Probability Theory

Rebecca Sela

September 28, 2005

1 Sets, Fields, Measures, and Probability Measures

Definition Let Ω be a probability space. We call $\omega \in \Omega$ a *sample point* and $S \subset \Omega$ an *event*.

Definition A class $\mathcal{F}$ of subsets of Ω is called a *field* or *algebra* if:

- $\Omega \in \mathcal{F}$
- If $A \in \mathcal{F}$ then $A^C \in \mathcal{F}$.
- If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$. (This is called *finite additivity.*)

The class is a *σ-field* or a *σ-algebra* if the following condition holds as well:

- If $A_1, A_2, ... \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ as well. (This is called *countable additivity.*)

An set that is an element of $\mathcal{F}$ is called an *$\mathcal{F}$-set*, and is said to be *measurable* $\mathcal{F}$. The σ-field *generated by* a class of sets, $\mathcal{A}$, $\sigma(\mathcal{A})$, is the intersection of all σ-fields that contain $\mathcal{A}$.

Definition The *extended real line* is $[-\infty, \infty]$; it includes both positive and negative infinity.

Definition Let $\mathcal{R}^k$ be the σ-field generated by the bounded rectangles $[x = (x_1, ..., x_k) : a_i \leq x_i \leq b_i, i = 1, ..., k] \subset R^k$. The elements of $\mathcal{R}^k$ are called the k-dimensional *Borel sets*. Note that $\mathcal{R}^k$ contains all the open and closed sets (among other things). $\mathcal{R}^1$ is sometimes written as $\mathcal{B}$.

Theorem 1.1 *If $\mathcal{A}$ is a class of sets in Ω and $\Omega_0 \subset \Omega$, let $\mathcal{A} \cap \Omega_0 = [A \cap \Omega_0 : A \in \mathcal{A}]$. If $\mathcal{F}$ is a σ-field in Ω then $\mathcal{F} \cap \Omega_0$ is a σ-field in Ω_0. If $\mathcal{A}$ generates the σ-field $\mathcal{F}$ in Ω then $\mathcal{A} \cap \Omega_0$ generates the σ-field $\mathcal{F} \cap \Omega_0$ in Ω_0. That is, $\sigma(\mathcal{A} \cap \Omega_0) = \sigma(\mathcal{A}) \cap \Omega_0$.*

Definition A class $\mathcal{P}$ of subsets of Ω is a *π-system* if whenever $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P}$ as well.

Definition A class $\mathcal{L}$ of subsets of Ω is a λ-*system* if:

1. $\Omega \in \mathcal{L}$
2. If $A, B \in \mathcal{L}$ and $A \subset B$, then $B - A \in \mathcal{L}$.
3. If $A_1, A_2, ... \in \mathcal{L}$ and $A_n \uparrow A$ then $A \in \mathcal{L}$.

Note that any class of subsets that is both a π-system and a λ-system is a σ-field.

Theorem 1.2 *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system and $\mathcal{P} \subset \mathcal{L}$ then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Definition A *set function* is a real-valued function defined on a class of subsets of Ω. A set function $\mu : \mathcal{F} \to R$, where $\mathcal{F}$ is a field in Ω, is a *measure* if:

1. $\mu(A) \in [0, \infty]$ for all $A \in \mathcal{F}$
2. $\mu(\emptyset) = 0$
3. If $A_1, A_2, ... \in \mathcal{F}$ are disjoint and $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, then $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$. (Note that this sum may be infinite.)

μ is *finite* if $\mu(\Omega) < \infty$ and *infinite* if $\mu(\Omega) = \infty$. If $\Omega = A_1 \cup A_2 \cup ...$ (where this is either a finite or a countable sequence of $\mathcal{F}$-sets, which need not be disjoint) with $\mu(A_k) < \infty$ for all k, then μ is σ-*finite*. If $A_1, A_2, ... \in \mathcal{A}$, then we say μ is σ-finite on $\mathcal{A}$. If μ is a measure on a σ-field $\mathcal{F}$ in Ω, we say that the triple, $(\Omega, \mathcal{F}, \mu)$ is a *measure space*. If $\mu(A^C) = 0$ for $A \in \mathcal{F}$, then A is a *support* of μ, and we say that μ is *concentrated on* A. If $\mathcal{F}$ is a σ-field on Ω, then we say that $(\Omega, \mathcal{F})$ is a *measurable space*.

Definition A measure μ on $(\Omega, \mathcal{F})$ is *discrete* if there are countably many $\omega_i \in \Omega$ such that $\mu(A) = \sum_{\omega_i \in A} \mu(\{\omega_i\})$ for all $A \in \mathcal{F}$.

Some facts about measures:

- Monotonicity: If $A \subset B$ then $\mu(A) \leq \mu(B)$.
- Inclusion-Exclusion: $\mu(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i \cap A_j) + ... + (-1)^{n+1} \mu(A_1 \cap ... \cap A_n)$
- Finite Subadditivity: $\mu(\bigcup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} \mu(A_k)$
- If $\mu(A_n) = 0$ for all A_n then $\mu(\bigcup_n A_n) = 0$.

Definition For an infinite sequence of real numbers, $x_1, x_2, ...$, in $[0, \infty]$, we say that $x_k \uparrow x$ if $x_k \leq x_{k+1} \leq x$ and $x_k \to x$ (either in the usual sense, or if $x = \infty$ and $x_k = \infty$ for some k). We say that $x_k \downarrow x$ if $x_k \geq x_{k+1} \geq x$ and $x_k \to x$.

Definition For an infinite sequence of sets, $A_1, A_2, ...$, we say that $A_n \uparrow A$ if $A_1 \subset A_2 \subset ...$ and $A = \bigcup_n A_n$. We say that $A_n \downarrow A$ if $A_1 \supset A_2 \supset ...$ and $A = \bigcap_n A_n$.

Theorem 1.3 *Let μ be a measure on a field $\mathcal{F}$. Then it has the following properties:*

1. *Continuity from Below: If $A_n, A \in \mathcal{F}$ and $A_n \uparrow A$ then $\mu(A_n) \uparrow \mu(A)$.*
2. *Continuity from Above: If $A_n, A \in \mathcal{F}$ and $A_n \downarrow A$ and $\mu(A_1) < \infty$ then $\mu(A_n) \downarrow \mu(A)$.*
3. *If $A_1, A_2, ... \in \mathcal{F}$ and $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ then $\mu(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} \mu(A_k)$.*
4. *If μ is σ-finite on $\mathcal{F}$, then $\mathcal{F}$ cannot contain an uncountable, disjoint collection of sets of positive measure.*

Definition Let $\lambda(a, b] = b - a$ for any $a, b \in R$. Note that λ is finitely additive and countably subadditive on the collection of finite intervals in R. The extension of this measure to the σ-field of all linear Borel sets in R^1, $\mathcal{B}$, defines the *Lebesgue measure* on R. The Lebesgue measure is indeed a measure on this set, and its extension from intervals to the entire class of sets is unique.

Theorem 1.4 (Extension Theorem.) *Suppose P is a measure on a field $\mathcal{F}_0$ of subsets of Ω. Let $\mathcal{F} = \sigma(\mathcal{F}_0)$. Then there is a unique probability measure, Q on $\mathcal{F}$ such that $Q(A) = P(A)$ for all $A \in \mathcal{F}_0$.*

Theorem 1.5 *Suppose P_1 and P_2 are probability measures on $\sigma(\mathcal{P})$, with $\mathcal{P}$ a π-system. If P_1 and P_2 agree on $\mathcal{P}$ then they agree on all of $\sigma(\mathcal{P})$.*

Definition Let $\lambda_k[x = (x_1, ..., x_k) : a_i \leq x_i \leq b_i, i = 1, ..., k] = \prod_{i=1}^{k}(b_i - a_i)$ for any bounded rectangle in R^k. The extension of this to the Borel subsets of R^k defines the k-dimensional Lebesgue measure.

