University of Illinois
Spring 2017

Department of Economics
Roger Koenker

Economics 574
Appendix to 13 Ways

This Appendix provides an overview of some of the background that is left undeveloped in “13 Ways of looking at a random variable.” It is neither necessary or sufficient to grasp the essentials of that lecture. Pollard (2002) constitutes an unmeasureably more thorough and sophisticated treatment. The marginal comments are in the spirit of Graham, Knuth and Patashnik (1989) but a pale imitation, at best.

1. Instant Measure Theory

Just add water, how much?

Our first question, which is not strictly necessary, or relevant, but is a useful bit of general education is: What is the difference between Riemann and Lebesgue integral? A straightforward answer may be provided with the aid of a couple of pictures. Riemann integral works by dividing up the *domain* and approximating via mean value theorem like this

$$RS_m = \sum_{i=1}^{m} f(x_{mi}^*)(x_{mi} - x_{mi-1}).$$

This works well for nonnegative, continuous functions, but fails in the sense that the approximation may fail to converge, i.e., there exist sequences $f_n \to f$ such that

$$\int_a^b f_n(x)dx \to \int_a^b f(x)dx$$

fails.

The Lebesgue integral proceeds by subdividing the range rather than the domain,

$$LS_m = \sum_{k=1}^{2^m} \frac{k-1}{2^m} \times \mu(\{x : \frac{k-1}{2^m} \leq f(x) < \frac{k}{2^m}\})$$

so rather than having fixed-width intervals of the domain and multiplying by heights as in elementary calculus, we have fixed width division of the range and we weight by the width, or length, or *measure* of the set of x such that $f(x)$ lies in these intervals. So the Lebesgue approximation is the area below the red curve in Figure 2. To see this start with the top block we get all area below this block. Then we get the two neighboring strips, etc. etc.

Do *you* see any red curve?

Note that we can perturb f at a few isolated points and this doesn't affect the Lebesgue sum, since these heights are multiplied by zero.

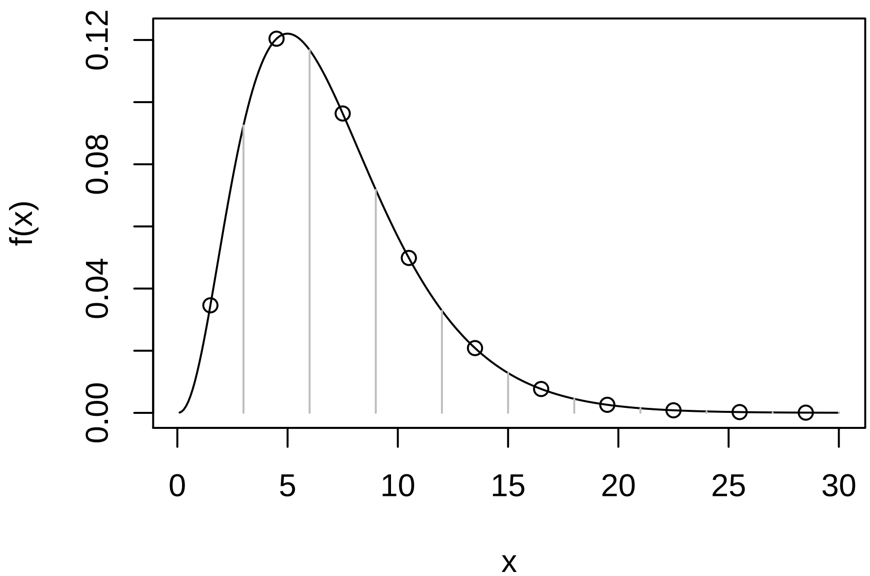

FIGURE 1. Riemann approximation divides the domain into pieces and computes an approximate area for each piece.

So why does this work better than the Riemann scheme, or when does it work better? For our function, f, we need to be able to find the length, measure of the sets

$$\{x : \frac{k-1}{2^m} \leq f(x) < \frac{k}{2^m}\}$$

(so to speak).

