Measures

Definition 1.1.1. A **σ-algebra** $\mathscr{F}$ over a set Ω is a non-empty class of subsets having these properties:

i. Ω and $\emptyset$ belong to $\mathscr{F}$.

ii. If $A \in \mathscr{F}$, then $\Omega n \in \mathscr{F}$.

iii. If (A_n) is countable and $A_n \in \mathscr{F}$, then $\cup_{n=1}^{\infty} A_n \in \mathscr{F}$.

■

Definition 1.1.2. The σ-algebra *generated by* a class of sets $\mathscr{A}$ is the intersection of all σ-algebras containing $\mathscr{A}$. If $(\Omega, \mathcal{T})$ is a *topological space*, the *Borel* σ-algebra $\mathscr{B}$ is that generated by the topology, $\mathscr{B} = \sigma(\mathcal{T})$. ■

Example 1.1.1. The Borel σ-algebra on $\mathbb{R}$ is that generated by the open intervals $\{(a,b) : a \leq b\}$, by the left open right closed intervals $\{(a,b] : a \leq b\}$, and by the semi-infinite intervals $\big\{(-\infty, b]\big\}$. ♠

Definition 1.1.3. A **measure** μ over $(\Omega, \mathscr{F})$ is a set-function $\mu : \mathscr{F} \to [0, \infty]$ satisfying:

i. (*Null measure*): $\mu(\emptyset) = 0$; and

ii. (*Disjoint additivity*) If (A_n) is a countable disjoint collection:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

■

Discussion 1.1.1. The *Borel measure* μ_B defined on the Borel σ-algebra $\mathscr{B}(\mathbb{R})$ is the *unique* measure that assigns to each interval $(a, b]$ its length $\mu_B\{(a,b]\} = b - a$. The Lebesgue σ-algebra $\mathscr{L}(\mathbb{R})$ is a refinement of the Borel σ-algebra that includes all subsets of measure-zero sets. The *Lebesgue measure* λ defined on $\mathscr{L}(\mathbb{R})$ is the likewise the unique measure that assigns to each interval $(a, b]$ its length $\lambda\{(a,b]\} = b - a$. The Borel and Lebesgue measures agree on all sets for which both are defined. However, there exist sets that are Lebesgue but not Borel measurable. ♠

Theorem 1.1.1 (Caratheodory Extension)**.** Suppose $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is a (semi-)algebra and that $P : \mathcal{A} \to [0, 1]$ is a σ-additive set function s.t. $P(\Omega) = 1$; there exists a unique extension of P onto $\sigma(\mathcal{A})$. □

Example 1.1.2. Define an equivalence relation $\sim$ where $x \sim y$ if $x - y \in \mathbb{Q}$. For each $x \in [0, 1]$, assign x to its equivalence class $[x]$. Observe that all rational numbers belong to the same equivalence class, while all numbers of the form $(t + q) \cap [0, 1]$, with $t \in \mathbb{I}$ irrational and $q \in \mathbb{Q}$ rational, belong to another. Now each $x \in [0, 1]$ belongs to exactly one equivalence class, and there are infinitely many such classes. Form a new set V by selecting one representative from each equivalence class. The resulting **Vitali set** is unmeasurable w.r.t. λ. ♠

1.1 Measurability

Definition 1.1.4. A mapping f between *measurable spaces* $(S, \mathscr{S})$ and $(T, \mathscr{T})$ is **measurable** if the pre-image $f^{-1}(B)$ of each measurable set B in the σ-algebra $\mathscr{T}$ on the target space belongs to the σ-algebra on the source space $\mathscr{S}$. ■

Theorem 1.1.2. Suppose $\mathscr{U}$ is a class of subsets generating the σ-algebra $\mathscr{T}$ on the target space: $\sigma(\mathscr{U}) = \mathscr{T}$. A mapping f is measurable $\iff$ $f^{-1}(\mathscr{U}) \subset \mathscr{S}$. □

Lemma 1.1.1. Every continuous and piece-wise continuous function is measurable. ■

Lemma 1.1.2. Suppose (f_n) is a sequence of measurable functions, then:

- $\sup_n f_n(s)$ and $\inf_n f_n(s)$ are measurable.
- $\limsup_n f_n(s)$ and $\liminf_n f_n(s)$ are measurable.
- If $\lim_{n\to\infty} f_n(s)$ exists for $\forall s$ in the domain, then $\lim_{n\to\infty} f_n(s)$ is measurable.

