3 Integration and Expectation

3.1 Construction of the Lebesgue Integral

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space (not necessarily a probability space). Our objective will be to define the *Lebesgue integral* $\int f \, d\mu$ for measurable functions $f : \Omega \to \mathbb{R}$ that meet some additional criteria to be specified later. When the measure μ is a *probability* measure, we will usually write X instead of f and call it a *random variable* rather than a *measurable function*, and we will write EX in place of $\int X \, d\mu$. You should be aware, though, that there are important measure spaces that are not probability spaces.

Example 3.1. Let $\mathbb{N}$ be the set of natural numbers and $\mathscr{F} = 2^{\mathbb{N}}$ be the power set; thus, every function $f : \mathbb{N} \to \mathbb{R}$ is measurable. *Counting measure* on $(\mathbb{N}, 2^{\mathbb{N}})$ is the measure μ defined by $\mu(F) = \#F$, where # denotes cardinality. In this case (as we will show later) the integral will coincide with ordinary summation:

$$\int f \, d\mu = \sum_{n=1}^{\infty} f_n \quad \text{provided} \quad \sum_{n=1}^{\infty} |f_n| < \infty. \tag{3.1}$$

Example 3.2. Let $\Omega = \mathbb{R}$ and $\mathscr{F} = \mathscr{B}_{\mathbb{R}}$. The *Lebesgue measure* on $\mathscr{F}$ is the unique measure whose restriction to each interval $[n, n+1]$ (where $n \in \mathbb{Z}$) is a copy of Lebesgue measure on $[0,1]$, that is

$$\lambda(F) = \sum_{n=-\infty}^{\infty} \lambda|(F \cap [n, n+1]) - n|.$$

In this case the Lebesgue integral $\int f \, d\lambda$ will be an extension of the Riemann integral.

Definition 3.3. A *simple function* $f : \Omega \to \mathbb{R}$ is a measurable function that takes only finitely many (distinct) possible values $a_1, a_2, \ldots, a_m \in \mathbb{R}$. Thus,

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i} \tag{3.2}$$

where $A_1, A_2, \ldots, A_n$ is a measurable partition of Ω. This decomposition is not unique, unless we further require that the scalars a_i are distinct. If f is a *nonnegative* simple function (i.e., the values $a_i \in [0, \infty)$) then we define

$$\int f \, d\mu = \sum_{i=1}^{n} a_i \mu(A_i), \tag{3.3}$$

with the convention that $0 \times \infty = 0$ and $a \times \infty = \infty$ for $a > 0$. (Thus, $\int f \, d\mu$ might be $+\infty$. For *probability* measures μ, it will always be the case that $\int f \, d\mu < \infty$ for nonnegative simple random variables f.) This definition does not depend on the particular decomposition (3.2) (as you should check).

Lemma 3.4. *If $f, g : \Omega \to \mathbb{R}$ are nonnegative simple functions, then for any two nonnegative scalars a, b,*

$$\int (af + bg)\,d\mu = a\int f\,d\mu + b\int g\,d\mu.$$

Proof. Exercise. Note that the restriction to *nonnegative* simple function is to ensure against having an integral be defined as $\infty - \infty$. This difficulty does not arise when μ is a probability (or more generally a finite) measure. □

Definition 3.5. If $f : \Omega \to [0, \infty)$ is a nonnegative, measurable function then its integral is defined by

$$\int f\,d\mu = \sup_{0 \le g \le f} \int g\,d\mu \tag{3.4}$$

where the sup is over all nonnegative *simple* functions g that do not exceed f. Observe that if f is itself simple then the sup is attained by $g = f$, so the definition (3.4) is a valid extension of the definition (3.3) for simple functions.

