NOTES

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

A USEFUL CONVERGENCE THEOREM FOR PROBABILITY DISTRIBUTIONS

By Henry Scheffé

University of California at Los Angeles

In problems of establishing limiting distributions it is often apparent that the probability density $p_n(x)$ of a random variable X_n has a limit $p(x)$; throughout this paper $n = 1, 2, 3, \cdots$, and all limits are taken as $n \to \infty$. If $p(x)$ is the density of a random variable X, what we really care about then is whether the limits apply to probabilities, which involve integrals of the densities: Does $\lim Pr\{X_n \text{ in } S\} = Pr\{X \text{ in } S\}$ for all[1] Borel sets S, or, does

$$\lim \int_S p_n(x)\, dx = \int_S p(x)\, dx \text{ ?} \tag{1}$$

The question is thus one of taking a limit under an integral sign. Perhaps the most widely used justification of such a process is the following theorem of Lebesgue [1, p. 47; 2, p. 29]: If for a sequence $\{f_n(x)\}$ of integrable functions, $\lim f_n(x) = f(x)$ for almost all x in S, then a sufficient condition that

$$\lim \int_S f_n(x) dx = \int_S f(x)\, dx$$

is that there exist an integrable function $g(x)$ which uniformly dominates the sequence $\{f_n(x)\}$, that is, $|f_n(x)| \le g(x)$ for all n and all x in S, and $\int_S g(x)\, dx < \infty$. For example, in the excellent new treatise by Cramér the limiting form of the t-distribution is treated as follows [1, p. 252; other examples on pp. 369, 371]: For n degrees of freedom the t-variable has the density

$$p_n(x) = c_n(1 + x^2/n)^{-\frac{1}{2}(n+1)}, \tag{2}$$

where

$$c_n = (n\pi)^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}(n + 1))/\Gamma(\tfrac{1}{2}n). \tag{3}$$

It is shown fairly easily that $\lim p_n(x) = p(x)$, the density of $N(0, 1)$, where

[1] In defining the convergence of a sequence of distributions to the distribution of a discontinuous random variable X it is desirable to modify this requirement so that it is demanded only of sets S which are continuity intervals of X [1, p. 83]. We are concerned here however only with the "absolutely continuous case" where X has a probability density $p(x)$.

$N(m, \sigma^2)$ denotes the normal distribution with mean m and variance σ^2. Then to prove

$$\lim \int_{-\infty}^{\xi} p_n(x)\, dx = \int_{-\infty}^{\xi} p(x)\, dx,$$

Cramér shows that $\{p_n(x)\}$ is uniformly dominated by an integrable function.

It is instructive to consider some examples where

$$\lim \int_{-\infty}^{\xi} p_n(x)\, dx \tag{4}$$

does not equal

$$\int_{-\infty}^{\xi} \lim p_n(x)\, dx. \tag{5}$$

In the examples (*i*), (*ii*), (*iii*), $\lim p_n(x) = 0$ for all x and hence (5) is zero for all ξ.

(*i*) $p_n(x) = 1$ for $-n - 1 < x < -n$, zero elsewhere. Then (4) equals 1 for all ξ.

(*ii*) $p_n(x) = 1/n$ for $-\frac{1}{2}n < x < \frac{1}{2}n$, zero elsewhere. Here (4) equals $\frac{1}{2}$ for all ξ.

(*iii*) $p_n(x) = 2n^2x$ for $0 < x < 1/n$, zero elsewhere. Now (4) is zero for $\xi \leq 0$, unity for $\xi > 0$.

An example in which $\lim p_n(x) \not\equiv 0$ is

(*iv*) $p_n(x) = \frac{1}{2}[h_n(x) + p_0(x)]$, where h_n is the p_n of one of the above examples and p_0 is a fixed density. Then $\lim p_n(x) = \frac{1}{2}p_0(x)$. Now (4) exceeds (5) by half the amount it did in the corresponding above example.