Theorem 1.6 Translation Invariance *If $A \in \mathcal{R}^k$ and we define $A + x = [a + x : a \in A]$, then $A + x \in \mathcal{R}^k$ and $\lambda_k(A) = \lambda_k(A + x)$.*

Theorem 1.7 *If $T : R^k \to R^k$ is linear and nonsingular, then for all $A \in \mathcal{R}^k$, $T(A) \in \mathcal{R}^k$ and $\lambda_k(T(A)) = |det(T)|\lambda_k(A)$.*

Definition A set function $P : \mathcal{F} \to R$ is a *probability measure* if:

1. $0 \leq P(A) \leq 1$ for all $A \in \mathcal{F}$
2. $P(\emptyset) = 0$ and $P(\Omega) = 1$
3. If $A_1, A_2, ... \in \mathcal{F}$ are disjoint and $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, then $P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$.

The triple, $(\Omega, \mathcal{F}, P)$ is called a *probability (measure) space.* A *support* of P is any set $A \in \mathcal{F}$ such that $P(A) = 1$.

Theorem 1.8 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then the following three properties hold:*

1. *Continuity from Below: Suppose $A, A_n \in \mathcal{F}$ with $A_n \uparrow A$. Then, $P(A_n) \uparrow P(A)$.*
2. *Continuity from Above: Suppose $A, A_n \in \mathcal{F}$ with $A_n \downarrow A$. Then, $P(A_n) \downarrow P(A)$.*
3. *Countable Subadditivity: If $A_k \in \mathcal{F}$ for all k and $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, then $P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)$.*

If $P(A_n) = 1$ for all A_n, then $P(\bigcap_n A_n) = 1$.

Definition Let $\Omega = \{1, 2, ...\}$ and $\mathcal{F}$ be the set of all subsets of Ω. If $\mu(A)$ is the number of elements in A, we call μ the *counting measure*.

Definition A property G is true *almost everywhere* or *almost surely* if $\mu[\omega : G \text{ does not hold for } \omega] = 0$.

Definition Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. $A \times B$ is a *measurable rectangle* if $A \in \mathcal{X}$ and $B \in \mathcal{Y}$. $\mathcal{X} \times \mathcal{Y}$ is the σ-field generated by the measurable rectangles.

Theorem 1.9 *If $E \in \mathcal{X} \times \mathcal{Y}$, then for all $x \in X$, $[y : (x, y) \in E] \in \mathcal{Y}$ and for all $y \in Y$, $[x : (x, y) \in E] \in \mathcal{X}$. (These are called the* sections *of E determined by x or y.) If f is measurable $\mathcal{X} \times \mathcal{Y}$ then for any fixed x, $f(x, .)$ is measurable $\mathcal{Y}$, and for any fixed y, $f(., y)$ is measurable $\mathcal{X}$. (These are the* sections *of f determined by x and y.)*

Definition Suppose $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ are measure spaces, with μ, ν finite. Let $E \in \mathcal{X} \times \mathcal{Y}$. Define

$$\begin{aligned} \pi_1(E) &= \int_X \nu[y : (x, y) \in E]\mu(dx) \\ \pi_2(E) &= \int_Y \nu[x : (x, y) \in E]\mu(dy) \end{aligned}$$

In particular, for a measurable rectangle $A \times B$, $\pi_1(A \times B) = \pi_2(A \times B) = \mu(A)\nu(B)$. This measure is called the *product measure*.

Theorem 1.10 *If $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ are σ-finite measure spaces, then $\pi_1(E) = \pi_2(E) = \pi(E)$ defines a σ-finite measure on $\mathcal{X} \times \mathcal{Y}$. It is the only measure such that $\pi(A \times B) = \mu(A)\nu(B)$.*

Measurable Functions and Mappings

Definition Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces. For $T : \Omega \to \Omega'$, define $T^{-1}(A') = [\omega \in \Omega : T(\omega) \in A']$ (this is the inverse image of $A' \subset \Omega'$). T is *measurable* $\mathcal{F}/\mathcal{F}'$ if $T^{-1}(A') \in \mathcal{F}$ for all $A' \in \mathcal{F}'$. If $f : \mathcal{F} \to R$ is measurable $\mathcal{F}/R$, then we say it is measurable $\mathcal{F}$.

Theorem 1.11 *If $T^{-1}(A') \in \mathcal{F}$ for all $A' \in \mathcal{A}'$ and $\mathcal{A}'$ generates $\mathcal{F}'$, then T is measurable $\mathcal{F}/\mathcal{F}'$. If T is measurable $\mathcal{F}/\mathcal{F}'$ and T' is measurable $\mathcal{F}'/\mathcal{F}''$, then $T' \circ T$ is measurable $\mathcal{F}/\mathcal{F}''$.*

Definition If $f : R^i \to R^k$ is measurable $\mathcal{R}^i/\mathcal{R}^k$, then it is called a *Borel function.*

Theorem 1.12 *If $f(\omega) = (f_1(\omega), ..., f_k(\omega))$, f is measurable $\mathcal{F}$ if and only if each $f_j(\omega)$ is measurable $\mathcal{F}$.*

Theorem 1.13 *If $f : R^i \to R^k$ is continuous, then it is measurable. Since compositions of measurable functions are measurable, sums, products, maxima, and other continuous functions of measurable functions are also measurable.*

Theorem 1.14 *Suppose $f_1, f_2, ...$ are measurable $\mathcal{F}$. Then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, and $\liminf_n f_n$ are all measurable $\mathcal{F}$. If $\lim_n f_n$ exists everywhere, then it is measurable $\mathcal{F}$. The set where $\{f_n(\omega)\}$ converges lies in $\mathcal{F}$. If f is also measurable $\mathcal{F}$, then the set where $f_n(\omega) \to f(\omega)$ lies in $\mathcal{F}$.*

Definition A *simple* real function is a real function with finite range, $\{x_1, ..., x_n\}$. Then we may write it as $f = \sum_{i=1}^n x_i I(A_i)$, where the A_i decompose Ω. This function is measurable if $A_i \in \mathcal{F}$ for $i = 1, ..., n$.

Theorem 1.15 *If f is real and measurable $\mathcal{F}$, there exists a sequence $\{f_n\}$ of simple functions that are measurable $\mathcal{F}$ such that when $f(\omega) \geq 0$ $0 \leq f_n(\omega)$ and $f_n(\omega) \uparrow f(\omega)$ and when $f(\omega) \leq 0$ $0 \geq f_n(\omega)$ and $f_n(\omega) \downarrow f(\omega)$.*

Definition Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable with $T : \Omega \to \Omega'$ measurable $\mathcal{F}/\mathcal{F}'$. Given a measure μ on $\mathcal{F}$, define $\mu T^{-1} : \mathcal{F}' \to R$ by $\mu T^{-1}(A') = \mu(T^{-1}(A'))$ for all $A' \in \mathcal{F}'$. Note that μT^{-1} is a measure. If μ is a probability measure, so is μT^{-1}.

2 Integrals

Definition Let $f = \sum_{i=1}^n x_i I_{A_i}$ with $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^n A_i = \Omega$. We define:

$$\int f d\mu = \sum_{i=1}^n x_i \mu(A_i)$$

If B is an $\mathcal{F}$-set, then we define $\int_B f d\mu = \int_\Omega (f)(I_B) d\mu = \sum_{i=1}^n x_i \mu(A_i \cap B)$.

Definition Let f be a non-negative measurable function. Consider a sequence of simple functions, $\{f_n\}_{n=1}^{\infty}$, such that $0 \leq f_n$ and $f_n \uparrow f$. We define $\int f\, d\mu = \lim_{n\to\infty} \int f_n d\mu$.

