NOTES

This section is devoted to brief research and expository articles and other short items.

CONVERGENCE OF DISTRIBUTIONS

By Herbert Robbins

University of North Carolina

Let $f_n(x)$ $(n = 0, 1, 2, \cdots)$ be frequency functions

$$f_n(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_n(x)\, dx = 1. \tag{1}$$

There are various ways in which the sequence of distributions corresponding to the $f_n(x)$ $(n = 1, 2, \cdots)$ may be said to converge to the distribution corresponding to $f_0(x)$. The definition customarily adopted in mathematical statistics (see e.g. [1]) is equivalent to the condition

$$\lim_{n \to \infty} \int_{-\infty}^{\xi} f_n(x)\, dx = \int_{-\infty}^{\xi} f_0(x)\, dx \qquad \text{for every } \xi.^1 \tag{a}$$

We shall also consider the two further conditions

$$\lim_{n \to \infty} \int_S f_n(x)\, dx = \int_S f_0(x)\, dx \qquad \text{for every Borel set } S, \tag{b}$$

and

$$\lim_{n \to \infty} \int_S f_n(x)\, dx = \int_S f_0(x)\, dx \qquad \text{uniformly for all Borel sets } S. \tag{c}$$

It is clear that (c) implies (b) and that (b) implies (a). That the converse implications do not hold is shown by the following examples.

Example 1. Let $f_0(x) = 1$ for $0 \leq x \leq 1$ and 0 elsewhere. Choose and fix any $0 < \epsilon < 1$, set $\delta_n = \epsilon/n \cdot 2^n$, and for $n = 1, 2, \cdots$ let $f_n(x) = 1/n \cdot \delta_n$ for $i/n - \delta_n \leq x \leq i/n$ $(i = 1, 2, \cdots, n)$ and 0 elsewhere. If we denote by S_n the set of all x for which $f_n(x) > 0$ it is easy to see that for $n = 1, 2, \cdots$

$$0 \leq \int_{-\infty}^{\xi} f_0(x)\, dx - \int_{-\infty}^{\xi} f_n(x)\, dx < 1/n \qquad \text{for every } \xi, \tag{2}$$

$$\int_{S_n} f_0(x)\, dx = \epsilon/2^n, \qquad \int_{S_n} f_n(x)\, dx = 1. \tag{3}$$

[1] From a well kown theorem of Pólya the convergence is then necessarily uniform for all ξ.

Hence for the Borel set $S = \sum_1^\infty S_n$ it follows that

$$\int_S f_0(x)\,dx \leq \sum_1^\infty \int_{S_n} f_0(x)\,dx = \epsilon, \tag{4}$$

$$\int_S f_n(x)\,dx = \int_{S_n} f_n(x)\,dx = 1, \qquad (n = 1, 2, \cdots). \tag{5}$$

From (2) we see that (a) holds (uniformly for all ξ), and from (4) and (5) that (b) fails about as badly as possible.

This construction can be modified to apply to any $f_0(x)$; thus choosing $f_0(x) = (2\pi e^{x^2})^{-1/2}$ we can construct $f_n(x)$ $(n = 1, 2, \cdots)$ and a Borel set S such that

$$\lim_{n\to\infty} \int_{-\infty}^{\xi} f_n(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} e^{-x^2/2}\,dx \qquad \text{uniformly for all } \xi,$$

while

$$\frac{1}{\sqrt{2\pi}} \int_S e^{-x^2/2}\,dx = .01, \qquad \int_S f_n(x)\,dx = 1, \qquad (n = 1, 2, \cdots).$$

It is conceivable that some time a statistician, failing to consider such a possibility, will be led to approximate .01 by 1.

If X_n is a random variable with frequency function $f_n(x)$, if $y = g(x)$ is a Borel function, and if (a) holds, then it follows from Example 1 that the distribution function $H_n(y)$ of $Y_n = g(X_n)$, equal to the integral of $f_n(x)$ over the set S_y of all x such that $g(x) \leq y$, need not converge to the distribution function $H_o(y)$ of $Y_o = g(X_o)$. It is easily seen that this possibility is excluded if, as commonly occurs in applications, $g(x)$ is such that for every y, the intersection of S_y with any finite interval is the sum of a finite number of intervals (e.g., if $g(x) = \sin x$).

