RANDOM SAMPLING IN THE EVALUATION OF A LEBESGUE INTEGRAL

C. R. ADAMS AND A. P. MORSE

The purpose of this note is to show that in the evaluation of a Lebesgue integral a sort of random sampling scheme is permissible. It is possible, therefore, that the results may have application to questions of probability or statistics.

The chief positive result established here is embodied in Theorem 4. With the object of indicating clearly the question at issue, however, we formulate two particular cases in the preliminary Theorems 1 and 2. In connection with the first of these particular cases a negative result, exhibiting the need for a limitation which is placed upon the sampling process, will be shown. Theorem 3, a third particular case of Theorem 4, will be established as an aid to the proof of Theorem 4.

Let $x(t)$ be summable on the interval $0 \leq t \leq 1$, that is, $x \, \varepsilon \, L([0, 1])$; let k be a fixed integer greater than or equal to 1; let $[0, 1]$ be divided into n equal parts $(n = 1, 2, 3, \cdots)$ each of length $1/n$; and let the ith one $(i = 1, 2, \cdots, n)$ of these parts in turn be divided into k equal parts; let m_{n,i,j_i}, $(i = 1, 2, \cdots, n; j_i = 1, 2, \cdots, k$ for each $i)$, designate the L-integral mean of $x(t)$ on the j_ith of the k equal parts into which the ith subinterval of $[0, 1]$ has been divided; and form the "Riemann-Lebesgue" sum

$$kS_n = \sum_{i=1}^{n} m_{n,i,j_i}/n.$$

Clearly this sum is multiple-valued, of multiplicity k^n.

THEOREM 1. *Under the conditions stated, we have*

$$\lim_{n \to \infty} kS_n = \int_0^1 x(t)dt.$$

This can be proved by showing that if S_n' and S_n'' denote arbitrary values of S_n obtainable by some choices of the j_i, we have

$$|S_n' - S_n''| \leq \int_0^{1 - 1/(kn)} |x(t + 1/(kn)) - x(t)| \, dt,$$

which tends to zero with $1/n$. The limit of kS_n may then easily be identified with $\int_0^1 x(t)dt$.

Before proceeding to the second preliminary theorem we show that, if k is allowed to be a function of n, $k(n)$, with

$$\limsup_{n\to\infty} k(n) = \infty,$$

there always exists an $x \in L([0, 1])$ (in fact, a *bounded* summable function) for which Theorem 1 does not hold. Indeed, under this condition on $k(n)$, there exists a non-decreasing sequence $\{N_n\}$ such that $k(N_n)$ is greater than or equal to 2^{n+1} for every n. Let A_n be the subset of $[0, 1]$ defined by*

$$A_n = \sum_{i=0}^{N_n-1} [i/N_n, i/N_n + 1/(N_n k(N_n))].$$

Clearly A_n has measure less than or equal to $1/2^{n+1}$ for each n, and $A = \sum_{n=1}^{\infty} A_n$ has measure less than or equal to $1/2$. Let $x(t)$ be the characteristic function of the set A. Then if we take j_i always as 1, it is clear that

$$k(N_n)S_{N_n} = \sum_{i=1}^{N_n} m_{N_n, i, 1}/N_n = 1 \qquad \text{for every } n,$$

whence, as $n \to \infty$,

$$k(N_n)S_{N_n} \to 1 > 1/2 \geqq \int_0^1 x(t)dt.$$

Incidentally, it may be noted that there exist other selections of n and the j_i under which $k(n)S_n$ tends to a limit less than or equal to

$$\int_0^1 x(t)dt.$$

On the other hand, we observe that for each fixed $x \in L([0, 1])$, there exists a function $k(n)$ with $\lim_{n\to\infty} k(n) = \infty$ for which we do have

$$\lim_{n\to\infty} k(n)S_n = \int_0^1 x(t)dt.$$

For, by Theorem 1, to each integer $m > 0$ corresponds an integer $N_m > 0$ such that

$$\left| mS_n - \int_0^1 x(t)dt \right| < 1/m$$

---

* The symbol $[a, b]$ stands for the closed interval $a \leqq t \leqq b$.

for all $n > N_m$. If, in addition, the N_m are chosen to satisfy the condition $N_m < N_{m+1}$ for every m, we may define $k(n) = n$ for $N_n < n \leq N_{n+1}$, $(n = 1, 2, 3, \cdots)$, and obtain the property asserted.

A complete extension of Theorem 1 to what is usually thought of as a Riemann integral set-up is provided by the following theorem.