Proposition 3.6. *The integral has the following properties:*

(a) ***Monotonicity:*** *If $0 \le f \le h$ then $\int f\,d\mu \le \int h\,d\mu$.*
(b) ***Zero function:*** *If $\mu\{f \neq 0\} = 0$ then $\int f\,d\mu = 0$.*
(c) ***Scaling:*** *If f is a nonnegative measurable function and $a \ge 0$ is a scalar then*

$$\int (af)\,d\mu = a\int f\,d\mu.$$

(d) ***Additivity:*** *If f, h are nonnegative measurable functions then*

$$\int (f + h)\,d\mu = \int f\,d\mu + \int h\,d\mu.$$

Proof. Properties (a), (b), (c) are trivial. Property (d) follows from the corresponding property for simple functions (see Lemma 3.4) and the *Monotone Convergence Theorem*, to which we turn next. □

Theorem 3.7. *(Monotone Convergence Theorem) Let $0 \le f_1 \le f_2 \le \cdots$ be a nondecreasing sequence of nonnegative, measurable functions with (pointwise) limit $f := \lim_{n\to\infty} f_n$. Then*

$$\int f\,d\mu = \uparrow \lim_{n\to\infty} \int f_n\,d\mu. \tag{3.5}$$

The proof will rely on a simple case of a general principle of real analysis known (at least to old-timers) as *Littlewood's Third Principle*.[2] The version that we will use is

[2]Littlewood, speaking about measure theory and its use in real analysis, claimed that in fact the whole subject really boiled down to 3 principles:

(1) Every measurable set is nearly a finite union of intervals.
(2) Every measurable function is nearly a continuous function.
(3) Every pointwise convergent sequence of measurable functions is nearly uniformly convergent.

See Royden, *Real Analysis*, sec. 3.6. The first two principles are specific to the real line, but the third holds for general measure spaces.

encapsulated in the following lemma, a weak form of what is sometimes called *Egorov's Theorem.*

Lemma 3.8. *(Egorov's Lemma) Let $A \in \mathscr{F}$ be a set of finite measure $\mu(A) < \infty$, and assume that f_n is a measurable functions that converge pointwise to a (finite) function f on A. Then for every $\varepsilon > 0$ there is a (measurable) set $B \subset A$ with measure $\mu(B) < \varepsilon$ and an integer $n_\varepsilon < \infty$ such that for all $n \geq n_\varepsilon$ and all $\omega \in A \setminus B$,*

$$|f_n(\omega) - f(\omega)| < \varepsilon.$$

Proof. Without loss of generality, we may assume that $A = \Omega$ and that $\mu(\Omega) < \infty$. Let $B_n = \{\omega : \sup_{m \geq n} |f(\omega) - f_m(\omega)| \geq \varepsilon\}$. Since by hypothesis $f_n \to f$ pointwise, it must be the case that $\cap_n B_n = \emptyset$; consequently, since finite measures are continuous from above,

$$\downarrow \lim_{n\to\infty} \mu(B_n) = \mu(\emptyset) = 0.$$

Now choose n_ε so large that $\mu(B_{n_\varepsilon}) < \varepsilon$. □

Proof of the Monotone Convergence Theorem. By hypothesis, the functions f_n converge monotonically up to f, so by the monotonicity of the integral (Proposition 3.6, part (a)) the integrals $\int f_n \, d\mu$ are non-decreasing in n, and hence

$$\lim_{n\to\infty} \int f_n \, d\mu := L \leq \infty$$

exists. Moreover, since $f_n \leq f$ for each n, we must have $L \leq \int f \, d\mu$. Consequently, to prove (3.5) it is enough to show that $L \geq \int f \, d\mu$. But by definition, $\int f \, d\mu$ is the supremum of $\int g \, d\mu$ over all *simple* nonnegative functions $g \leq f$; therefore, to prove the theorem it will suffice to prove that for every simple function $0 \leq g \leq f$,

$$L \geq \int g \, d\mu.$$

Because g is simple, it has the form $g = \sum_{i \leq m} a_i \mathbf{1}_{A_i}$ where each A_i is a measurable set and each $a_i > 0$ is a positive scalar. We must consider two cases: (i) where each set A_i has finite measure, and (ii) where (say) $\mu(A_1) = \infty$. In the first case $\int g < \infty$, whereas in the second $\int g = \infty$.