■

Definition 1.1.5. The σ-algebra generated by a random variable $X : (\mathcal{X}, \mathscr{A}) \to (\mathbb{R}, \mathscr{L})$ is the collection of sets:

$$\sigma(X) = \Big[\big\{\omega : X(\omega) \in B\big\} : B \in \mathscr{L}\Big].$$

■

Lemma 1.1.3. If $\mathscr{U}$ is a class of subsets generating $\mathscr{L}$, then:

$$\sigma(X) = \Big[\big\{\omega : X(\omega) \in U\big\} : U \in \mathscr{U}\Big].$$

■

1.2 Probability

Definition 1.1.6. A *probability measure* $\mathbb{P} : \mathscr{F} \to [0, 1]$ is a measure that assigns $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$. A *probability space* if the triple of an even space Ω, a σ-algebra $\mathscr{F}$, and a probability measure $\mathbb{P}$. ■

Definition 1.1.7. A **random variable** is a measurable mapping from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathbb{R}, \mathscr{L})$. The *distribution* induced by a random variable is the set function $P_X(B) = \mathbb{P}\{X^{-1}(B)\}$ for $\forall B \in \mathscr{L}$. ■

Discussion 1.1.2. The distribution P_X is a probability measure on $(\mathbb{R}, \mathscr{L})$, while the base probability measure $\mathbb{P}$ is defined on $(\Omega, \mathscr{F})$. The requirement that the random variable X is measurable ensure that the pre-image $X^{-1}(B)$ of any measurable set $B \in \mathscr{L}$ belongs to the σ-algebra $\mathscr{F}(\Omega)$. ♠

1.3 Density

Definition 1.1.8. Suppose $(\mathcal{X}, \mathscr{A})$ is a measurable space. Let ν and μ denote two measures over $\mathscr{A}$. ν is **absolutely continuous** w.r.t. μ, denoted $\nu \ll \mu$, ν assigns measure zero whenever μ assigns measure zero:

$$\mu(A) = 0 \implies \nu(A) = 0.$$

■

Lemma 1.1.4. $\nu \ll \mu \iff$ for $\forall \epsilon > 0$ there $\exists \delta > 0$ s.t. if $\mu(A) < \delta \implies \nu(A) < \epsilon$. ■

Definition 1.1.9. Measure space $(\mathcal{X}, \mathscr{A}, \mu)$ is *finite* if $\mu(\mathcal{X}) < \infty$, and σ-finite if there exists a countable union $\cup_{n=1}^{\infty} \mathcal{X}_n$ of finite sets such that $\mathcal{X} = \cup_{n=1}^{\infty} \mathcal{X}_n$ and $\mu(\mathcal{X}_n) < \infty$. ■

Theorem 1.1.3 (Radon-Nikodym). Suppose $(\mathcal{X}, \mathscr{A})$ is a measure space, that ν and μ are σ-finite measures on $\mathscr{A}$, and that ν is absolutely continuous w.r.t. μ: $\nu \ll \mu$. Then there exists an a.s. unique, non-negative, measurable *density* f such that:

$$\nu(A) = \int_A f d\mu.$$

f is called the *Radon-Nikodym derivative* of ν w.r.t. μ, and is denoted $f = \frac{d\nu}{d\mu}$. □

Remark 1.1.1. A random variable X that admits a density f_X w.r.t. the Lebesgue measure λ on $(\mathbb{R}, \mathscr{L})$ is described as *absolutely continuous.* A *discrete* random variable is supported on a countable set, and absolutely continuous w.r.t. the counting measure. ♦