Example 2. Let $f_0(x)$ be defined as in the previous example, and for $n = 1, 2, \cdots$ let $f_n(x) = 1 + \sin(2\pi n x)$ for $0 \leq x \leq 1$ and 0 elsewhere. By the Riemann-Lebesgue theorem it follows that (b) holds. But let S_n denote the set of all x for which $f_n(x) \geq 1$; then

$$\int_{S_n} f_0(x)\,dx = \tfrac{1}{2}, \qquad \int_{S_n} f_n(x)\,dx = \tfrac{1}{2} + 1/\pi, \qquad (n = 1, 2, \cdots),$$

so that (c) does not hold.

It follows from these examples that (a), (b), and (c) are successively stronger definitions of convergence. We shall now give some definitions equivalent to (b) and (c).

First we recall that the non-negative, completely additive, and absolutely continuous set functions

$$P_n(S) = \int_S f_n(x)\,dx, \qquad (n = 1, 2, \cdots), \tag{6}$$

are said to be uniformly absolutely continuous if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for any S and any $n = 1, 2, \cdots$,

$$m(S) < \delta \text{ implies } P_n(S) < \epsilon. \tag{7}$$

We shall denote the condition that the $P_n(S)$ be uniformly absolutely continuous by (u.a.c.), and we shall now prove that (b) is equivalent to

(b′) (a) and (u.a.c.).

PROOF. (A) Suppose (b) holds. It is clear that (a) holds, and we shall show by contradiction that (u.a.c.) holds also. For if not then there would exist an $\epsilon > 0$ such that for any $\eta > 0$ we could find a set S and an integer n such that

$$m(S) < \eta, \qquad P_n(S) \geq \epsilon. \tag{8}$$

Moreover, since the set function

$$P_0(S) = \int_S f_0(x)\, dx$$

is absolutely continuous, there exists a $\delta > 0$ such that

$$m(S) < \delta \quad \text{implies} \quad P_0(S) < \epsilon/2. \tag{9}$$

Now by (8) there exists an S_1 with $m(S_1) < \delta/2$ and a k_1 such that $P_{k_1}(S_1) \geq \epsilon$. Next, there exists an S_2 with $m(S_2) < \delta/2^2$ and a k_2 such that $P_{k_2}(S_2) \geq \epsilon$, and it is easy to see that we may assume that $k_2 > k_1$. Proceeding in this way we find a sequence of integers $k_1 < k_2 < \cdots$ and of sets $S_1, S_2, \cdots$ such that

$$m(S_n) < \delta/2^n, \qquad P_{k_n}(S_n) \geq \epsilon, \qquad (n = 1, 2, \cdots). \tag{10}$$

Let $S = \sum_1^\infty S_n$; then by (10), $m(S) \leq \sum_1^\infty m(S_n) < \delta$, so that by (9),

$$P_0(S) < \epsilon/2. \tag{11}$$

But by (10),

$$P_{k_n}(S) \geq P_{k_n}(S_n) \geq \epsilon, \qquad (n = 1, 2, \cdots). \tag{12}$$

From (11) and (12) we conclude that (b) does not hold, which is a contradiction. Hence (b) implies (b′).

(B) Suppose (b′) holds. We shall show first that (b) holds for any set S_1 of finite measure. Choose any $\epsilon > 0$; then from (u.a.c.) it follows that there exists a $\delta > 0$ such that

$$m(S) < \delta \text{ implies } P_n(S) < \epsilon/8 \qquad (n = 0, 1, 2, \cdots). \tag{13}$$

It is known from the theory of measure that corresponding to S_1 and to δ we can find a set S_2 which is the sum of a finite number of disjoint intervals, such that

$$m((S_1 - S_2) + (S_2 - S_1)) < \delta. \tag{14}$$

From (13), (14), and the relations

(15) $P_n(S_1) = P_n(S_2) + P_n(S_1 - S_2) - P_n(S_2 - S_1), \qquad (n = 0, 1, 2, \cdots),$

it follows that

$$(16) \quad \begin{aligned} | P_0(S_1) - P_n(S_1) | &\leq | P_0(S_2) - P_n(S_2) | + P_n(S_1 - S_2) + P_n(S_2 - S_1) \\ &+ P_0(S_1 - S_2) + P_0(S_2 - S_1) \leq | P_0(S_2) - P_n(S_2) | + \epsilon/2, \end{aligned}$$

and from (a) that for large enough n,

$$(17) \qquad | P_0(S_2) - P_n(S_2) | < \epsilon/2.$$

Thus from (16) and (17) it follows that for large enough n,

$$| P_0(S_1) - P_n(S_1) | < \epsilon,$$

which proves (b) for the case $m(S) < \infty$.