Case (i): Assume first that each set A_i has finite measure $\mu(A_i) < \infty$. Then $A = \cup_{i=1}^m A_i$ has finite measure, and so Egorov's Lemma can be applied. Egorov implies that for every $\varepsilon > 0$ there exists $B_\varepsilon \subset A$ of measure $\mu(B_\varepsilon) < \varepsilon$ and an integer $n_\varepsilon \geq 1$ such that for all $n \geq n_\varepsilon$,

$$f_n > f - \varepsilon \geq g - \varepsilon \quad \text{on } A \setminus B_\varepsilon.$$

Consequently, for all $n \geq n_\varepsilon$,

$$\begin{aligned}\int f_n\,d\mu &\geq \int f_n(\mathbf{1}_A - \mathbf{1}_{B_\varepsilon})\,d\mu + \int f_n \mathbf{1}_{B_\varepsilon}\,d\mu \\ &\geq \int (g-\varepsilon)(\mathbf{1}_A - \mathbf{1}_{B_\varepsilon})\,d\mu \\ &\geq \int g\,d\mu - \varepsilon\mu(A \setminus B_\varepsilon) + \int g\mathbf{1}_{B_\varepsilon}\,d\mu \\ &\geq \int g\,d\mu - \varepsilon\mu(A) + \varepsilon\|g\|_\infty,\end{aligned}$$

where $\|g\|_\infty = \max_{i \leq m}|a_i|$. Here we have used the fact that each function f_n is nonnegative, and that the support of g is A. Since $\varepsilon > 0$ can be made arbitrarily small, it follows that

$$L = \lim_{n\to\infty}\int f_n\,d\mu \geq \int g\,d\mu.$$

Case (ii): Exercise.

□

The definition of the integral (3.4) as a sup is useful in establishing some of the integral's basic properties, but for other purposes it is sometimes more useful to be able to identify the integral *explicitly* as a limit. The Monotone Convergence Theorem implies that the integral $\int f$ can be approximated from below by $\int f_n$ for *any* increasing sequence f_n converging to f. One such sequence is the following:

$$\begin{aligned} f_n(x) &= \lfloor 2^n f(x) \rfloor / 2^n \quad \text{if } 0 \leq f(x) \leq 2^n; \\ &= 0 \quad \text{otherwise.}\end{aligned} \tag{3.6}$$

Each of the functions f_n is simple (f_n only takes values $k/2^n$ for integers $0 \leq k \leq 2^{2n}$), and clearly $f_n(x)$ increases to $f(x)$ as $n \to \infty$.

Proof of Proposition 3.6 (d). Take any sequences f_n and h_n of nonnegative simple functions converging pointwise to f and h respectively, for instance the sequences defined by (3.6). Then for each n the function $f_n + h_n$ is nonnegative and simple, and the sequence $f_n + h_n$ converges to $f + h$. Monotone convergence implies that

$$\begin{aligned}\int f_n &\to \int f; \\ \int h_n &\to \int h; \quad \text{and} \\ \int (f_n + h_n) &\to \int (f+h).\end{aligned}$$

Since additivity of the integral is easy for simple functions, additivity in general follows.

□

Extension to Arbitrary Measurable Functions. If $f : \Omega \to \mathbb{R}$ is a measurable real-valued function, its positive and negative parts are defined as follows:

$$f_+ = f\mathbf{1}_{\{f\geq 0\}} \quad \text{and} \quad f_- = -f\mathbf{1}_{\{f\leq 0\}} = f - f_+$$

Definition 3.9. The integral of a measurable real-valued function f is defined by

$$\int f\,d\mu = \int f_+\,d\mu - \int f_-\,d\mu$$

unless both of these are $+\infty$, in which case $\int f\,d\mu$ is not defined. Clearly, if $f \geq 0$ then this definition coincides with the definition (3.4). The integral of a measurable *complex*-valued function $f = u + iv$ is defined by

$$\int f\,d\mu = \int u\,d\mu + i\int v\,d\mu$$

provided both integrals on the right are well-defined and finite. If the integral $\int |f|\,d\mu$ is finite then we write $f \in L^1(\mu)$ or $f \in L^1$, and we define the L^1*−norm* of f to be

$$\|f\|_1 = \int |f|\,d\mu.$$

Proposition 3.10. *The sum of two L^1 functions (real or complex) is in L^1, and the L^1−norm satisfies the triangle inequality:*

$$\|f+g\|_1 \leq \|f\|_1 + \|g\|_1.$$

Furthermore, the integral is linear: for any scalars a, b and any measurable functions $f, g \in L^1$,

$$\int (af+bg)\,d\mu = a\int f\,d\mu + b\int g\,d\mu$$

Proof. The triangle inequality follows from the monotonicity of the integral for nonnegative functions (Proposition 3.6 (a)) and the fact that the absolute value for complex numbers satisfies the triangle inequality $|a+b| \leq |a|+|b|$. The linearity of the integral follows routinely from properties (c)-(d) of Proposition 3.6. □