Definition Let f be any measurable function. Define $f^+(\omega) = f(\omega)I(f(\omega) > 0)$ and $f^-(\omega) = -f(\omega)I(f(\omega) < 0)$. (Note that $f = f^+ - f^-$.) If $\int f^+\, d\mu < \infty$ and $\int f^-\, d\mu < \infty$, then we say that f is *integrable* and define:

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

Some properties of integrals:

- If $f \leq g$ almost everywhere, then $\int f d\mu \leq \int g d\mu$.
- $|\int f d\mu| \leq \int |f| d\mu$
- If $f = 0$ almost everywhere, then $\int f d\mu = 0$.
- If $\mu(A) = 0$ then $\int_A f d\mu = 0$.
- If $\int f d\mu < \infty$ then $f < \infty$ almost everywhere.
- If α, β are finite and f, g are integrable, then $\int(\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$.

Theorem 2.1 Monotone Convergence Theorem *If* $0 \leq f_n$ *and* $f_n \uparrow f$ *almost everywhere, then* $\int f_n d\mu \uparrow \int f d\mu$.

Theorem 2.2 Fatou's Lemma *If* $0 \leq f_n$ *then*

$$\int (\liminf_{n\to\infty} f_n) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$$

Theorem 2.3 Lebesgue's Dominated Convergence Theorem *If* $|f_n| \leq g$ *almost everywhere,* g *is integrable, and* $f_n \to f$ *almost everywhere, then* $\{f_n\}$ *and* f *are integrable and* $\int f_n d\mu \to \int f d\mu$.

Definition A sequence of functions, f_n is *uniformly bounded* if there exists $K < \infty$ such that $|f_n(\omega)| < K$ for all ω and n.

Theorem 2.4 Bounded Convergence Theorem *If* $\mu(\Omega) < \infty$, $\{f_n\}$ *is uniformly bounded, and* $f_n \to f$ *almost everywhere, then* $\int f_n d\mu \to \int f d\mu$.

Definition A sequence $\{f_n\}$ is *uniformly integrable* if

$$\lim_{\alpha\to\infty} (\sup_n \int_{|f_n| \geq \alpha} |f_n| d\mu) = 0$$

Theorem 2.5 *If* $\sup_n \int |f_n|^{1+\epsilon} d\mu < \infty$ *for some* $\epsilon > 0$ *then* $\{f_n\}$ *is uniformly integrable.*

Theorem 2.6 *Let $\mu(\Omega) < \infty$ and $f_n \to f$ almost everywhere. If the f_n are uniformly integrable, then f is integrable and $\int f_n d\mu \to \int f d\mu$.*

Definition A real measurable function, f, is *Lebesgue integrable* if it is integrable with respect to the Lebesgue measure, λ. The integral is usually written as $\int f d\lambda = \int f(x)dx$.

Definition A real function, f, on an interval $(a, b]$ is *Riemann integrable* with integral r if for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|r - \sum_j f(x_j)\lambda(J_j)| < \epsilon$$

where $\{J_j\}$ is a finite partition of $(a, b]$ into subintervals of length at most δ and $x_j \in J_j$.

A bounded function on a bounded interval if Riemann integrable if and only if its set of discontinuities has Lebesgue measure 0. If f is Riemann integrable, then the Riemann integral equals the Lebesgue integral.

Theorem 2.7 *Suppose $\int |f|dx < \infty$. Then for all $\epsilon > 0$ there exists a step function, $g_\epsilon = \sum_{i=1}^k y_i I_{A_i}$ (where the A_i are bounded intervals) such that $\int |f - g|dx < \epsilon$. In addition, there is a continuous integrable function, h_ϵ, which is 0 outside a bounded interval, such that $\int |f - h|dx < \epsilon$.*

Theorem 2.8 Fundamental Theorem of Calculus. *Suppose F is a function with continuous derivative, $F' = f$. Then, $\int_a^b f(x)dx = \int_a^b F'(x)dx = F(b) - F(a)$.*

Theorem 2.9 *Suppose $\{x_i\}$ is a non-negative sequence. Define $f(i) = x_i$. Then $\sum_{i=1}^\infty x_i = \int f d\mu$, where μ is the counting measure.*

Theorem 2.10 Fubini's Theorem *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be σ-finite measure spaces. Suppose f is a non-negative function. Then, $\int_Y f(x, y)\nu(dy)$ and $\int_X f(x, y)\mu(dx)$ are measurable $\mathcal{X}$ and $\mathcal{Y}$ respectively, and:*

$$\begin{aligned}
\int_{X\times Y} f(x, y)\pi(d(x, y)) &= \int_X [\int_Y f(x, y)\nu(dy)]\mu(dx) \\
&= \int_Y [\int_X f(x, y)\mu(dx)]\nu(dy)
\end{aligned}$$

Let f be an arbitrary function that is integrable with respect to π. Then, $\int_Y f(x, y)\nu(dy)$ and $\int_X f(x, y)\mu(dx)$ are measurable and finite except on sets of μ- and ν-measure 0 respectively, and the iterated integrals above continue to hold.

3 Random Variables

Definition Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to R$. X is a *simple random variable* if it has a finite range and $[\omega : X(\omega) = x] \in \mathcal{F}$ for all $x \in R$. Let $A_i = \{\omega : X(\omega) = x_i\}$ for each x_i in the range of X. Then we may represent X as a sum of indicator functions:

$$X = \sum_{i=1}^{n} x_i I(A_i)$$

In this case, μ has mass $p_i = P[X = x_i] = \mu\{x_i\}$ at the points in the range of X.

Definition A *random variable* on a probability space $(\Omega, \mathcal{F}, P)$ is a real-valued function $X = X(\omega)$ which is measurable $\mathcal{F}$. A *random vector* is a mapping $X : \Omega \to R^k$ that is measurable $\mathcal{F}$; this is a k-tuple of random variables.

Definition If $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$ then a simple random variable X is *measurable* $\mathcal{G}$ if $[\omega : X(\omega) = x] \in \mathcal{G}$ for each $x \in R$. In general, a k-dimensional random vector, X, is *measurable* $\mathcal{G}$ if $[\omega : X(\omega) \in H] \in \mathcal{G}$ for every measurable $H \in R^k$.

Definition The σ-field, $\sigma(X)$, *generated by* X is the smallest σ-field that X is measurable with respect to. If $X_1, X_2, ...$ is a finite or infinite sequence of random variables, then $\sigma(X_1, X_2, ...)$ is the smallest σ-field with respect to which each X_i is measurable.

Theorem 3.1 *Let $X = (X_1, ..., X_k)$ be a random vector. Then, $\sigma(X) = \sigma(X_1, ..., X_k)$ consists exactly of the sets $[\omega : X(\omega) \in H] = [\omega : (X_1(\omega), ..., X_n(\omega)) \in H]$ for all $H \in \mathcal{R}^k$. Y is measurable $\sigma(X)$ if and only if there is some $f : R^k \to R$ such that $Y(\omega) = f(X_1(\omega), ..., X_k(\omega))$.*

Definition Let $X : \Omega \to R$ be a random variable. The *distribution* (or law) of X is the probability measure on $\mathcal{R}$ given by PX^{-1}. That is, $\mu(A) = P(X \in A)$ for all $A \in \mathcal{R}$. The *support* for μ is any Borel set S such $\mu(S) = 1$. A random variable (and its distribution) are *discrete* if μ has countable support.

If X has distribution μ and $g : R \to R$, then $P(g(X) \in A) = P(X \in g^{-1}(A)) = \mu(g^{-1}A)$, and $g(X)$ has distribution μg^{-1}.