Now given any $\epsilon > 0$ choose α, β so that, setting $A = \{\alpha \leq x \leq \beta\}$, we have

$$(19) \qquad P_0(A) > 1 - \epsilon/4.$$

Then it follows from (a) that for large enough n,

$$(20) \qquad P_n(A) > 1 - \epsilon/2.$$

Then for any Borel set S we have for large enough n,

$$P_n(S) - P_0(S) = P_n(SA) + P_n(S - A) - P(SA) - P(S - A),$$

$$\begin{aligned} | P_n(S) - P_0(S) | &\leq | P_n(SA) - P_0(SA) | + P_n(S - A) + P_0(S - A) \\ &\leq | P_n(SA) - P_0(SA) | + \epsilon/2 + \epsilon/4. \end{aligned}$$

But by the previous case, since $m(SA) < \infty$, for large enough n we shall have $| P_n(SA) - P_0(SA) | < \epsilon/4$. Hence for large enough n,

$$| P_n(S) - P_0(S) | < \epsilon,$$

so that (b) holds in this case also. This completes the proof.

We shall say that $\lim_{n \to \infty} f_n(x) = f_0(x)$ in measure if for every $\epsilon > 0$ and for every set A such that $m(A) < \infty$, the measure of the set of all x in A for which $| f_n(x) - f_0(x) | > \epsilon$, tends to 0 as n increases. (For a space of finite measure this reduces to the usual definition.) We now observe that (c) is equivalent to

$$(c') \qquad \lim_{n \to \infty} f_n(x) = f_0(x) \qquad \text{in measure.}$$

In fact, it is easy to show that (c) is equivalent to convergence in the mean of order 1,

$$(c'') \qquad \lim_{n \to \infty} \int_{-\infty}^{\infty} | f_n(x) - f_0(x) | \, dx = 0,$$

which implies (c′), and a theorem of Scheffé [2] states that (c′) implies (c).[2] Finally, it is not hard to show that the condition

$$\text{(d)} \qquad \lim_{n \to \infty} f_n(x) = f_0(x) \qquad \text{almost everywhere}$$

implies (c′) but not conversely.

Summing up, we arrive at the following complete set of implication relations among the various modes of convergence which we have considered:

$$\text{(20)} \qquad \text{(d)} \rightarrow \text{(c}'') \rightleftarrows \text{(c}') \rightleftarrows \text{(c)} \rightarrow \text{(b}') \rightleftarrows \text{(b)} \rightarrow \text{(a)}.$$

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press. 1946, pp. 58–60.

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

[3] E. J. McShane, *Integration.* Princeton Univ. Press, 1944, p. 168.

ON RANDOM VARIABLES WITH COMPARABLE PEAKEDNESS

By Z. W. Birnbaum

University of Washington

The quality of a distribution usually referred to as its peakedness has often been measured by the fourth moment of the distribution. It is known, however, that there is no definite connection between the value of the fourth moment and what one may intuitively consider as the amount of peakedness of a distribution.[1] In the present paper a definition of relative peakedness is proposed and it is shown that this concept has properties which may make it practically applicable.

Definition. *Let Y and Z be real random variables and Y_1 and Z_1 real constants. We shall say that Y is more peaked about Y_1 than Z about Z_1 if the inequality*

$$P(| Y - Y_1 | \geqq T) \leqq P(| Z - Z_1 | \geqq T)$$

is true for all $T \geqq 0$.

If, for example, Y and Z are normal random variables with expectations Y_1 and Z_1 and standard deviations σ_y and σ_z, and if $\sigma_y < \sigma_z$, then Y is more peaked about Y_1 than Z about Z_1. Similarly, if Y is a random variable such that $P(Y < a) = P(Y > b) = 0$ for $a < b$, and if Z is the discrete random variable with $P(Z = a) = P(Z = b) = \frac{1}{2}$, then Y is more peaked about $\frac{1}{2}(a + b)$ than Z about the same point.

[2] Scheffé actually proves that (d) implies (c), but the Lebesgue convergence theorem on which his proof is based holds for convergence in measure (see e.g. [3]).

[1] I. Kaplansky, "A common error concerning kurtosis," *Am. Stat. Assn. Jour.*, Vol. 40 (1945), p. 259.