ON INTERCHANGING LIMITS AND INTEGRALS

JOHN W. PRATT[1]

Harvard University

One frequently wants to show $\lim \int f_n = \int \lim f_n$; that is, knowing $f_n \to f$ pointwise, one wants to show $\int f_n \to \int f$. Commonly used criteria are those of the Lebesgue (dominated or bounded) convergence theorem [1, Theorem 26.D; 2, Theorem 7.2C; etc.] and Scheffé's "Useful Convergence Theorem for Probability Distributions" [3]. The following criterion sometimes applies more directly and is never much harder to apply. Informally, the criterion is that f_n shall be bounded above and below by functions which converge pointwise and in integral; or, in other words, a convergent sequence permits exchange of lim and $\int$ if it is bracketed by two sequences which permit this exchange. Specifically,

THEOREM 1. *If*

(i) $f_n \to f$, $g_n \to g$, $G_n \to G$,

(ii) $g_n \leqq f_n \leqq G_n$ *for all* n,

(iii) $\int g_n \to \int g$ *and* $\int G_n \to \int G$ *with* $\int g$ *and* $\int G$ *finite,*

then $\int f_n \to \int f$ *and* $\int f$ *is finite.*

(i) and (ii) may be interpreted as holding at each point and the integrals as ordinary (Lebesgue) integrals over a fixed (Lebesgue measurable) subset of the real line or k-dimensional Euclidean space.

More generally, it is assumed throughout this note that all integrals are taken with respect to the same measure μ on a Borel field $\mathcal{B}$, all sets mentioned are measurable, all functions mentioned are measurable from $\mathcal{B}$ to the class of Borel sets, inequalities like (ii) hold almost everywhere $[\mu]$, and convergence of functions, as in (i), is either almost everywhere $[\mu]$ or in measure $[\mu]$. Proofs will be given for the case of convergence almost everywhere. The more general case follows, since every subsequence of a sequence which converges in measure has a subsubsequence which converges almost everywhere.

In Theorem 1, Corollary 1, and Corollary 4, $\int$ may be replaced by $\int_B$ for a fixed set B (and $\int_S$ by $\int_{S \cap B}$) provided all integrals are so replaced. This is not a real generalization, being the result of substituting μ_1 for μ, where

$$\mu_1(S) = \mu(B \cap S)$$

for all S.

PROOF OF THEOREM 1: $0 \leqq f_n - g_n \to f - g$ and $0 \leqq G_n - f_n \to G - f$. We

Received February 13, 1959; revised August 17, 1959.

[1] This research has been supported in part by the United States Navy through the Office of Naval Research, under contract Nonr 1866(37). Reproduction in whole or in part is permitted for any purpose of the United States Government.

can thus apply Fatou's Lemma [1, Theorem 27.F, etc.] (The Lemma says $\int \liminf h_n \leqq \liminf \int h_n$ for $h_n \geqq 0$.) We obtain:

$$\int f - \int g = \int \lim (f_n - g_n) \leqq \liminf \int (f_n - g_n) = \liminf \int f_n - \int g;$$

$$\int G - \int f = \int \lim (G_n - f_n) \leqq \liminf \int (G_n - f_n) = \int G - \limsup \int f_n.$$

Therefore, $\limsup \int f_n \leqq \int f \leqq \liminf \int f_n$, so $\int f_n \to \int f$, q.e.d.

COROLLARY 1. *If* (i)–(iii) *hold and, in addition*

(iv) $g_n \leqq 0 \leqq G_n$ *for all* n,

then

(a) $\int |f_n - f| \to 0$;

(b) $\int_S f_n \to \int_S f$ *uniformly in* S *(*S *measurable)*;

(c) $\int h f_n \to \int h f$ *for all bounded functions* h, *uniformly in* h *for each bound.*

(a)–(c) are equivalent, as is well known, since it is immediate that (a) implies (c), (c) implies (b), and (b) implies (a). To prove (i)–(iv) imply (a), note that, by (ii) and (iv),

$$0 \leqq |f_n - f| \leqq |f_n| + |f| \leqq G_n - g_n + G - g \to 2(G - g),$$

while, by (iii), $\int (G_n - g_n + G - g) \to \int 2(G - g)$ which is finite. Thus Theorem 1 applies with $|f_n - f|$ for f_n, 0 for g_n, and $G_n - g_n + G - g$ for G_n. (b) and (c) without uniformity are perhaps even more direct applications of Theorem 1.

Conditions (i)–(iii) alone do not imply (a)–(c), nor can the region of integration in the conclusion of Theorem 1 be different from that in (iii) in general. For trivial counter-examples, let $f_1(x) = -1$ for $-1 < x < 0$, $=1$ for $0 < x < 1$, and $=0$ otherwise; let $f = g = G = 0$; let $f_n(x) = g_n(x) = G_n(x) = n f_1(nx)$ or $n^{-1} f_1(n^{-1}x)$; and let the integrals be ordinary (Lebesgue) integrals from -1 to 1 or $-\infty$ to ∞. Then (i)–(iii) hold but (a)–(c) do not, nor can the integrals in the conclusion of Theorem 1 be taken over positive x only. The choice $n f_1(nx)$ from -1 to 1 gives finite measure and the choice $n^{-1} f_1(n^{-1}x)$ from $-\infty$ to ∞ gives uniformly bounded functions. If the measure is finite and the functions are uniformly bounded, of course, $f_n \to f$ implies (a)–(c) without further conditions.

Theorem 1 and Corollary 1 reduce to the Lebesgue convergence theorem when $G_n = G = -g = -g_n \geqq 0$.

The next corollary is Scheffé's theorem.