Theorem 3.2 *Let $T : \Omega \to \Omega'$ be measurable $\mathcal{F}/\mathcal{F}'$. Given a measure μ on $\mathcal{F}$, define μT^{-1} on $\mathcal{F}'$ by $\mu T^{-1}(A) = \mu(T^{-1}A)$. If f is non-negative, then $\int_\Omega f(T\omega)\mu(d\omega) = \int_{\Omega'} f(\omega')\mu T^{-1}(d\omega')$. A function f is integrable with respect to μT^{-1} if and only if fT is integrable with respect to μ. In that case, $\int_{T^{-1}A'} f(T\omega)\mu(d\omega) = \int_{A'} f(\omega')\mu T^{-1}(d\omega')$.*

Definition Let μ be a measure on $\mathcal{R}^1$ that assigns a finite measure to any bounded set. Define:

$$F(x) = \begin{cases} \mu(0, x] & \text{if } x \geq 0 \\ -\mu(x, 0] & \text{if } x \leq 0 \end{cases}$$

If μ is a probability measure, we call F a *(cumulative) distribution function* of the random variable X with distribution μ. If μ is finite, then we may define $F(x) = \mu(-\infty, x]$ instead; in this case, $F(x) = P(X \leq x)$.

Some facts about F as defined above:

- F is finite because μ is finite on bounded sets.
- F is non-decreasing.
- Right-continuous: If $x_n \downarrow x$ then $F(x_n) \downarrow F(x)$.
- $\mu(a, b] = F(b) - F(a)$
- If μ is the Lebesgue measure, then $F(x) = x$.

Theorem 3.3 *If F is a non-decreasing, right-continuous real function on R, there exists on $\mathcal{R}$ a unique measure, μ such that $\mu(a, b] = F(b) - F(a)$ for all $a, b \in R$.*

Definition The *jump* (or saltus) in a distribution function, F is given by $F(x) - \lim_{y \uparrow x} F(y) = F(x) - F(x-) = \mu(\{x\}) = P(X = x)$.

Definition Let δ be a non-negative measurable function. Define a measure by $\nu(A) = \int_A \delta d\mu$ for all $A \in \mathcal{F}$. Then we say that ν has *density* δ with respect to μ.

Definition A random variable, X, and its distribution, μ, have *density* f with respect to the Lebesgue measure if $f \geq 0$ is a Borel function on R and $P(x \in A) = \mu(A) = \int_A f(x)dx$ for all $a \in \mathcal{R}$.

Theorem 3.4 *Let X be a random variable. Let g be a non-negative function such that $\int_R g(x)dx = 1$ and $\mu_X(B) = \int_B g(x)dx$ for all measurable sets B. Then, for any function f, $\int f(x)d\mu_x = \int f(x)g(x)dx$, and g is the probability density function of the random variable X.*

Some properties of densities:

- If $\mu(A) = 0$ then $\nu(A) = 0$.
- ν is finite if and only if δ is integrable with respect to μ.
- If $\delta = \delta'$ almost everywhere, then δ and δ' induce the same density.
- If μ is σ-finite and δ and δ' induce the same density then $\delta = \delta'$ almost everywhere.

- $\int_R f(x)dx = P(X \in R) = 1$.
- f is determined only up to a set of Lebesgue measure 0.
- $F(b) - F(a) = \int_a^b f(x)dx$.
- Suppose f is a continuous density and $g : R \to R$ is increasing. Let $T = g^{-1}$. Then, $P(g(X) \leq x) = P(X \leq T(x)) = F(T(x))$. If T is differentiable, then $\frac{d}{dx}P(g(X) \leq x) = f(T(x))|T'(x)|$.

Theorem 3.5 *If ν has density δ with respect to μ and g is a non-negative function, then $\int g d\nu = \int g\delta d\mu$. An arbitrary g is integrable with respect to ν if and only if $g\delta$ is integrable with respect to μ. In that case, $\int_A g\, d\nu = \int_A g\delta\, d\mu$ for any A.*

Theorem 3.6 *Suppose $\nu_n(A) = \int_A \delta_n d\mu$ for each n and $\nu(A) = \int_A \delta d\mu$. If $\nu_n(\Omega) = \nu(\Omega) < \infty$ for all n and $\delta_n \to \delta$ except on a set of μ-measure 0, then*

$$\sup_{A \in \mathcal{F}} |\nu(A) - \nu_n(A)| \leq \sup_{A \in \mathcal{F}} \int_A |\delta - \delta_n| d\mu \to 0$$

Definition If $X = (X_1, ..., X_k)$ is a random vector, the (joint) distribution and distribution functions are given by $\mu(A) = P((X_1, ..., X_k) \in A)$ and $F(x_1, ..., x_k) = P(X_1 \leq x_1, ..., X_k \leq x_k) = \mu(S_x)$.

Definition Let $X = (X_1, ..., X_k)$ be a random vector. Let $g_j(X) = X_j$. Then, $\mu_j = \mu g_j^{-1}$ is defined by $\mu_j(A) = \mu((X_1, ..., X_k) : X_j \in A) = P(X_j \in A)$. $\{\mu_j\}$ are called the *marginal distributions* of μ. If f is the density of μ, then the marginal density is given by $f_j(y) = \int_{R^{k-1}} f(x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_k)dx_1...dx_{j-1}dx_{j+1}...dx_k$.

Some facts about joint distributions:

- F is non-decreasing in each variable and continuous from above (that is, $\lim_{h \downarrow 0} F(x_1 + h, ..., x_k + h) = F(x_1, ..., x_k)$).
- If any $x_i \to -\infty$ then $F(x_1, ..., x_k) \to 0$.
- If all $x_i \to \infty$, then $F(x_1, ..., x_k) \to 1$.
- F is continuous at $(x_1, ..., x_k)$ if and only if $\lim_{h \downarrow 0} F(x_1 - h, ..., x_k - h) = F(x_1, ..., x_k)$.
- F may have uncountably many discontinuities, but the set of points at which F is continuous are dense in R^k.
- If X is a random vector and $g : R^k \to R^i$ is a measurable function, then $g(X)$ is an i-dimensional random vector with distribution μg^{-1}. (Marginal distributions are a special case of this.)
- Suppose $g : V \to U$ is a one-to-one, onto and continuously differentiable map of open sets. Let $T = g^{-1}$. Suppose that the Jacobian of T, $J(x)$, does not vanish. If X has density f and support V, then for any $A \subset U$, $P(g(X) \in A) = P(X \in T(A)) = \int_{T(A)} f(y)dy = \int_A f(T(x))|J(x)|dx$, and the density of $g(X)$ is $f(T(x))|J(x)|I(x \in U)$.

4 Independence

Definition A finite collection of events, $A_1, ..., A_n$, is *independent* if $P(A_{k_1} \cap ... \cap A_{k_j}) = P(A_{k_1}) \cdot ... \cdot P(A_{k_j})$ for all subsets of distinct indices. An infinite collection of events is independent if each finite subcollection is independent.

Definition Let $\mathcal{A}_1, ..., \mathcal{A}_n \subset \mathcal{F}$ be classes of events. These classes are *independent* if, for any choices of $A_i \in \mathcal{A}_i$ for $i = 1, ..., n$, the events $A_1, ..., A_n$ are independent. That is, $P(A_1 \cap ... \cap A_n) = P(A_1) \cdots P(A_n)$ whenever $A_i \in \mathcal{A}_i \cup \{\Omega\}$.

Definition The infinite collection of classes, $[\mathcal{A}_\theta : \theta \in \Theta]$, is *independent* if the sets $A_\theta \in \mathcal{A}_\theta$ for each $\theta \in \Theta$ are always independent. Equivalently, the infinite collection is independent if each finite subcollection of $[\mathcal{A}_\theta : \theta \in \Theta]$ is independent.