Integration

2.1 Lebesgue Integral

Discussion 1.2.1. The idea of Lebesgue integration is to partition the *range*, rather than the domain, of a bounded measurable function $f : S \to T$ into a finite number of disjoint intervals, $T_i = (t_{i-1}, t_i]$, then approximate the area under the graph $\big(s, f_s\big)$ using a tag-point $t_i^* \in T_i$ and the measure of the pre-image $\mu\{f^{-1}(T_i)\}$. The lower and upper Lebesgue sums are:

$$L(f,T) = \sum_{i=1}^{n} t_{i-1}\mu\{f^{-1}(T_i)\} \qquad U(f,T) = \sum_{i=1}^{n} t_i\mu\{f^{-1}(T_i)\}.$$

A function is Lebesgue integrable if the infimum of the upper sum, across all possible partitions of the image, matches the supremum of the lower sum. ♠

Remark 1.2.1. For evaluating Lebesgue integrals, $0 \times \infty = 0$. ♦

Definition 1.2.1 (Indicator Function)**.** The *Lebesgue integral* of an indicator I_A for a measurable set A is:

$$\int_S I_A d\lambda = \int_A d\lambda = \lambda(A).$$

■

Definition 1.2.2 (Simple Function)**.** A *simple function* $f : S \to T$ is one whose image is a *finite* set of non-negative reals $T = \{t_1, \cdots, t_n\}$. Define $S_i = f^{-1}(t_i) = \{s : f(s) = t_i\}$ as the pre-image of t_i. Note that the S_i are *disjoint* and partition the source space: $\cup_{i=1}^{n} S_i = S$. If each S_i is measurable, then the *Lebesgue integral* of f is:

$$\int_S f d\lambda = \sum_{i=1}^{n} t_i\mu(S_i),$$

and the integral over $A \subseteq S$ is:

$$\int_S f I_A d\lambda = \int_A f d\lambda = \sum_{i=1}^{n} t_i\mu(A \cap S_i).$$

■

Definition 1.2.3 (Non-Negative Function)**.** Let $\mathcal{G}$ denote the set of measurable simple functions on S. If f is a non-negative measurable function, then the *Lebesgue integral* is defined as:

$$\int_S f d\lambda = \sup_{g\in\mathcal{G}} \left\{ \int_S g \; d\lambda : 0 \leq g \leq f \right\}.$$

■

Theorem 1.2.1. Suppose f is a non-negative measurable function, then:

- (*Approximation*) There exists a sequence of simple functions (f_n) s.t. $f_n \uparrow f$.
- (*Vanishing*) $f = 0$ a.e. on $A \iff \int_A f d\lambda = 0$.
- (*Monotonicity*) If $f \geq g$ on A, then $\int_A f d\mu \leq \int_A g d\mu$.
- (*Bounding*): $\inf_A(f)\mu(A) \leq \int_A f d\mu \leq \sup_A(f)\mu(A)$.
- (*Triangle inequality*): $|\int_A f d\mu| \leq \int_A |f| d\mu$.

□

Definition 1.2.4 (General Function)**.** Define the positive and negative parts of f as:

$$f^{+}(\omega) = f(\omega) \cdot I\big\{f(\omega) \geq 0\big\}, \qquad f^{-1}(\omega) = -f(\omega) \cdot I\big\{f(\omega) < 0\big\}.$$

Note that the positive and negative parts of f are each non-negative. If f is a measurable function and either the positive or the negative integral is finite:

$$\min\left(\int_S f^{+} d\lambda, \int_S f^{-} d\lambda\right) < \infty,$$

then the *Lebesgue integral* exists and is defined as:

$$\int_S f d\lambda = \int_S f^{+} d\lambda - \int_S f^{-} d\lambda.$$

If the positive $\int_S f^{+} d\lambda$ and negative $\int_S f^{-} d\lambda$ integrals are both finite, then f is described as **integrable**, denoted $f \in \mathcal{L}_1(\Omega)$. ■

Lemma 1.2.1. Suppose $(\Omega, \mathscr{F}, \mu)$ is a measure space and f is a non-negative function, then the mapping $\nu(A) = \int_A f d\mu$ for $A \in \mathscr{F}$ defines a measure $\nu \ll \mu$. ■