If we can assure ourselves that these sets are *measurable*, then we are set. This leads us to the concept of σ-fields.

What about Boxers?

Set Theory in Brief

Let $\mathcal{A}$ be a nonempty set of subsets, A, of a nonempty set Ω. Recall,

A^c	denotes the complement of A
$A \cup B$	denotes the union of A and B
$A \cap B$	denotes the intersection of A and B
$A \subset B$	means A is contained in B, i.e. is a subset of B
$A_n \nearrow$	means the sequence A_n is increasing so $A_n \subset A_{n+1}$ for all $n \geq 1$
$A_n \searrow$	means $A_n \supset A_{n+1}$ for all $n \geq 1$.

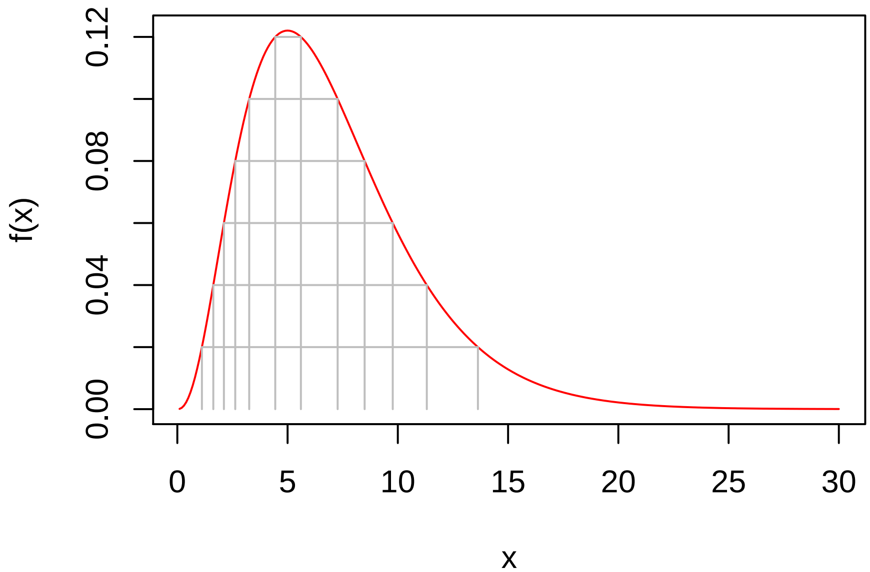

FIGURE 2. Lebesgue approximation divides the range into pieces and computes an approximate area for each piece.

Def. $\mathcal{A}$ is a *field* if it is closed under complements and unions, so $A \subset \mathcal{A}, B \subset \mathcal{A} \Rightarrow A^c \subset \mathcal{A}, A \cup B \subset \mathcal{A}$.

Def. $\mathcal{A}$ is a σ-*field* if it is closed under complements and countable unions so $A_i \subset \mathcal{A}\ \ i = 1, 2, \ldots\ \ \Rightarrow A_i^c \subset \mathcal{A}$ and $\bigcup_{i=1}^{\infty} A_i \subset \mathcal{A}$.

Remark σ-fields are also closed under intersections since $A \cap B = (A^c \cup B^c)^c$.

Def. If $\mathcal{A}$ is a σ-field and $\mu : \mathcal{A} \to [0, 1]$ is a set function that is countably additive in the sense that for disjoint A_i

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i),$$

then μ is called a *measure*, or a *countably additive measure* on $(\Omega, \mathcal{A})$.

Examples:

1)	Lebesgue:	length of the set A
2.)	Counting:	#(A) cardinality of A, i.e. number of elements of A

3.) Indicator: $\delta_{\omega_0}(A) = I_{\omega_0}(A) = \begin{cases} 1 & \text{if } \omega_0 \in A \\ 0 & \text{otherwise.} \end{cases}$

Borel Sets

It is convenient to have a minimal σ-field containing a specified class of sets, ξ. We say ξ is the generator of a σ-field and write,

$$\sigma[\xi] \equiv \bigcap\{\mathcal{F}_\alpha : \mathcal{F}_\alpha \text{ is a } \sigma\text{-field of subsets of } \Omega \text{ for which } \xi \subset \mathcal{F}_\alpha\}$$

Since it is the intersection it has to be minimal.