The essential features of these examples could be obtained with normal distributions but would involve a little more computation, for instance, $N(-n, 1)$, $N(0, n^2)$, $N(1/n, 1/n^4)$, for examples (*i*), (*ii*), (*iii*), respectively.

We note that in none of these examples is $\lim p_n(x)$ a density. This suggests that the trouble might perhaps be prevented by requiring that $\lim p_n(x)$ be a density—which happens in the case from which we started. This surmise is correct. We may formalize the situation as follows:

DEFINITION. *A function $f(x)$ will be called a density if it is non-negative and $\int_R f(x)\, dx = 1$. Here R denotes the whole space of x.*

The reader may think of a univariate density, where x is a real variable and R is the real axis, but theorem and proof run the same for a k-variate density, where x is a point in a k-dimensional Euclidean space R.

THEOREM[2]. *If for a sequence $\{p_n(x)\}$ of densities*

$$\lim p_n(x) = p(x)$$

[2] The hypotheses of this theorem, while perfectly adapted to applications in probability and statistics, would not seem the "natural" ones in real variable or measure theory. Professor A. P. Morse has remarked to the writer that, if the theorem has not been stated in this form before, it is at least an easy corollary of some more general results known in that field. Nevertheless our direct proof based only on the familiar Lebesgue theorem and using only

for almost all x in R, then a sufficient condition that

$$\lim \int_S p_n(x)\, dx = \int_S p(x)\, dx,$$

uniformly for all Borel sets S in R, is that $p(x)$ be a density.

PROOF. Let us write the difference

$$p_n(x) - p(x) = \delta_n(x). \tag{6}$$

Then

$$\delta_n(x) \to 0 \tag{7}$$

for almost all x in R. Also

$$\int_S \delta_n\, dx = \int_S p_n\, dx - \int_S p\, dx, \tag{8}$$

and so it suffices to prove that $\int_S \delta_n\, dx \to 0$ uniformly for all S in R, where S henceforth denotes a Borel set. If in (8) we let $S = R$ we get

$$\int_R \delta_n\, dx = 0 \tag{9}$$

since p_n and p are densities. We now split the difference $\delta_n(x)$ into its positive and negative parts: Let

$$\delta_n^+ = \tfrac{1}{2}(\delta_n + |\delta_n|), \qquad \delta_n^- = \tfrac{1}{2}(\delta_n - |\delta_n|), \tag{10}$$

so that

$$\delta_n = \delta_n^+ + \delta_n^-, \qquad \delta_n^+ \geq 0, \qquad \delta_n^- \leq 0.$$

From (7) and (10), we find

$$\delta_n^- \to 0 \tag{11}$$

for almost all x in R, and from (9),

$$\int_R \delta_n^+\, dx + \int_R \delta_n^-\, dx = 0. \tag{12}$$

---

very simple manipulations may be of interest to readers of the *Annals*. Professor Morse also pointed out that the stronger result $\lim \int_S |p_n(x) - p(x)|\, dx = 0$ uniformly for all S, may be stated. This follows from our proof since

$$\int_S |p_n - p|\, dx = \int_S \delta_n^+\, dx - \int_S \delta_n^-\, dx.$$

By virtue of (6), $\delta_n \geq -p$. Now if $\delta_n \leq 0$, $\delta_n^- = \delta_n \geq -p$, and if $\delta_n > 0$, $\delta_n^- = 0 \geq -p$, and hence in every case $0 \geq \delta_n^- \geq -p$. Since we now have $| \delta_n^-(x) | \leq p(x)$ and $\int_R p(x)\, dx = 1$, we may apply[3] the Lebesgue theorem to get

$$\lim \int_R \delta_n^-\, dx = \int_R \lim \delta_n^-\, dx.$$

The right member is zero because of (11). It then follows from (12) that $\lim \int_R \delta_n^+\, dx$ is also zero. The relations

$$0 \leq \int_S \delta_n^+\, dx \leq \int_R \delta_n^+\, dx \to 0,$$

$$0 \geq \int_S \delta_n^-\, dx \geq \int_R \delta_n^-\, dx \to 0$$

guarantee that the quantities $\int_S \delta_n^+\, dx$ and $\int_S \delta_n^-\, dx$ have the limit zero uniformly for all S, and hence the same is true of their sum (8).