2.2 Monotone and Dominated Convergence

Remark 1.2.2. The major theorems of Lebesgue integration are the monotone convergence theorem (MCT) and the dominated convergence theorem (DCT). Each theorem allows for the interchange of a limiting process with integration. ♦

Theorem 1.2.2 (**Monotone Convergence**)**.** If (f_n) is a sequence of non-negative measurable functions with $f_n \leq f$ for $\forall s$ in the domain, and $f_n \uparrow f$ as $n \to \infty$, then:

$$\int_S f_n d\lambda \uparrow \int_S f d\lambda.$$

□

Remark 1.2.3. Recall that $\liminf_{n\to\infty} f_n$ is $\lim_{n\to\infty} g_n$ where $g_n(s) = \inf_{k\geq n} f_k(s)$. Similarly, $\limsup_{n\to\infty} f_n$ is $\lim_{n\to\infty} h_n$ where $h_n(s) = \sup_{k\geq n} f_k(s)$. The limit superior always exceeds the limit inferior:

$$\liminf_{n\to\infty} f_n \leq \limsup_{n\to\infty} f_n.$$

If the limit inferior matches the limit superior, then $\lim_{n\to\infty} f_n$ exists and:

$$\lim_{n\to\infty} f_n = \liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n.$$

♦

Lemma 1.2.2 (Fatou)**.** For a sequence (f_n) of non-negative measurable functions:

$$\int_S \left(\liminf_{n\to\infty} f_n\right) d\lambda \leq \liminf_{n\to\infty} \left(\int_S f_n d\lambda\right).$$

■

Theorem 1.2.3 (**Dominated Convergence**)**.** Let (f_n) and g denote measurable functions. Suppose $f_n \to f$ as $n \to \infty$, and that g is non-negative and integrable. If $|f_n| \leq g$, then (f_n) and f are integrable, and:

$$\lim_{n\to\infty} \left(\int_S f_n d\lambda\right) = \int_S \left(\lim_{n\to\infty} f_n\right) d\lambda = \int_S f d\lambda.$$

□

Corollary 1.2.1. Suppose the sequence (f_n) converges uniformly to f on $[a, b]$:

$$\lim_{n\to\infty} \sup_{s\in[a,b]} \big|f_n(s) - f(s)\big| = 0.$$

If each f_n is integrable, then f is integrable, and:

$$\lim_{n\to\infty} \left(\int_a^b f_n d\lambda\right) = \int_a^b \left(\lim_{n\to\infty} f_n\right) d\lambda = \int_a^b f d\lambda.$$

♣

Theorem 1.2.4 (**Beppo Levi**)**.** Suppose (f_n) is a sequence of measurable functions, and that:

$$\sum_{n=1}^{\infty} \left(\int_S |f_n| d\mu\right) < \infty.$$

Then the series $\sum_{n=1}^{\infty} f_n$ converges absolutely and uniformly, and:

$$\int_S \left(\sum_{n=1}^{\infty} f_n\right) d\mu = \sum_{n=1}^{\infty} \left(\int_S f_n d\mu\right).$$

□

2.3 Reduction to Riemann

Theorem 1.2.5. Suppose $f : [a, b] \to \mathbb{R}$ is bounded:

- f is Riemann integrable $\iff$ it is a.e. continuous w.r.t. the Lebesgue measure. Equally, f is Riemann integrable $\iff$ the set of discontinuities has Lebesgue measure zero.
- If f is Riemann integrable on $[a, b]$, then f is Lebesgue integrable, and the two integrals agree.

