[2] Rao, C. R., *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.

[3] Okamoto, M. and Ishii, G., "Independence in intraclass contingency tables," to be published.

[4] Chernoff, H., "Large-sample theory: parametric case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1–22.

Errata

Annals, Vol. X, No. 1.

Page	Line	Read	Instead of
48	13	$V_{(6)}=15$	$V_{(6)}=16$
"	14	(8 17)	(8 7)

Vol. X, No. 3.

Page	Line	Read	Instead of
200	17—18	$\cdots \sum_{j=0}^{\infty} \frac{\Gamma(p/2+1+j)}{j!}\left(\frac{\delta}{\gamma}\right)^{2j} \times\left[\left(\eta+\frac{p}{2}-j\right)\cdots\right.$	$\cdots \sum_{j=0}^{\infty} \frac{\Gamma(p/2+1+j)}{j!} \times\left[\left(\eta+\frac{p}{2}-j\right)\cdots\right.$

Vol. XI, No. 2.

Page	Line	Read	Instead of
108	16	$[X]_n$	$[X_n]$
112	11	while X_{ii} are defined by$\cdots$	X_{ij} defined by$\cdots$