THEOREM 2. *Let* $x \in L([0, 1])$ *and* k *be any fixed real number greater than or equal to* 1; *and for* $0 < \delta \leq 1$ *let* $H(\delta)$ *stand for the aggregate of measurable subsets* A *of* $[0, 1]$ *defined by the condition* $A \in H(\delta)$ *if and only if there exist numbers* $a_0, a_1, \cdots, a_p$ *with* $0 = a_0 < a_1 < \cdots < a_p = 1$ *such that**

$$\left| [a_{j-1}, a_j] \cdot A \right| = (a_j - a_{j-1})/k, \quad a_j - a_{j-1} < \delta, \quad j = 1, 2, \cdots, p.$$

Then we have

$$\lim_{\delta \to 0} \inf_{A \in H(\delta)} k \int_A x(t) dt = \int_0^1 x(t) dt = \lim_{\delta \to 0} \sup_{A \in H(\delta)} k \int_A x(t) dt.$$

Still further generalization, however, is possible. As a first step toward it we establish the following lemma, in which $C([0, 1])$ stands for the class of functions continuous on $[0, 1]$.

LEMMA. *Let* $x \in C([0, 1])$. *Corresponding to any* $\epsilon > 0$, *there exists a* $\delta > 0$ *such that if* $\{A_n\}$ *and* $\{B_n\}$ *are sequences (finite or infinite) of measurable subsets of* $[0, 1]$ *satisfying the conditions*

$$A_n \subset B_n, \quad 0 < |A_n|, \quad \text{diameter } (B_n) < \delta, \quad n = 1, 2, 3, \cdots;$$

$$|B_m \cdot B_n| = 0, \quad m \neq n; \qquad \sum_n B_n = [0, 1],$$

then we have

$$\left| \int_0^1 x(t) dt - \sum_n |B_n| \cdot \int_{A_n} x(t) dt / |A_n| \right| < \epsilon.$$

PROOF. There exists $\delta > 0$ such that $|t_1 - t_2| < \delta$ implies $|x(t_1) - x(t_2)|$ $< \epsilon/2$. On each set B_n of diameter less than δ, the integrals†

$$\int_{B_n} x / |B_n|, \qquad \int_{A_n} x / |A_n|$$

both lie between the greatest lower and least upper bounds of $x(t)$ on

---

* We use $|E|$ to denote the measure of a measurable set E.

† Henceforth we shall often write, for example, $\int_{B_n} x / |B_n|$ in place of $\int_{B_n} x(t) dt / |B_n|$, when such abbreviation seems likely to cause no confusion.

the set B_n, whence these integrals differ by $\epsilon/2$ at most. Therefore we infer

$$\left| \int_0^1 x - \sum_n |B_n| \cdot \int_{A_n} x/|A_n| \right|$$

$$= \left| \sum_n |B_n| \cdot \int_{B_n} x/|B_n| - \sum_n |B_n| \cdot \int_{A_n} x/|A_n| \right|$$

$$\leqq \sum_n |B_n| \cdot \left| \int_{B_n} x/|B_n| - \int_{A_n} x/|A_n| \right| \leqq \frac{\epsilon}{2} \sum_n |B_n| < \epsilon.$$

Remarks. It is now a simple matter to extend this lemma to the case of $x \, \epsilon \, R([0, 1])$, where R represents the class of Riemann-integrable functions. For an $x \, \epsilon \, R$ can be approximated in the mean within $\epsilon/2$ by a function $y \, \epsilon \, C$ with $y(t) \geqq x(t)$ on $[0, 1]$. Then if δ corresponds to $\epsilon/2$ for y in accordance with the lemma, we have

$$\sum_n |B_n| \cdot \int_{A_n} x/|A_n|$$

$$\leqq \sum_n |B_n| \cdot \int_{A_n} y/|A_n| < \int_0^1 y + \epsilon/2 < \int_0^1 x + \epsilon.$$

Similarly, by approximating x by y from below, we obtain the other half of the desired inequality. Clearly the lemma holds also for any function x which is almost everywhere equal to a function $y \, \epsilon \, R([0, 1])$.

That the lemma cannot be extended to the case of $x \, \epsilon \, L([0, 1])$ is evident; for whenever x is not essentially bounded, the A_n can be so chosen that the corresponding sum is arbitrarily large, and whenever x is essentially bounded but not almost everywhere equal to a function $y \, \epsilon \, R([0, 1])$, the A_n and B_n can be so chosen that the corresponding sum approximates either the essential upper or the essential lower Darboux integral of x.

The following theorem extends Theorem 2 in three distinct directions.