□

Lemma 1.2.3. Suppose $f : [a, b] \to \mathbb{R}$ is bounded and monotone, then f is Riemann (and therefore Lebesgue) integrable. ■

2.4 Product Measures

Discussion 1.2.2. Suppose $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ are σ-finite measure spaces. Let $\Omega = \Omega_1 \times \Omega_2$ and define the **product $\boldsymbol{\sigma}$-algebra** as:

$$\mathscr{F} = \sigma\big\{A_1 \times A_2 : A_1 \in \mathscr{F}_1, A_2 \in \mathscr{F}_2\big\}.$$

The goal is to construct a measure μ on $\mathscr{F}$ such that if $A_1 \in \mathscr{F}_1$ and $A_2 \in \mathscr{F}_2$, then:

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

♠

Theorem 1.2.6. Suppose μ_1, μ_2 are σ-finite. For $A \in \mathscr{F}$, define the *sections*:

$$A(\omega_1, \bullet) = \big\{\omega_2 : (\omega_1, \omega_2) \in A\big\}, \qquad A(\bullet, \omega_2) = \big\{\omega_1 : (\omega_1, \omega_2) \in A\big\},$$

Then the following integrals are equivalent:

$$I_1 = \int_{\Omega_1} \mu_2\big\{A(\omega_1, \bullet)\big\} d\mu_1(\omega_1), \qquad I_2 = \int_{\Omega_2} \mu_1\big\{A(\bullet, \omega_2)\big\} d\mu_2(\omega_2).$$

and this common quantity $\mu(A)$ is defined as the **product measure**:

$$\mu(A) \equiv I_1 = I_2.$$

□

Theorem 1.2.7 (Fubini). Suppose μ_1, μ_2 are σ-finite and $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is product measurable. If either of the following integrals is finite:

$$J_1 = \int_{\Omega_1} \left(\int_{\Omega_2} \big|f(\omega_1, \omega_2)\big| d\mu_2(\omega_2) \right) d\mu_1(\omega_1),$$
$$J_2 = \int_{\Omega_2} \left(\int_{\Omega_1} \big|f(\omega_1, \omega_2)\big| d\mu_1(\omega_1) \right) d\mu_2(\omega_2),$$

then both are finite, the following integrals are equivalent:

$$I_1 = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1), \quad I_2 = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2)$$

and this common quantity is defined as the **product integral**:

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d\mu(\omega_1, \omega_2) \equiv I_1 = I_2.$$

□

2.5 Lebesgue-Stieltjes Integral

Theorem 1.2.8.

i. Suppose μ is a σ-finite measure on $(\mathbb{R}, \mathscr{L})$. The function G defined, up to an additive constant, by $G(b) - G(a) = \mu(a, b]$ is *cádlág* and non-decreasing.

ii. Suppose $G : \mathbb{R} \to \mathbb{R}$ is *cádlág* and non-decreasing. For any finite interval $(a, b]$, let $\mu(a, b] = G(b) - G(a)$. There exists a unique extension of μ onto $(\mathbb{R}, \mathscr{L})$.

□

Definition 1.2.5. Suppose $f : \mathbb{R} \to \mathbb{R}$ is measurable, G is *cádlág* non-decreasing, and μ_G is the measure on $(\mathbb{R}, \mathscr{L})$ induced by G. The **Lebesgue-Stieltjes integral** of f w.r.t. the *integrator* G is defined as the Lebesgue integral of f w.r.t. the measure μ_G:

$$\int_A f(\omega) dG(\omega) = \int_A f(\omega) d\mu_G(\omega).$$

■

Definition 1.2.6. Define the **total variation** of a function G on an interval $[a, b]$ as:

$$TV_G[a, b] = \sup_{\mathcal{P}[a,b]} \left\{ \sum_{k=1}^{n} \big|G(\omega_k) - G(\omega_{k-1})\big| \right\},$$

where the supremum is taken across all finite partitions $a = \omega_0 < \cdots < \omega_n = b$ of the interval $[a, b]$. The function G is of **bounded variation** on $[a, b]$ if the total variation is finite: $TV_G[a, b] < \infty$. ■

Discussion 1.2.3. If G is a *cádlág* function of *bounded variation* then there exists a decomposition of G as $G_1 - G_2$, were G_1 and G_2 are both *cádlág* and non-decreasing. The *Lebesgue-Stieltjes* integral of a bounded, measurable function f w.r.t. G is then:

$$\int_A f(\omega)dG(\omega) = \int_A f(\omega)dG_1(\omega) - \int_A f(\omega)dG_2(\omega).$$

♠

Example 1.2.1.