Theorem 4.1 *If $\mathcal{A}_1, ..., \mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-system, then $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent. If $\mathcal{A}_\theta$ for $\theta \in \Theta$ are independent and each $\mathcal{A}_\theta$ is a π-system, then $\sigma(\mathcal{A}_\theta)$ for $\theta \in \Theta$ are independent.*

Definition A sequence $X_1, X_2, ...$ of random variables is *independent* if $\sigma(X_1), \sigma(X_2), ...$ are independent. Equivalently, we have $P[X_1 \in H_1, ..., X_n \in H_n] = P[X_1 \in H_1]...P[X_n \in H_n]$ for any set, $H_1, ..., H_n$. Since $[X_i = x]$ for all $x \in R$ (together with the empty set) form a π-system that generates $\sigma(X_i)$, $P[X_{k_1} = x_1, ..., X_{k_n} = x_n] = P[X_{k_1} = x_1] \cdot ... \cdot P[X_{k_n} = x_n]$, for any subset of distinct indices, is sufficient for independence as well.

If $(X_1, ..., X_k)$ has distribution μ, distribution function F, marginal distributions μ_i, and marginal distribution functions F_i, the following are equivalent:

- The random variables $X_1, ..., X_k$ are independent.
- $\mu = \mu_1 \cdot ... \cdot \mu_k$
- $F(x_1, ..., x_k) = F_1(x_1) \cdot ... \cdot F_k(x_k)$.

If these random variables have densities f and f_i, then $f(x) = f_1(x_1) \cdot ... \cdot f_k(x_k)$.

In addition, if $\mathcal{G}_1, ..., \mathcal{G}_k$ are independent σ-fields and each X_i is measurable $\mathcal{G}_i$, then $X_1, ..., X_k$ are independent as well.

Theorem 4.2 *Suppose an array, (A_{ij}) of events is independent. Let $\mathcal{F}_i$ be the σ-field generated by the i^{th} row. Then, $\mathcal{F}_1, \mathcal{F}_2, ...$ are independent.*

Theorem 4.3 *Suppose X_{ij} is an independent collection of random vectors. Let $\mathcal{F}_i$ be the σ-field generated by the i^{th} row. Then, $\mathcal{F}_1, \mathcal{F}_2, ...$ are independent.*

Theorem 4.4 *If X and Y are independent random variables with distributions μ and ν in R^j and R^k, then, for all $B \in \mathcal{R}^{j+k}$ and $A \in \mathcal{R}^j$,*

$$\begin{aligned} P((X,Y) \in B) &= \int_{R^j} P((x,Y) \in B)\mu(dx) \\ P(X \in A \text{ and } (X,Y) \in B) &= \int_A P((x,Y) \in B)\mu(dx) \end{aligned}$$

Theorem 4.5 *Let $\{\mu_n\}$ be a sequence of probability measures on the class of all subsets of the line, each having finite (discrete) support. There exists some probability space, $(\Omega, \mathcal{F}, P)$, with an independent sequence of random variables $\{X_n\}$ of simple random variables all on that space such that each X_n has distribution μ_n.*

Theorem 4.6 *If $\{\mu_n\}$ is a finite or infinite sequence of probability measures on $\mathcal{R}^1$, there exists on some probability space, $(\Omega, \mathcal{F}, P)$, an independent sequence of random variables, $\{X_n\}$ such that X_n has distribution μ_n.*

Theorem 4.7 *If the random variables $X_1, ..., X_n$ are independent, and $f_1, ..., f_n$ are measurable functions, then $f_1(X_1), ..., f_n(X_n)$ are also independent.*

Definition Let m, n be integers with $1 \leq m < n$. Let $\Pi_{mn}(H) = \{(x_1, ..., x_n) : (x_1, ..., x_m) \in H\}$. We call this the *projection map* of $\mathcal{R}^m$ into $\mathcal{R}^n$.

Theorem 4.8 *For each $n \geq 1$, let μ_n be a probability measure on $(R^n, \mathcal{R}^n)$ such that for all $m < n$, $\mu_n \circ \Pi_{mn} = \mu_m$. Then there exists a probability space, $(\Omega, \mathcal{F}, P)$and a sequence of random variables $\{X_j\}$ on it such that μ_n is the measure of $(X_1, ..., X_n)$ for all n.*

5 Expected Value

Definition A simple random variable, X, has an *expected value* (mean value) given by $E(X) = E(\sum_i x_i I(A_i)) = \sum_i x_i P(A_i)$. Equivalently, we may write $E(X) = \sum_{x \in R} xP(X = x)$.

Definition Let X be a random variable on (Ω, P). Then, the expected value of X is given by $E(X) = \int_\Omega X dP$.

Some facts about expected values:

- $E(X) = E(Y)$ if $P(X = Y) = 1$
- $E(I(A)) = P(A)$
- If $X(\omega) = \alpha$ for all $\omega \in \Omega$, then $E(X) = \alpha$.
- If α and β are constants, then $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$.
- If X and Y are independent, then $E(XY) = E(X)E(Y)$.
- If $X(\omega) \leq Y(\omega)$ on a set of probability one, then $E(X) \leq E(Y)$.

Theorem 5.1 *If $\{X_n\}$ is uniformly bounded and if $X = \lim_{n\to\infty} X_n$ on an $\mathcal{F}$-set of probability 1, then $E(X) = \lim_{n\to\infty} E(X_n)$.*

Corollary 5.2 *If $X = \sum_{n=1}^{\infty} X_n$ on an $\mathcal{F}$-set of probability one and the partial sums, $\sum_{n=1}^{k} X_n$, are uniformly bounded, then $E(X) = \sum_{n=1}^{\infty} E(X_n)$.*

Theorem 5.3 *Let X be a random variable on $(\Omega, \mathcal{F}, P)$ and f a Borel function. Let μ_X be the probability measure induced by X. Then, $E(f(X)) = \int_\Omega f(x)dP = \int_R f(x)d\mu_X$.*

Theorem 5.4 *For any random variable X, $\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n)$. If X is non-negative and takes only integer values, then $E(X) = \sum_{n=1}^{\infty} P(X \geq n)$.*

Theorem 5.5 (Generalization.) *Let X be a non-negative random variable. Let $f : R^+ \to R^+$ be measurable with $f(0) = 0$, such that f is absolutely continuous on $[0, t]$ for all $t < \infty$. Then, $E(f(X)) = \int_0^\infty f'(t)P(X \geq t)dt$. In particular, $E(X) = \int_0^\infty P(X \geq t)dt$.*

Definition The *variance* of a random variable, X is defined as $Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$.

6 Limit Sets and Convergence

Definition Let $A_1, A_2, ...$ be a sequence of sets. Then,

$$\begin{aligned} \limsup_n A_n &= \overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = [A_n \text{ infinitely often}] \\ \liminf_n A_n &= \underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = [A_n \text{ almost always}] \end{aligned}$$

That is, ω is in $\limsup_n A_n$ if and only if it lies in infinitely many of the A_n and in $\liminf_n A_n$ if and only if it lies in all but finitely many of the A_n.

Some facts:

- $\bigcap_{k=n}^{\infty} A_k \uparrow \liminf_n A_n$
- $\bigcup_{k=n}^{\infty} A_k \downarrow \limsup_n A_n$
- $\liminf A_n \subset \limsup A_n$ and the two are equal if and only if $\lim A_n$ exists (and then all three are equal).

Theorem 6.1 *For any $\{A_n\}$,*

$$P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n)$$

If $A_n \to A$ then $P(A_n) \to P(A)$.

Theorem 6.2 First Borel-Cantelli Lemma *If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(\limsup_n A_n) = 0$.*

Theorem 6.3 Second Borel-Cantelli Lemma *If $\{A_n\}$ is a sequence of independent events and $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(\limsup_n A_n) = 1$.*

Definition Let $A_1, A_2, ...$ be a sequence of events in a probability space $(\Omega, \mathcal{F}, P)$. The *tail σ-field* associated with $\{A_n\}$ is given by $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, ...)$. The elements of this σ-field are called *tail events*.