THEOREM 3. *Let $x \, \epsilon \, L([0, 1])$ and k be any fixed real number greater than or equal to 1. Corresponding to any $\epsilon > 0$, there exists a $\delta > 0$ such that if $\{A_n\}$ and $\{B_n\}$ are sequences (finite or infinite) of measurable subsets of $[0, 1]$ satisfying the conditions*

$$A_n \subset B_n, \quad 0 < |B_n| \leqq k|A_n|, \quad \text{diameter } (B_n) < \delta, \quad n = 1, 2, 3, \cdots ;$$

$$|B_m \cdot B_n| = 0, \quad m \neq n; \qquad \sum_n B_n = [0, 1],$$

then we have

$$\left| \int_0^1 x(t)dt - \sum_n |B_n| \cdot \int_{A_n} x(t)dt / |A_n| \right| < \epsilon.$$

PROOF. Let $y \ \varepsilon \ C([0, 1])$ be such that $\int_0^1 |x-y| < \epsilon/(2k+2)$, and let δ correspond to $\epsilon/2$ for y in accordance with the lemma. Then we have

$$\left| \int_0^1 x - \sum_n |B_n| \cdot \int_{A_n} x / |A_n| \right|$$

$$= \left| \int_0^1 x - \int_0^1 y + \int_0^1 y - \sum_n |B_n| \cdot \int_{A_n} y / |A_n| \right.$$

$$\left. + \sum_n |B_n| \cdot \int_{A_n} (y - x) / |A_n| \right|$$

$$\leqq \int_0^1 |x - y| + \left| \int_0^1 y - \sum_n |B_n| \cdot \int_{A_n} y / |A_n| \right|$$

$$+ \sum_n |B_n| \cdot \int_{A_n} |x - y| / |A_n|$$

$$< \epsilon/(2k + 2) + \epsilon/2 + k \int_0^1 |x - y| < \epsilon.$$

Finally, we have the following theorem.

THEOREM 4. *Theorem 3 holds for $x \ \varepsilon \ L(E)$, where E is an arbitrary measurable set, bounded or unbounded.*

PROOF. When E is bounded, the validity of this result may be seen at once. For the above lemma holds if x is *uniformly* continuous on E. And if $E \subset [a, b]$, x may be defined as zero on $[a, b] - E$, and a function $y \ \varepsilon \ C([a, b])$ approximating x in the mean within $\epsilon/(2k+2)$ on $[a, b]$ is uniformly continuous on E and approximates x in the mean within $\epsilon/(2k+2)$ on E; so that the argument used above in the proof of Theorem 3 applies without change. Moreover, the reasoning shows clearly that the following statement is valid.

(i) *If δ corresponds to x, E, k, ϵ for a bounded set E, and E_1 is any measurable subset of E, the same δ corresponds to x, E_1, k, ϵ.*

To dispose of the case of E unbounded, let H_M stand for the set $[-M, M]$, $M > 0$; and $\tilde{H}_M$ for the complement of H_M with respect to the infinite interval $-\infty < t < +\infty$. First, let M be chosen so that

$\int_{E \cdot \tilde{H}_M} |x(t)|\, dt < \epsilon/(2k+2)$. Secondly, in accordance with the theorem for a bounded set, let δ, $(\delta < 1)$, correspond to x, $E \cdot H_{M+1}$, k, $\epsilon/2$. Thirdly, let $\{A_n\}$ and $\{B_n\}$ be any sequences satisfying the hypotheses; and allow $\sum_n'$ to stand for a sum extended over all values of n for which $|B_n \cdot H_M| > 0$, and $\sum_n''$ for the rest of the sum $\sum_n$. Then we have

$$\left| \int_E x - \sum_n |B_n| \cdot \int_{A_n} x/|A_n| \right|$$

$$= \left| \int_{\Sigma_n'' B_n} x + \int_{\Sigma_n' B_n} x - \sum_n' |B_n| \cdot \int_{A_n} x/|A_n| \right.$$

$$\left. - \sum_n'' |B_n| \cdot \int_{A_n} x/|A_n| \right|$$

$$\leqq \int_{\Sigma_n'' B_n} |x| + \left| \int_{\Sigma_n' B_n} x - \sum_n' |B_n| \cdot \int_{A_n} x/|A_n| \right|$$

$$+ k \int_{\Sigma_n'' B_n} |x|$$

$$< (k+1) \int_{E \cdot \tilde{H}_M} |x| + \epsilon/2 < \epsilon$$

in consequence of the relation $\sum_n' B_n \subset E \cdot H_{M+1}$, the choice of δ, and the remark (i).

Similar results obtain for a function of several variables, since the above proofs (employing only the uniform continuity of a function continuous over a closed interval and the possibility of approximating in the mean a summable function by a continuous one) are essentially independent of dimensionality.

BROWN UNIVERSITY AND
    THE INSTITUTE FOR ADVANCED STUDY