The triangle inequality implies that $\|\cdot\|_1$ is a *pseudo-norm* on the vector space L^1 of all measurable functions with finite integrals. When referring to elements of L^1 one should always keep in mind that they are really *equivalence classes* of functions; however, it is customary to abuse the terminology and refer to individual functions a elements of L^1.

3.2 The Fatou and Dominated Convergence Theorems

The Fatou Lemma and the Dominated Convergence Theorem are, together with the Monotone Convergence Theorem, the basic results in the limit theory of the integral. The Fatou Lemma is itself an easy consequence of the Monotone Convergence Theorem.

Theorem 3.11. *(Fatou's Lemma) For any sequence f_n of nonnegative, measurable functions*

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu.$$

Proof. For each n define $g_n = \inf_{m\ge n} f_m$. Clearly, $0 \le g_n \le f_n$ for each n; furthermore, the sequence g_n is non-decreasing in n, with

$$\lim_{n\to\infty} g_n = \liminf f_n.$$

This latter fact, together with the Monotone Convergence Theorem, implies that

$$\lim_{n\to\infty} \int g_n \, d\mu = \int \liminf_{n\to\infty} f_n \, d\mu.$$

Now for every $n \ge 1$,

$$\int g_n \, d\mu \le \int f_n \, d\mu,$$

since $0 \le g_n \le f_n$; consequently,

$$\lim_{n\to\infty} \int g_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu.$$

□

You should be aware that the inequality in Fatou's Lemma can be strict. Here is the basic example: let the measure space be the unit interval with Lebesgue measure, and set

$$f_n = n\mathbf{1}_{[0,1/n]}.$$

Clearly, $f_{n\to 0}$ everywhere except at $x = 0$, but $\int f_n = 1$ for every n. Thus, additional conditions are needed to ensure that pointwise convergence of functions (or random variables) implies convergence of their integrals (or expectations).

Theorem 3.12. *(Dominated Convergence Theorem) If f_n are measurable functions such that $f = \lim_{n\to\infty} f_n$ exists pointwise and such that for some nonnegative measurable function $g \in L^1$,*

$$|f_n(x)| \le g(x) \quad \textit{for all } x \in \Omega, \tag{3.7}$$

then

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu \quad \textit{and} \quad \lim_{n\to\infty} \int |f_n - f| \, d\mu = 0 \tag{3.8}$$

Proof. Since each f_n is dominated by g the functions $g - f_n$ are nonnegative, and so are the functions $g + f_n$. Since $f_n \to f$ pointwise, the sequence $g - f_n$ converges to $g - f$ pointwise, and the sequence $g + f_n$ converges to $g + f$. Therefore, Fatou's Lemma can be applied to each sequence; this (together with the linearity of the integral) gives

$$\int g\,d\mu - \int f\,d\mu \le \liminf\left(\int g\,d\mu - \int f_n\,d\mu\right) \quad \text{and}$$
$$\int g\,d\mu + \int f\,d\mu \le \liminf\left(\int g\,d\mu + \int f_n\,d\mu\right).$$

Thus,

$$\limsup \int f_n\,d\mu \le \int f\,d\mu \quad \text{and}$$
$$\liminf \int f_n\,d\mu \ge \int f\,d\mu.$$

This proves the first statement in (3.8).