Theorem 6.4 Kolgomorov's Zero-One Law *If $A_1, A_2, ...$ are an independent sequence of events, then for each event, A, in the tail σ-field, $\mathcal{T}$, $P(A) = 0$ or $P(A) = 1$.*

Definition $\{X_n\}$ *converges* to X *with probability one* (*converges almost everywhere*, *converges almost surely* or *converges strongly*) if $P(\lim_n X_n(\omega) = X(\omega)) = 1$. Equivalently, $\{X_n\}$ converges almost everywhere to the random variable X if there exists a set A such that $P(A) = 0$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in A^C$.

Theorem 6.5 *$X_n \to X$ almost everywhere if and only if for all $\epsilon > 0$, $\lim_{m\to\infty} P(|X_n - X| \leq \epsilon$ for all $n \geq m) = 1$.*

Theorem 6.6 *$X_n \to 0$ almost surely if and only if for all $\epsilon > 0$, $P(|X_n| > \epsilon$ infinitely often$) = 0$. Equivalently, $\{X_m\}$ does not converge to X with probability one if $P(\bigcup_\epsilon [|X_n - X| \geq \epsilon i.o.]) > 0$.*

Definition Let $X_1, X_2, ...$ be a sequence of random variables on a probability space, $(\Omega, \mathcal{F}, P)$. They are *identically distributed* if their distributions (that is, $P(X \in A)$ for any set A) are all the same. We define $S_n = X_1 + ... + X_n$ and $\overline{X}_n = \frac{S_n}{n}$.

Theorem 6.7 Borel's Strong Law of Large Numbers. *Let $\{X_n\}$ be independent and identically distributed with $E(X_i) = 0$ and $E(X_i^4) < \infty$. Then $\overline{X}_n \to 0$ almost surely.*

Definition Two sequences, $\{X_n\}$ and $\{Y_n\}$, are *tail-equivalent* if $\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$.

Theorem 6.8 Kolgomorov's Three Series Criterion. *Let $\{X_n\}$ be a sequence of independent random variables. $S_n = \sum_{j=1}^{n} X_j$ converges almost everywhere if and only if, for some $a > 0$,*

1. $\sum_{n=1}^{\infty} P(|X_n| \geq a) < \infty$
2. $\sum_{n=1}^{\infty} E(X_n I(|X_n| \leq a)) < \infty$
3. $\sum_{n=1}^{\infty} Var(X_n I(|X_n| \leq a)) < \infty$

Theorem 6.9 *If $\{X_n\}$ and $\{Y_n\}$ are tail-equivalent then $\sum_{n=1}^{\infty}(X_n - Y_n)$ converges almost everywhere, and if $a_n \uparrow \infty$ then $\frac{1}{a_n}\sum_{j=1}^{n}(X_j - Y_j) \to 0$ almost everywhere.*

Lemma 6.10 Kronecker's Lemma. *Let $\{x_k\}$ be a sequence of real numbers and $\{a_n\}$ a sequence of real numbers with $a_k \uparrow \infty$. If $\sum_{j=1}^{\infty} \frac{x_j}{a_j} < \infty$ then $\frac{1}{a_n}\sum_{j=1}^{n} x_j \to 0$.*

Theorem 6.11 Kolgomorov's Strong Law of Large Numbers. *Let $\{X_n\}$ be independent and identically distributed, with $E(|X_i|) < \infty$ and $E(X_i) = 0$. Then, $\overline{X}_n \to 0$ almost surely.*

Theorem 6.12 Marankiewicz-Lygmund Strong Law of Large Numbers. *Let $\{X_j\}$ be independent and identically distributed with $E(X_1) = 0$ and $E(|X_1|^p) < \infty$ for some $1 < p < 2$. Let $S_n = \sum_{i=1}^n X_i$. Then, $\frac{S_n}{n^{1/p}} \to 0$ almost surely.*

Definition A sequence of random variables, $\{X_n\}$ *converges in probability* to X (that is, $X_n \to_p X$) if $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$ for all $\epsilon > 0$. This is also called *weak convergence*.

Theorem 6.13 *X_n converges to X in probability if and only if each subsequence, $\{X_{n_k}\}$ contains a further subsequence $\{X_{n_{k_i}}\}$ such that $X_{n_{k_i}} \to X$ with probability 1 as $i \to \infty$.*

Theorem 6.14 *If X_n converges almost everywhere to X, then X_n converges in probability to X.*

Definition For $0 < p < \infty$, X_n *converges in L^p* to X if $E(|X_n|^p) < \infty$, $E(|X|^p) < \infty$, and $\lim_{n\to\infty} E(|X_n - X|^p) = 0$.

Theorem 6.15 *If $X_n \to X$ in L^p then $X_n \to X$ in probability. If $X_n \to X$ in probability and there exists $Y \in L^p$ with $|X_n| \leq Y$ then $X_n \to X$ in L^p.*

Let $\Omega = [0, 1]$. Then we have examples of different kinds of convergence.

- In probability (and in L^p) but not almost surely: Write $n = k + 2^v$ where $0 \leq k \leq 2^v$ (this representation is unique). Define $X_n = 1$ if $\omega \in [k2^{-v}, (k+1)2^{-v})$ and $X_n = 0$ otherwise. This sequence converges to 0 in probability and in L^1, but not almost surely, since $X_n \neq 0$ infinitely often for all ω. (However, there is a subsequence that converges almost surely.)
- Almost surely but not in L^p: If the p^{th} moment does not exist, then a sequence cannot converge in L^p.

Theorem 6.16 *If $X_n \to X$ almost surely, then $E(|X|^r) \leq \liminf E(|X_n|^r)$. If $X_n \to X$ in L^r, then $E(|X_n|^r) \to E(|X|^r)$.*

Theorem 6.17 *If $X_n \to X$ in L^p for some $0 < p < \infty$, then $X_n \to X$ in L^q for all $0 < q < p$.*

Theorem 6.18 *Let $f : R \to R$ be continuous. If $X_n \to X$ almost surely then $f(X_n) \to f(X)$ almost surely. If $X_n \to X$ in probability, then $f(X_n) \to f(X)$ in probability.*

Corollary 6.19 *If $X_n \to X$ in probability and $Y_n \to Y$ in probability, then $X_n + Y_n \to X + Y$ in probability and $X_nY_n \to XY$ in probability.*

Theorem 6.20 *If $X_n \to X$ and $Y_n \to Y$ in L^p, then $X_n + Y_n \to X + Y$ in L^p.*

Definition If F_n and F are distribution functions, F_n *converges weakly* to F ($F_n \Rightarrow F$) if $\lim_{n\to\infty} F_n(x) = F(x)$ for each x at which F is continuous. The random variables, $\{X_n\}$, *converge in distribution* to X (that is, $X_n \to_D X$) if their distributions converge weakly to the distribution of X. Equivalently, $P(X_n \leq x) \to P(X \leq x)$ for all x such that $P(X = x) = 0$.

(The variable to which the X_n converge in distribution need not be defined on the same probability space.)

Definition Let Δ be defined by $\Delta(x) = I(x \geq 0)$. This is the distribution of the random variable $X(\omega) = 0$ for all $\omega \in \Omega$.

Definition The distribution functions, F and G, are of the same *type* if there exist constants a and b such that $F(ax + b) = G(x)$. A distribution function is *degenerate* if it has the form $\Delta(x - b)$ for some b. Otherwise, it is called *non-degenerate*.