i. Suppose the integrator G is a non-decreasing step function with jumps at (ω_k). Define $\Delta G_n(\omega_n)$ as $G(\omega_n) - G(\omega_n-) > 0$. Then the measure μ_G induced by G is discrete, and the Lebesgue-Stieltjes integral of f w.r.t. G is:

$$\int_a^b f(\omega)dG(\omega) = \sum_{\{\omega_k \in (a,b]\}} f(\omega_k)\Delta G(\omega_k).$$

ii. Suppose the integrator G has derivative $g > 0$ at each $\omega \in (a, b]$, then the measure μ_G induced by G is *absolutely continuous* w.r.t. the Lebesgue measure, and the Lebesgue-Stieltjes integral of f w.r.t. G is:

$$\int_a^b f(\omega)dG(\omega) = \int_a^b f(\omega) \cdot g(\omega)d\omega.$$

iii. Suppose the integrator G is increasing and differentiable except for at most countably many jump discontinuities (ω_k). The Lebesgue-Stieltjes integral of f w.r.t. G is:

$$\int_a^b f(\omega)dG(\omega) = \int_a^b f(\omega)g(\omega)d\omega + \sum_{\{\omega_k \in (a,b]\}} f(\omega_k)\Delta G(\omega_k).$$

♠

Statistical Properties

3.1 Independence

Definition 1.3.1. Suppose $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space. Events A and B are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. ■

Theorem 1.3.1 (Borel Cantelli)**.**

- If (A_n) is any sequence of events and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then:

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$

- If (A_n) are *independent* events and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then:

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

□

Definition 1.3.2. The tail σ-field generated by a sequence (X_n) of random variables is:

$$\mathscr{T} = \bigcap_{n=1}^{\infty} \sigma\big(X_n, X_{n+1}, \cdots\big).$$

■

Theorem 1.3.2 (Kolmogorov's Zero-One Law)**.** If (X_n) is a sequence of *independent* random variables, and $A \in \mathscr{T}$ is a *tail field* event, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. □

Corollary 1.3.1.

- The event that $\sum_{n=1}^{\infty} X_n$ converges has probability 0 or 1.
- The $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are constant with probability 1.

♣

Definition 1.3.3. Suppose X_1 maps $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\mathcal{X}_1, \mathscr{A}_1)$ and X_2 maps $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\mathcal{X}_2, \mathscr{A}_2)$. X_1 is **independent** of X_2 if:

$$\mathbb{P}(X_1 \in A_1 \cap X_2 \in A_2) = \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2).$$

■

Definition 1.3.4. A **cumulative distribution function** (CDF) F is a *cádlág* (continuous on the right with limits on the left), non-decreasing function, satisfying:

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1.$$

■

Lemma 1.3.1. If $F : \mathbb{R} \to [0, 1]$ is a CDF, there exists a random variable $X(\omega)$ on $([0, 1], \mathscr{L}, \lambda)$ whose distribution is F. ■

Theorem 1.3.3 (Factorization). The random variables X_1 and X_2 are independent $\iff$ the joint distribution function factors as the product of marginals:

$$P(X_1 \le \xi_1, X_2 \le \xi_2) = P(X_1 \le \xi_1)P(X_2 \le \xi_2).$$

□

3.2 Expectation

Definition 1.3.5. The **expectation** of a random variable X mapping from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to the measure space $(\mathbb{R}, \mathscr{L})$ is defined as:

$$E(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega).$$

If $A \in \mathscr{F}$, then the expectation over A is:

$$E(XI_A) = \int_A X(\omega) d\mathbb{P}(\omega).$$

■

Discussion 1.3.1. Let $F_X(x) = \mathbb{P}\big\{X^{-1}(-\infty, x]\big\}$ denote the CDF of a random variable X. The expectation is expressible as:

$$E(X) = \int_\mathcal{X} x \cdot dF_X(x).$$

If the distribution P_X admits a density f_X w.r.t. measure μ on $(\mathbb{R}, \mathscr{L})$, then:

$$E(X) = \int_\mathcal{X} x \cdot f_X(x) d\mu(x).$$

♠

Example 1.3.1. Suppose X is a simple random variable on $(\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{L})$, then X is expressible as a linear combination of indicators:

$$X(\omega) = \sum_{j=1}^{J} x_j I_{A_j}(\omega),$$

where $A_j \subseteq \Omega$ are disjoint measurable sets, and $\cup_{j=1}^{J} A_j = \Omega$. The expectation is:

$$E(X) = \int_{\mathcal{X}} X(\omega) d\mathbb{P}(\omega) = \sum_{j=1}^{J} x_j \int I_{A_j}(\omega) d\mathbb{P}(\omega) = \sum_{j=1}^{J} x_j \mathbb{P}(A_j).$$