The proof of the second assertion is similar. Since each f_n is dominated by g, so is f; consequently, by the triangle inequality, $|f_n - f| \le 2g$. Now $\lim_{n\to\infty}(2g - |f_n - f|) = 2g$ pointwise, so by Fatou's Lemma,

$$\int 2g\,d\mu \le \liminf \int (2g - |f_n - f|\,d\mu) = \int 2g\,d\mu - \limsup \int |f_n - f|\,d\mu \quad \Longrightarrow$$
$$0 \ge \limsup \int |f_n - f|\,d\mu.$$

But since each integrand $|f_n - f|$ is nonnegative, the liminf of the sequence $\int |f_n - f|\,d\mu$ must be nonnegative also, and so it follows that

$$0 = \lim \int |f_n - f|\,d\mu.$$

This proves the second part of (3.8).

□

Sets of Measure Zero. The hypotheses of Theorems 3.7 and 3.12 require that the sequence f_n converge *pointwise* to the limit function f. In fact, it is enough that they converge *almost everywhere* (that is, at all points except for those in a set of measure 0). This follows, in essence, from Proposition 3.6 (b), which states that if a function is 0 except on a set of measure 0 then its integral is 0. The argument for the monotone convergence theorem goes as follows; the other theorems can be handled in similar fashion.

Assume that f_n is a sequence of measurable real-valued functions on Ω such that

$$0 \le f_1 \le f_2 \le \cdots \quad \text{on } G$$

where G is a measurable set whose complement G^c has measure 0. Define a new sequence of functions $F_n = f_n \mathbf{1}_G$; then each F_n is a measurable function that agrees with f_n except on G^c, and so $\int F_n = \int f_n$, by Proposition 3.6 (b) and the linearity of the integral. Furthermore,

$$0 \leq F_1 \leq F_2 \leq F_3 \leq \cdots \quad \textit{everywhere};$$

thus, the hypotheses of Theorem 3.7 are satisfied, and so

$$\lim_{n\to\infty} \int f_n \, d\mu = \lim_{n\to\infty} \int F_n \, d\mu = \int \lim_{n\to\infty} F_n \, d\mu.$$

This proves that the conclusions of the Monotone Convergence Theorem hold for the sequence f_n.

□

3.3 Measures Defined by Likelihood Ratios

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and $g \geq 0$ a nonnegative measurable function on Ω. The function g can be used as a *density* against μ to define another measure ν on Ω as follows:

$$\nu(F) = \int g \mathbf{1}_F \, d\mu \tag{3.9}$$

Proposition 3.13. *The set function ν on $\mathscr{F}$ is a measure. Moreover, a nonnegative measurable function h on Ω is integrable with respect to ν if and only if (hg) is integrable with respect to μ, and*

$$\int h \, d\nu = \int (gh) \, d\mu \tag{3.10}$$

Proof. The first assertion follows by the Monotone Convergence Theorem. For the second, first check that the formula (3.10) holds for nonnegative simple functions, and then use Monotone Convergence to get it for everything else. □

Observe that if $\int g \, d\mu = 1$ then the measure defined by (3.9) is a *probability* measure. Many of the important parametric families of probability distributions in statistics are built this way. A special case of particular interest is that of an *exponential family* of probability measures. First, a necessary inequality:

Proposition 3.14. *(Jensen's inequality) Let $(\Omega, \mathscr{F}, P)$ be a probability space and $X \in L^1$ an integrable real-valued random variable. If $\varphi : \mathbb{R} \to \mathbb{R}$ is* convex *then the integral of $\varphi(X)$ is well-defined (although possibly infinite) and*

$$\varphi(EX) \leq E\varphi(X).$$

Proof. Support lines. □

Example: Exponential Families. Suppose now that X is a real-valued random variable on a probability space $(\Omega, \mathscr{F}, P)$; then the set of real numbers θ for which $Ee^{\theta X} < \infty$ is an *interval* J containing $\theta = 0$ (by Jensen). For any $\theta \in J$, define

$$\varphi(\theta) = e^{\psi(\theta)} = Ee^{\theta X};$$

then φ is called the *moment generating function* of X and ψ is the *cumulant generating function* of X. By Jensen's inequality, the moment generating function φ and the cumulant generating function ψ are both convex functions. Now fix $\theta \in J$ and define

$$P_\theta(F) = Ee^{\theta X - \psi(\theta)} \mathbf{1}_F \quad \text{for } F \in \mathscr{F}.$$