Theorem 6.21 *Suppose $F_n(u_n x + v_n) \Rightarrow F(x)$ and $F_n(a_n x + b_n) \Rightarrow G(x)$ with $u_n, a_n > 0$ and F, G non-degenerate. Then, there exists a, b with $a > 0$ such that $\frac{a_n}{u_n} \to a$, $\frac{b_n - v_n}{u_n} \to b$, and $F(ax + b) = G(x)$.*

Theorem 6.22 *If $X_n \to_P X$ then $X_n \to_D X$.*

Theorem 6.23 *Let b be a constant. $X_n \to_P b$ if and only if $X_n \to_D b$.*

Theorem 6.24 *If $X_n \to_D X$ and $Y_n \to_D 0$ then $X_n + Y_n \to_D X$.*

Theorem 6.25 Skorohod's Device. *Suppose F_n and F are distribution functions on R with $F_n \to_D F$. Then there exist random variables Y_n and Y on a common probability space, $(\Omega, \mathcal{F}, P)$, such that each Y_n has distribution function F_n, Y has distribution F, and $Y_n \to Y$ for all $\omega \in \Omega$.*

Theorem 6.26 *Let $h : R \to R$ be measurable. Let $D_h = \{x : h$ is not continuous at $x\}$. Let $X_n \to_D X$ and $P(X \in D_h) = 0$. Then, $h(X_n) \to_D h(X)$.*

Corollary 6.27 *If $a_n \to a$, $b_n \to b$, and $X_n \to_D X$, then $a_n X_n + b_n \to_D aX + b$.*

Theorem 6.28 *$F_n \to_D F$ if and only if $\int f \, d\mu_{F_n} \to \int f \, d\mu_F$ for every bounded, continuous, real-valued f. Equivalently, $X_n \to_D X$ if and only if $E(f(X_n)) \to E(f(X))$ for all bounded, continuous, real-valued f.*

Theorem 6.29 Helly's Theorem. *For every sequence $\{F_n\}$ of distribution functions there exists a subsequence $\{F_{n_k}\}$ and a non-decreasing right-continuous function F such that $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ wherever F is continuous. Note that F may not be a distribution function.*

Definition A sequence of distribution functions, $\{F_n\}$, if *tight* if for every $\epsilon > 0$ there exists a finite interval, $[a, b]$ such that $F_n(b) - F_n(a) > 1 - \epsilon$ for all n.

Theorem 6.30 *Let $\{F_n\}$ be a sequence of distribution functions. $\{F_n\}$ is tight if and only if for every subsequence, $\{F_{n_k}\}$, there exists a further subsequence, $\{F_{n_{k_i}}\}$, and a distribution function F such that $F_{n_{k_i}} \to_D F$ as $i \to \infty$.*

Corollary 6.31 *If $\{F_n\}$ is tight and any subsequence converges to the same distribution function F, then $F_n \to_D F$.*

Theorem 6.32 *If $X_n \to_D X$ then $E(|X|) \leq \liminf_n E(|X_n|)$.*

Theorem 6.33 *If $X_n \to_D X$ and $|X_n|^r$ is uniformly integrable, then $E(X_n^r) \to E(X^r)$.*

Theorem 6.34 *Let $\{(X_n, Y_n)\}$ be a sequence of pairs of random variables. Let c be a constant. Then,*

- *If $X_n \to_D X$ and $Y_n \to_P c$ then $X_n \pm Y_n \to_D X_n \pm c$.*
- *If $X_n \to_D X$ and $Y_n \to_P c$ then $X_n Y_n \to_D cX$ if $c \neq 0$ and $X_n Y_n \to_P 0$ if $c = 0$.*
- *If $X_n \to_D X$ and $Y_n \to_P c$ then $\frac{X_n}{Y_n} \to_D \frac{X}{c}$ if $c \neq 0$.*

7 Characteristic Functions

Definition The *moment-generating function* for a random variable X is defined as $M_X(t) = E(e^{tX})$.

Theorem 7.1 *Since $E(e^{tX}) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$, $\frac{d^k}{dt^k} M_x(0) = E(X^k)$.*

Definition The *characteristic function* of a random variable, X, is defined for $t \in R$ as:

$$\phi_X(t) = E(e^{itX}) = \int_\Omega e^{itx} dP$$

Some facts about characteristic functions:

- $\phi_X(0) = 1$
- $\phi_X(t)$ exists and $|\phi_X(t)| \leq 1$ for all t.
- $\phi_X(t)$ is uniformly continuous, because $|\phi_X(t+h) - \phi_X(t)| \leq \int |e^{iht} - 1| d\mu$.
- Using a Taylor expansion, we find that $|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}| \leq \min\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\}$. This is useful for taking expectations and calculating moments.

Theorem 7.2 *If $E(|X|^k) < \infty$, then $\frac{d^k}{dt^k} \phi_X(0) = i^k E(X^k)$.*

Theorem 7.3 *If $X_1, ..., X_n$ are independent, then $\phi_{\sum X_i}(t) = \prod_{i=1}^n \phi_{X_i}(t)$.*

Theorem 7.4 Inversion and Uniqueness Theorem. *If a probability measure, μ, has a characteristic function, ϕ, and if $\mu(\{a\}) = \mu(\{b\}) = 0$, then*

$$\mu(a, b] = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{1}{it}(e^{-ita} - e^{itb})\phi(t)\, dt$$

(This shows that distinct measures must have distinct characteristic functions.) Furthermore, if $\int_{-\infty}^{\infty} |\phi(t)|dt < \infty$, then $F(x) = \mu(-\infty, x]$ has a derivative given by $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt$.

Theorem 7.5 Continuity Theorem. *Let μ_n, μ be probability measures with characteristic functions ϕ_n, ϕ. $\mu_n \to_D \mu$ if and only if $\phi_n(t) \to \phi(t)$ for all t.*

Corollary 7.6 *Suppose $\lim_n \phi_n(t) = g(t)$ for all $t \in R$, and g is continuous at 0. Then there exists a measure, μ such that $\mu_n \to_D \mu$ and g is the characteristic function of μ.*

Corollary 7.7 *Suppose $\lim_n \phi_n(t) = g(t)$ for all $t \in R$, and $\{\mu_n\}$ is tight. Then there exists a measure μ such that $\mu_n \to_D \mu$ and g is the characteristic function of μ.*

Definition A function, f, is *positive definite* if for all $z_1, ..., z_n \in C$ and $t_1, ..., t_n \in R$, $\sum_{i=1}^n \sum_{j=1}^n f(t_i - t_j) z_i \overline{z_j} \geq 0$.

Theorem 7.8 Bochner's Theorem. *If $\phi(t)$ is a function with*

1. *$\phi(0) = 1$,*
2. *$\phi(t)$ continuous at 0,*
3. *$\phi(t)$ positive definite,*

then ϕ is a characteristic function of some probability distribution μ.

Central Limit Theorems

Definition We say that $a_n = O(b_n)$ if $|\frac{a_n}{b_n}| < M$ for all n and some $M < \infty$. We say that $a_n = o(b_n)$ if $\frac{a_n}{b_n} \to 0$. We say that $f(t) = o(g(t))$ as $t \to 0$ if $\frac{f(t)}{g(t)} \to 0$ as $t \to 0$.

Lemma 7.9 *Let $\phi_X(t)$ be the characteristic function X, with $E(X^2) < \infty$. Then, as $t \to 0$,*

$$\phi_X(t) = 1 + itE(X) - \frac{1}{2}t^2E(X^2) + o(t^2)$$

Theorem 7.10 Central Limit Theorem. *Suppose $\{X_n\}$ is a sequence of independent and identically distributed random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$. Let $S_n = \sum_{i=1}^n X_i$. Then,*

$$\frac{1}{\sigma\sqrt{n}}(S_n - n\mu) \to_D Normal(0,1)$$

Theorem 7.11 Lindberg-Feller Theorem. *For each n, assume that $X_{n,1}, X_{n,2}, ..., X_{n,r_n}$ are independent and $r_n \to \infty$. Let $S_n = \sum_{i=1}^{r_n} X_{n,i}$. Suppose $E(X_{n,k}) = 0$ and $E(X_{n,k}^2) = \sigma_{n,k}^2$. Define $s_n^2 = \sum_{k=1}^{r_n} \sigma_{n,k}^2$. Assume that the* Lindberg condition *holds. That is, for all $\epsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{n,k}| \geq \epsilon s_n} X_{n,k}^2 dP = 0$$

Then, $\frac{S_n}{s_n} \to_D Normal(0,1)$.