Returning to the example (2), we remark that it is practically obvious that the second factor on the right has the limit $e^{-\frac{1}{2}x^2}$, but it is not quite so obvious that $\lim c_n = (2\pi)^{-\frac{1}{2}}$. This situation is typical of many applications where it is more difficult to evaluate the limit of "the" constant than the limit of the remaining factors, and one wonders after obtaining the latter limit whether the constant is not automatically forced toward the limit desired for it, and whether the direct calculation of its limit could not be avoided. Let us put the question as follows: Suppose that

$$\{p_n(x) = c_n f_n(x)\}$$

is a sequence of densities and that

$$p(x) = cf(x)$$

is also a density. Then if $\lim f_n(x) = f(x)$ for almost all x, may we conclude that $\lim c_n = c$? If so, we could then apply the above theorem without having evaluated the limit of the constant or produced a dominating function. Unfortunately the answer to this question is no, as shown by example (*iv*) above:

[3] Although our proof rests on the Lebesgue convergence theorem, this theorem is applied to $\delta(x)$ and not to $p_n(x)$. While in most cases of practical interest the sequence $\{p_n(x)\}$ is uniformly dominated by an integrable function, it is possible to devise a simple example where this is not true and yet our theorem applies: Let $p_n(x) = 1$ for $1/(n+1) \leq x \leq 1$ and for $a_n \leq x \leq a_{n+1}$, zero elswhere, where $a_n = \sum_{i=1}^{n} 1/i$. Then $\sup p_n(x) = 1$ for all $x > 0$, nevertheless $\lim p_n(x)$ is a density, namely that of the uniform distribution on $(0, 1)$.

If we let $f_n(x) = h_n(x) + p_0(x)$, and $f(x) = p_0(x)$, then $\lim f_n(x) = f(x)$, but $c_n = \frac{1}{2}$ and $c = 1$, hence $\lim c_n \neq c$. Employing the assumption that $p_n(x)$ and $p(x)$ are densities we see

$$1/c_n = \int_R f_n(x)\, dx, \qquad 1/c = \int_R f(x)\, dx,$$

and hence $\lim c_n = c$ if and only if

$$\lim \int_R f_n(x)\, dx = \int_R \lim f_n(x)\, dx. \tag{13}$$

It follows that in such cases if we wish to establish a limiting distribution in the sense (1), we may either prove $\lim c_n = c$, or we may justify (13), say by producing a suitable dominating function, but we need not do both. No doubt the first alternative would be preferable at all but the most advanced levels of teaching or exposition.

REFERENCES

[1] H. CRAMÉR, *The Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.
[2] S. SAKS, *Theory of the Integral*, Stechert, New York, 1937.

AN EXPLICIT REPRESENTATION OF A STATIONARY GAUSSIAN PROCESS

BY M. KAC[1] AND A. J. F. SIEGERT

Cornell University and *Syracuse University*

1. In a paper which will soon appear in the *Journal of Applied Physics* [1] the authors have introduced methods of calculating certain probability distributions which are of importance in the theory of random noise in radio receivers.

The complexity of the physical problem and occasional uses of heuristic reasonings may have obscured some of the mathematical points. For this reason the authors felt that it may be worth while to illustrate one of the basic ideas on a simple but important example.

2. A stationary Gaussian process is a one parameter family $x(t)$ of random variables such that:

(a). $x(t)$ is normally distributed; the mean and the variance being independent of t

(b). the joint probability distribution of $x(t_1), x(t_2), \cdots, x(t_r)$ is multivariate Gaussian whose parameters depend only on the differences $t_j - t_k$.

[1] John Simon Guggenheim Memorial Fellow.