For each $\theta \in J$ the set function P_θ is a probability measure on $(\Omega, \mathscr{F})$. The family of probabilities $\{P_\theta\}_{\theta \in J}$ is called a (one-dimensional) *exponential family.* The random variable X is the *sufficient statistic,* and θ is the *natural parameter.*

Exercise 3.15. Verify that each of the following is an exponential family:

(a) P_θ = normal distribution with mean θ and variance 1.
(b) P_θ = Poisson distribution with mean e^θ.
(c) P_θ = Binomial (n, p) with $e^\theta = p/(1-p)$.

Proposition 3.16. *Let θ be a point in the* interior *of J. Then for each $k = 1, 2, \dots,$*

$$E_\theta X^k = \left(\frac{d}{d\theta}\right)^k \varphi(\theta).$$

Proof. Dominated Convergence Theorem + Jensen. □

Corollary 3.17. *The mapping $\theta \mapsto \mu_\theta = E_\theta X$ is a smooth, nondecreasing function of θ on the natural parameter space J. If* $\operatorname{Var}_0(X) > 0$ *then $\theta \mapsto \mu_\theta$ is* strictly *increasing, and consequently maps J homeomorphically onto an interval.*

Proof. The first statement is a direct consequence of the preceding proposition. To prove the second assertion, you must show that if $\operatorname{Var}_0(X) > 0$ then $\varphi''(\theta) > 0$ for all $\theta \in J$. □

3.4 Uniform Integrability

Definition 3.18. Let $(\Omega, \mathscr{F}.P)$ be a probability space[3]. A family of $\{X_\theta\}_{\theta \in \Theta}$ of random variables on Ω is *uniformly integrable* if

$$\lim_{\alpha \to \infty} \sup_{\theta \in \Theta} E|X_\theta| \mathbf{1}\{|X_\theta| \geq \alpha\} = 0$$

[3]The notion of uniform integrability also makes sense in infinite measure spaces, but is less useful.

If the family $\{X_\theta\}_{\theta\in\Theta}$ is uniformly integrable then it is bounded in L^1 norm (exercise). If $X \in L^1$ then the singleton family $\{X\}$ is trivially uniformly integrable, by the dominated convergence theorem, because the sequence $|X|\mathbf{1}\{|X| \geq n\}$ converges to 0 pointwise and is dominated by the integrable random variable $|X|$.

Exercise 3.19. If there exists $X \in L^1$ such that $|X_\theta| \leq X$ pointwise for every $\theta \in \Theta$ then the collection $\{X_\theta\}_{\theta\in\Theta}$ is uniformly integrable.

Exercise 3.20. If there is a function $\varphi : [0,\infty) \to [0,\infty]$ such that $\varphi(x)/x \to \infty$ as $x \to \infty$ and if

$$\sup_{\theta\in\Theta} E\varphi(|X_\theta|) < \infty$$

then the collection $\{X_\theta\}_{\theta\in\Theta}$ is uniformly integrable. Thus, for example, if the second moments EX_θ^2 are uniformly bounded then the collection $\{X_\theta\}_{\theta\in\Theta}$ is uniformly integrable.

Theorem 3.21. *(Vitali's Convergence Theorem) Let X_n be a sequence of random variables such that $\lim_{n\to\infty} X_n = X$ pointwise. If the sequence $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable then $X \in L^1$, the collection $\{X\} \cup \{X_n\}_{n\geq 1}$ is uniformly integrable, and*

$$\lim_{n\to\infty} EX_n = EX \quad \textit{and} \quad \lim_{n\to\infty} E|X_n - X| = 0. \tag{3.11}$$

Exercise 3.22. Prove the first two assertions (that $E|X| < \infty$ and that the collection $\{X\} \cup \{X_n\}_{n\geq 1}$ is uniformly integrable).