Suppose $\Omega = \Re$, and ξ consists of all finite disjoint unions of intervals of the form

$$(a, b], (-\infty, b], \text{ and } (a, +\infty)$$

ξ is a σ-field, but not a field since we can't get (a, b) by finite unions but we can by countable unions: take $(a, b_n]$ with $b_n \nearrow b$. Then,

$$\mathcal{B} = \sigma[\xi]$$

is called the Borel subsets of $\Re$. And $\mu(A)$ is the countably additive measure assigning the length of the intervals composing A.

This can be extended to general metric spaces. Let (Ω, d) be a metric space and

$$U = \{ \text{ all } d\text{-open supsets of } \Omega\}$$

then $B = \sigma[U]$ are the Borel sets of (Ω, d) or the Borel σ-field of (Ω, d).

Expectations and the Lebesgue Integral

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and suppose that $X, X_1, X_2, \ldots$ are measurable functions from $(\Omega, \mathcal{A}, \mu)$ to $(\Re, \mathcal{B})$. If the sets $\{A_i : i = 1, \ldots, n\}$ are disjoint, it is convenient to write

$$\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} A_i$$

and if, further, $\sum A_i = \Omega$ we say that the A_i are a partition of Ω.

We may define the Lebesgue integral in a sequence of steps beginning with the simplest cases and building up from these.

(1) If $X = \sum_{i=1}^{n} x_i I_{A_i}$, we call it a simple (block) function provided $x_i \geq 0$ and the A_i's constitute a partition of Ω. Then

$$\int X d\mu \equiv \sum x_i \mu(A_i).$$

This is the simplest version of our original definition with only a finite number of possible values for X.

(2) If $X \geq 0$, then

$\int X d\mu = \sup\{\int Y d\mu : Y \text{ is a simple function such that } 0 \leq Y \leq X\}$

(3) For general measurable X,

$$\int X d\mu = \int X^+ d\mu - \int X^- d\mu$$

provided either $\int X^+ d\mu$ or $\int X^- d\mu$ is finite.

(4) For unmeasurable X, if X equals a measurable function Y on a set A such that $\mu(A^c) = 0$, has zero measure, then

$$\int X d\mu = \int Y d\mu$$

Properties

(1) Using a simple functions it is easy to show that
 (a) $\int (X+Y) d\mu = \int X d\mu + \int Y d\mu$
 (b) $\int cX d\mu = c \int X d\mu$
 (c) $X \geq 0 \Rightarrow \int X d\mu \geq 0$.

(2) (Monotone Convergence Theorem) Suppose $X_n \nearrow X$ a.c. for measurable functions $X_n \geq 0$, then

$$0 \leq \int X_n d\mu \nearrow \int X d\mu$$

(3) (Fatou's Lemma) For measurable $X_n \geq 0$ a.e.

$$\int \liminf X_n d\mu \leq \liminf \int X_n d\mu$$

(4) (Dominated Convergence Theorem) Suppose $|X_n| \leq Y$ a.e. for some Y such that $\int |Y| d\mu < \infty$. And assume either (i) $X_n \to X$ a.e., or (ii) $X_n \to_\mu X$. Then

$$\int |X_n - X| d\mu \to 0 \qquad \text{as } n \to \infty$$

Remark. Note, e.g. Shorack (2000, Chapter 3), that (4) implies that

$$\int X_n d\mu \to \int X d\mu$$

and that

$$\sup_{A \in \mathcal{A}} |\int_A X_n d\mu - \int_A X d\mu| \to 0.$$

This follows from the observation that

$$|\int X_n - \int X| \leq \int |X_n - X_n|$$

and thus uniformly for $A \in \mathcal{A}$,

$$|\int_A X_n - \int_A X| \leq \int_A |X_n - X| \leq \int |X_n - X| \to 0.$$