COROLLARY 2. *If all* f_n and f *are probability densities and* $f_n \to f$, *then* (a)–(c) *hold.*

PROOF. (i)–(iv) are satisfied by $g_n = g = 0$, $G_n = f_n$, $G = f$. Thus, Corollary 1 applies, q.e.d.

Suppose P_n and P are the probability measures given by the probability densities f_n and f; that is, $P_n(S) = \int_S f_n$, $P(S) = \int_S f$. It is an immediate consequence of Corollary 2 that, in Euclidean space, $f_n \to f$ implies P_n converges in distribution to P (in the usual sense that the c.d.f. of P_n approaches the c.d.f.

of P at points of continuity of the latter. $\mathcal{B}$ must include Borel sets, but μ need not be Lebesgue measure). It is more illuminating to compare consequences (b) and (c) of convergence of densities to a density with the following conditions, each of which is equivalent to convergence in distribution in Euclidean space. (This is well known; in fact, (c′) is often used to define convergence in distribution more generally.)

(b′) $P_n(S) \rightarrow P(S)$ for every set S whose boundary has P-measure 0.

(c′) $\int h dP_n \rightarrow \int h dP$ for every bounded *continuous* function h.

Provided open sets are measurable, (b′) and (c′) are obviously weaker than (b) and (c).

COROLLARY 3. *A density which is continuous in a parameter has continuous (in fact, equicontinuous) power and Type II error functions.*

PROOF. Suppose $f(x, \theta)$ is a density function for each θ and continuous in θ for each x (or, more generally, for almost all x at each value of θ). The power function of a test is $\alpha(\theta) = \int h(x)f(x, \theta)$ where $0 \leqq h \leqq 1$ and the integration is over x. Given any sequence $\theta_n \rightarrow \theta$, Corollary 2 applies with $f_n(x) = f(x, \theta_n)$, $f(x) = f(x, \theta)$, giving $\alpha(\theta_n) \rightarrow \alpha(\theta)$ uniformly in h, q.e.d.

Another consequence of Theorem 1 is that a function may be differentiated with respect to a parameter under the integral sign if its derivative is bracketed by the derivatives of two functions which permit differentiation under the integral sign. That is,

COROLLARY 4. *Suppose θ is a real parameter. Let D denote differentiation with respect to θ and $\int$ integration over x. If θ_0 is an interior point of an interval I and*

(i) $Dg(x, \theta)$, $Df(x, \theta)$, and $DG(x, \theta)$ exist for all $\theta \varepsilon I$,

(ii) $Dg(x, \theta) \leqq Df(x, \theta) \leqq DG(x, \theta)$ for all $\theta \varepsilon I$,

(iii) $D\int g(x, \theta_0) = \int Dg(x, \theta_0)$ and $D\int G(x, \theta_0) = \int DG(x, \theta_0)$ with all four quantities existing and finite,

then $D\int f(x, \theta_0) = \int Df(x, \theta_0)$ with both quantities existing and finite.

This follows from Theorem 1, since the difference quotients of f lie between the corresponding difference quotients of g and of G. It suffices that (i) and (ii) hold for almost all $x[\mu]$.

When $G(x, \theta) = -g(x, \theta) = \theta G(x)$, Corollary 4 reduces to the commonly given criterion that $| Df(x, \theta)| \leqq G(x)$ for G integrable.

The main advantage of the approach presented here is its simplicity. From the point of view of application, Theorem 1 is a single theorem which applies with trivial specialization to the situations for which Lebesgue's and Scheffé's theorems are tailor-made. Furthermore, Theorem 1 implies the following facts, for instance, more directly than the latter theorems do.

(1) If $f_n \rightarrow f$ and $\int | f_n | \rightarrow \int | f |$ finite, then $\int | f_n - f | \rightarrow 0$.

(2) If all f_n and f are densities, $f_n \rightarrow f$, all h_n and h are test functions (that is, $0 \leqq h_n \leqq 1$, $0 \leqq h \leqq 1$), and $h_n \rightarrow h$, then $\int h_n f_n \rightarrow \int hf$.

Pedagogically, I find Theorem 1 useful in reviewing measure theory briefly in a probability course. The most expeditious way I know to prove the Lebesgue Convergence Theorem is to prove Fatou's Lemma first [2, Section 7.2, for instance]. But the proof of Theorem 1 is a simple extension of a proof of Lebesgue's Theorem from Fatou's Lemma. Thus Corollaries 2 and 3 are obtained virtually without extra proof. Concepts not involved in the statement of the theorems (such as equicontinuity at the empty set) need never be introduced.

In fact, it is interesting to note that the equivalence of Theorem 1, the Lebesgue Convergence Theorem, Fatou's Lemma, and the Monotone Convergence Theorem [2, Theorem 7.2A, etc.] depends only on properties of measurable functions and $\int$ having to do with order and addition. Theorem 1 is especially natural in this context.

The fact that the foregoing theorems have short proofs is fortunate for the purposes mentioned. However, it means the individual theorems are in this sense not deep, and it makes it hard to verify that any particular one is new. I have not searched the literature thoroughly, but I have never seen even a statement of Theorem 1 or Corollary 1, although they unify important theorems on interchanging lim and $\int$, both conceptually and pedagogically. The only statement of Corollary 3 I know is Wald's [4, p. 133], although it is obviously important and frequently assumed tacitly. Corollary 4 I presume is new.

REFERENCES

[1] Paul R. Halmos, *Measure Theory*, D. Van Nostrand, New York, 1950.
[2] Michel Loève, *Probability Theory*, D. Van Nostrand, New York, 1955.
[3] Henry Scheffé, "A Useful Convergence Theorem for Probability Distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.
[4] Abraham Wald, *Statistical Decision Functions*, John Wiley & Sons, New York, 1950.