Proof. The strategy of the proof will be to *truncate* the random variables at some level m (i.e., to replace the original random variables X_n by the random variables $X_n \mathbf{1}_{[0,m](|X_n|)}$ for some large m and then use *bounded* convergence. Unfortunately, though, on the event that $|X| = m$ it is possible that

$$\lim_{n\to\infty} X_n \mathbf{1}_{[0,m]}(|X_n|) = 0 \quad \text{but} \quad X\mathbf{1}_{[0,m]}(|X|) = X.$$

To circumvent this difficulty, we will modify the usual truncation procedure. For each m, let g_m be the piecewise linear function

$$\begin{aligned} g_m(x) &= 1 \quad \text{if } x \leq m; \\ &= 0 \quad \text{if } x \geq m+1; \\ &= m+1-x \quad \text{if } m \leq x \leq m+1. \end{aligned}$$

Clearly, g_m is a continuous function; hence, for each $m \geq 1$ the sequence of random variables $X_n g_m(|X_n|)$ converges pointwise to the random variable $X g_m(|X|)$ as $n \to \infty$. Moreover,

$$|X_n| g(|X_n|) \leq |X_n| \mathbf{1}_{[0,m+1]}(|X_n|)),$$

so for each fixed $m \geq 1$ the random variables $X_n g_m(|X_n|)$ are uniformly bounded. Therefore, by the dominated convergence theorem,

$$\lim_{n\to\infty} EX_n g_m(|X_n|) = EX g_m(|X|).$$

Next, choose $\varepsilon > 0$ small. Since the collection $\{X\} \cup \{X_n\}_{n\geq 1}$ is uniformly integrable (exercise 3.22), there exists $m \geq 1$ so large that

$$\sup_n E|X_n|\mathbf{1}_{[m,\infty)}(|X_n|) + E|X|\mathbf{1}_{[m,\infty)}(|X|) \leq \varepsilon.$$

Now the function g_m coincides with $\mathbf{1}_{[0,m]}$ except in the interval $[m, m+1]$, where it takes values between 0 and 1; hence,

$$|x|(1 - g_m(|x)) \leq |x|(1 - \mathbf{1}_{[0,m](x)}).$$

It follows that

$$\sup_n E|X_n|(1 - g_m(|X_n|)) + E|X|(1 - g_m(|X|)) \leq \varepsilon.$$

Together with the result of the previous paragraph, this implies that

$$\begin{aligned}
&\limsup_{n\to\infty} |EX_n - EX| \\
&\qquad \leq \limsup_{n\to\infty} |EX_n g_m(|X_n|) - EX g_m(|X|)| \\
&\qquad + \sup_n E|X_n|(1 - g_m(|X_n|)) + E|X|(1 - g_m(|X|)) \\
&\qquad \leq 0 + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\lim_{n\to\infty} EX_n = EX$. A similar argument proves that $\lim_{n\to\infty} E|X_n - X| = 0$.

□

3.5 Convergence in Measure and Completeness of L^1

Definition 3.23. Let $(\Omega, \mathscr{F}), \mu$ be a measure space. For any real $\alpha \geq 1$ define the $L^\alpha-$*norm* of a measurable function $f : \Omega \to \mathbb{R}$ by

$$\|f\|_\alpha = \left(\int |f|^\alpha \, d\mu\right)^{1/\alpha}$$

This is not, strictly speaking, a *norm*, because $\|f\|_\alpha = 0$ does not imply that f is identically 0, it only implies that $f = 0$ a.e. For this reason, we shall consider two measurable functions f, g *equivalent* if the set on which they differ has measure 0. The space of equivalence classes of functions f with $\|f\|_\alpha < \infty$ is denoted by L^α (or if the measure space must be emphasized, $L^\alpha(\Omega, \mathscr{F}, \mu)$). It is obvious that for any scalar b,

$$\|bf\|_\alpha = |b|\|f\|_\alpha;$$

thus, to show that $\|\cdot\|_\alpha$ is a norm on L^α it suffices to establish the triangle inequality.

Lemma 3.24. *(Minkowski Inequality)* $\|f+g\|_\alpha \le \|f\|_\alpha + \|g\|_\alpha$.