♠

Remark 1.3.1. The MCT and DCT for random variables give conditions under which a limiting process may be exchanged with expectation. The MCT requires monotone convergence and non-negativity. The DCT relaxes monotonicity but requires the existence of a dominating, integrable random variable. ♦

Theorem 1.3.4 (Monotone Convergence). If $X(\omega) \leq Y(\omega)$, then $E(X) \leq E(Y)$. Moreover, if $0 \leq X_n(\omega)$ and $X_n(\omega) \uparrow X(\omega)$, then:

$$\lim_{n\to\infty} E(X_n) = E\left(\lim_{n\to\infty} X_n\right).$$

□

Theorem 1.3.5 (Dominated Convergence). If $X_n(\omega) \to X(\omega)$ and there exists random variable Z s.t. $E|Z| < \infty$, then:

$$\lim_{n\to\infty} E(X_n) = E\left(\lim_{n\to\infty} X_n\right),$$

and $\lim_{n\to\infty} E|X_n - X| = 0$. □

3.3 Conditional Expectation

Definition 1.3.6. Suppose $Y(\omega)$ is a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathscr{G} \subseteq \mathscr{F}$ is a sub-σ-algebra. A random variable $Z(\omega)$ that is:

i. measurable w.r.t. $\mathscr{G}$, i.e. for $\forall B \in \mathscr{L}$, $Z^{-1}(B) \in \mathscr{G}$; and

ii. satisfies the integral identity:

$$\int_G Y d\mathbb{P} = \int_G Z d\mathbb{P}, \text{ for } \forall G \in \mathscr{G}$$

is the **conditional expectation** of Y w.r.t. $\mathscr{G}$, denoted $Z = E(Y|\mathscr{G})$. ■

Example 1.3.2. Suppose $X(\omega)$ and $Y(\omega)$ are random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, and that $E(Y^2) < \infty$. Recall that the σ-algebra generated by X is:

$$\sigma(X) = \Big[\big\{\omega : X(\omega) \in B\big\} : B \in \mathscr{L}\Big].$$

Denote $\sigma(X)$ by $\mathscr{G}$, and let Z denote the orthogonal projection of Y onto the space $\mathcal{L}^2(\mathscr{G})$ of square-integrable, $\mathscr{G}$-measurable functions. Then for $\forall Z^* \in \mathcal{L}^2(\mathscr{G})$:

$$\langle Y - Z, Z^* \rangle = \int_\Omega (Y - Z) Z^* d\mathbb{P} = 0,$$

and in particular for $Z^* = I_G$ where $G \in \mathscr{G}$:

$$\int_G Y d\mathbb{P} = \int_G Z d\mathbb{P}.$$

Thus $Z = E(Y|\mathcal{G})$ may be interpreted as the orthogonal projection of Y onto the space of square-integrable, $\sigma(X)$-measurable functions. ♠

Remark 1.3.2. For a continuous random variable $Y(\omega)$, standard notation $E(Y|X)$ denotes the random variable:

$$\big\{E(Y|X)\big\}(\omega) = \int_\mathbb{R} y \cdot f\big\{y|X(\omega)\big\} dy$$

that is 1. measurable $\sigma(X)$, and 2. satisfies:

$$E\big\{Y I_A\big\} = E\big\{E(Y|X) I_A\big\},$$

for $\forall A \in \sigma(X)$. Since Ω always belongs to $\sigma(X)$, the law of iterated expectation follows immediately: $E(Y) = E\big\{E(Y|X)\big\}$. ♦