Proof. In view of the scaling property noted above, it suffices to consider the case where $\|f\|_\alpha = s \in (0,1)$ and $\|g\|_\alpha = 1-s$. By the convexity of the function $x \mapsto |x|^\alpha$,

$$\left|s\frac{f}{s} + (1-s)\frac{g}{1-s}\right|^\alpha \le s|f/s|^\alpha + (1-s)|g/(1-s)|^\alpha.$$

(NOTE: This is where the assumption that $\alpha \ge 1$ is needed, because these are the only values for which the function $x \mapsto |x|^\alpha$ is convex.) Integrating on both sides and using the scaling relations $\|f/s\|_\alpha = \|g/(1-s)\|_\alpha = 1$ gives

$$\int |f+g|^\alpha \le s\int |f/s|^\alpha + (1-s)\int |g/(1-s)|^\alpha = 1.$$

□

For any $\alpha \ge 1$ the space L^α with the norm $\|\cdot\|_\alpha$ is a *normed vector space*. The norm induces a *metric* on L^α by

$$d_\alpha(f,g) := \|f-g\|_\alpha.$$

Our next objective is to show that for every α the metric space $(L^\alpha, d_\alpha$ is *complete*, that is, every Cauchy sequence has a limit. Recall that a sequence f_n is *Cauchy* if for every $\varepsilon > 0$ there exists $n_\varepsilon < \infty$ such that for any two indices $n, m \ge n_\varepsilon$,

$$d_\alpha(f_n, f_m) < \varepsilon.$$

Lemma 3.25. *(Markov Inequality)* $\mu\{|f| \ge C\} \le \|f\|_\alpha^\alpha / C^\alpha$.

□

Proposition 3.26. *If a sequence f_n is* Cauchy in L^α *then there exists a measurable function $f \in L^\alpha$ such that*

$$\lim_{n\to\infty} \|f - f_n\|_\alpha = 0$$

Proof. Since any subsequence of a Cauchy sequence is Cauchy, and since a Cauchy sequence can have at most one limit, it suffices to show that there is a subsequence of f_n

that converges in L^α. For this, choose any subsequence f_k such that $\|f_k - f_{k+1}\|_\alpha \le 4^{-k}$. By the Markov inequality, for any $\varepsilon > 0$,

$$\mu\{|f_k - f_{k+1}| \ge \varepsilon/2^k\} \le \frac{1}{\varepsilon^\alpha 2^{k\alpha}}.$$

Since $\sum_{k\ge 1} 2^{-k\alpha} < \infty$, the Borel-Cantelli Lemma implies that, except on a set B of measure 0, the event $|f_k - f_{k+1}| \ge \varepsilon/2^k$ occurs for only finitely many k. Consequently, except on the set B the sequence f_k converges *pointwise*, and so we may define

$$\begin{aligned} f(\omega) &= \lim_{k\to\infty} f_k(\omega) \quad \text{for } \omega \in B^c, \\ &= 0 \quad \text{for } \omega \in B. \end{aligned}$$

By Fatou's Lemma, since $f_k \to f$ and $|f_k - f| \to 0$ almost everywhere (i.e., except on B),

$$\int |f - f_k|^\alpha \, d\mu \le \liminf_{m\to\infty} \int |f_k - f_{m_k}|^\alpha \, d\mu \le \left(\sum_{n=1}^{\infty} 4^{-n}\right)^\alpha.$$

Thus, $f_k \to f$ in L^α. □

Definition 3.27. A sequence of measurable functions defined on a measure space $(\Omega, \mathscr{F}, \mu)$ is said to *converge in measure* (or *in probability*, if the measure μ is a probability) to a measurable function f if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu\{|f_n - f| \ge \varepsilon\} = 0.$$

Convergence in probability is denoted $X_n \xrightarrow{P} X$. For any real number $\alpha \ge 1$ the sequence f_n *converges in* L^p to f if

$$\lim_{n\to\infty} \int |f_n - f|^p \, d\mu = 0.$$

Exercise 3.28. Show that if $f_n \to f$ almost surely then $f_n \to f$ in measure.

Exercise 3.29. Show that if $f_n \to f$ in L^p then $f_n \to f$ in measure.

Proposition 3.30. *If $f_n \to f$ in measure then there is a subsequence f_{n_k} such that $f_{n_k} \to f$ almost everywhere.*

Proof. Mimic the proof of Proposition 3.26. □