Theorem 7.12 *Suppose $X_{n,1}, X_{n,2}, ..., X_{n,r_n}$ are independent with $E(X_{n,k}) = 0$. If these variables satisfy the* Lyapunov condition, *that is, if for any $\delta > 0$*

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E(|X_{n,k}|^{2+\delta}) = 0$$

then they satisfy the Lindberg condition as well.

Corollary 7.13 *Suppose $\{X_n\}$ is independent and identically distributed with $E(X_n) = \mu$ and $Var(X_n) = \sigma^2$. Then, $\sqrt{n}\frac{\overline{X}-\mu}{s} \to_D Normal(0,1)$.*

Theorem 7.14 Multivariate Central Limit Theorem. *Suppose X_n is a k-dimensional random vector such that for any vector of constants, $(a_1, ..., a_k)^T$, $\sum_{i=1}^k a_i X_{ni} \to_D Normal(0, a^T \Sigma a)$. Then, $X_n \to_D Normal(0, \Sigma)$.*

Theorem 7.15 Multivariate Central Limit Theorem. *Let $X_n = (X_{n1}, ..., X_{nk})$ be independent random vectors all having the same distribution. Suppose that $E(X_{1u}^2) < \infty$. Let $c = E(X_1)$ and $\Sigma = E((X_1 - c)'(X_1 - c))$. Let $S_n = X_1 + ... + X_n$. Then, the distribution of the random vector $\frac{1}{\sqrt{n}}(S_n - nc)$ converges weakly to the multivariate normal distribution with mean zero and covariance matrix Σ.*

8 Conditional Expectation

Definition If $P(A) > 0$ then the *conditional probability* of B given A is $P(B|A) = \frac{P(A\cap B)}{P(A)}$.

Definition Suppose X is a random variable on $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ is a σ-field. Then there exists a random variable, $E(X\|\mathcal{G})$, called the *conditional expected value* of X given $\mathcal{G}$, such that:

- $E(X\|\mathcal{G})$ is measurable $\mathcal{G}$ and integrable
- $\int_G E(X\|\mathcal{G})dP = \int_G X\,dP$ for all $G \in \mathcal{G}$.

Since this definition is unique except for a set of probability zero, we called each a *version* of $E(X\|\mathcal{G})$.

Note that $E(X\|\{0,\Omega\}) = E(X)$ and $E(X\|\mathcal{F}) = X$, and conditional probabilities can be defined by $P(A\|\mathcal{G}) = E(I(A)\|\mathcal{G})$.

Definition If $\{X_t\}_{t\in T}$ is a collection of random variables on $(\Omega, \mathcal{F}, P)$, we define $E(X\|X_t, t \in T) = E(X\|\sigma(X_t, t \in T))$.

Theorem 8.1 *Let $\mathcal{P}$ be a π-system generating the σ-field $\mathcal{G}$. Suppose Ω is a finite or countable union of sets in $\mathcal{G}$. An integrable function, f, is a version of $E(X\|\mathcal{G})$ if it is measurable $\mathcal{G}$ and if $\int_G f\,dP = \int_G X\,dP$ for all $G \in \mathcal{G}$.*

Theorem 8.2 *Suppose X, Y, and X_n are integrable. Then, with probability one:*

- *If $X = a$ with probability one, then $E(X\|\mathcal{G}) = a$.*
- *If $a, b \in R$, $E(aX + bY\|\mathcal{G}) = aE(X\|\mathcal{G}) + bE(Y\|\mathcal{G})$.*
- *If $X \le Y$ with probability one, then $E(X\|\mathcal{G}) \le E(Y\|\mathcal{G})$.*
- $|E(X\|\mathcal{G})| \le E(|X|\|\mathcal{G})$
- *If $\lim_n X_n = X$ with probability one and $|X_n| \le Y$ (with Y integrable), then $\lim_n E(X_n\|\mathcal{G}) = E(X\|\mathcal{G})$ with probability one.*

Theorem 8.3 *If X is measurable $\mathcal{G}$ and if Y and XY are integrable, then $E(XY\|\mathcal{G}) = XE(Y\|\mathcal{G})$ with probability one.*

Theorem 8.4 *If X is integrable and $\mathcal{G}_1 \subset \mathcal{G}_2$ are σ-fields, then $E(E(X\|\mathcal{G}_2)\|\mathcal{G}_1) = E(X\|\mathcal{G}_1) = E(E(X\|\mathcal{G}_1)\|\mathcal{G}_2)$. (This is a generalization of the law of iterated expectations.)*

Theorem 8.5 Jensen's Inequality. *If ϕ is a convex function on the line and if both X and $\phi(X)$ are integrable, then $\phi(E(X\|\mathcal{G})) \le E(\phi(X)\|\mathcal{G})$ with probability one.*

Definition Let X be a random variable on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a σ-field in $\mathcal{F}$. Then there exists a function, $\mu(H, \omega)$ for $H \in \mathcal{R}$, $\omega \in \Omega$, such that:

- For each $\omega \in \Omega$, $\mu(\cdot, \omega)$ is a probability measure on $\mathcal{R}$.
- For each $H \in \mathcal{R}$, $\mu(H, \cdot)$ is a version of $P(X \in H\|\mathcal{G})$.

Such a function is called the *conditional distribution* of X given $\mathcal{G}$.

Theorem 8.6 *Let $\mu(\cdot, \omega)$ be a conditional distribution with respect to $\mathcal{G}$ of a random variable X. If $\phi : R \to R$ is a Borel function and $\phi(X)$ is integrable, then $\int_R \phi(x)\mu(dx, \omega)$ is a version of $E(\phi(X)\|\mathcal{G})$.*

Martingales

Definition Let $X_1, X_2, ...$ be a sequence of random variables on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_1, \mathcal{F}_2, ...$ be a sequence of σ-fields in $\mathcal{F}$. The sequence $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ is a *martingale* relative to the σ-fields $\{\mathcal{F}_n\}$ if:

- $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ (in which case, we say that $\{\mathcal{F}_n\}$ is a *filtration*),
- X_n is measurable $\mathcal{F}_n$ (in which case, we say that $\{X_n\}$ is *adapted to* the filtration),
- $E(|X_n|) < \infty$, and
- with probability one, $E(X_{n+1} \| \mathcal{F}_n) = X_n$.

Note that the smallest filtration to which a sequence $\{X_n\}$ is adapted is $\mathcal{F}_n = \sigma(X_1, ..., X_n)$.

Definition Let $\{X_n\}$ be a martingale. Define $\Delta_n = X_n - X_{n-1}$. Then we say that $\{\Delta_n\}$ is a *martingale difference*.

9 Miscellaneous

9.1 Convolution

Definition Let X and Y be independent random variables with distributions μ and ν. The *convolution* of μ and ν is $(\mu * \nu)(H) = \int_{-\infty}^{\infty} \nu(H - x)\mu(dx)$, for $H \in \mathcal{R}$.

Some facts about convolution:

- Convolution is commutative and associative.
- If X and Y are independent with distributions μ and ν, then $P(X + Y \in H) = (\mu * \nu)(H)$.

Definition If F and G are the distributions functions corresponding to μ and ν, then the distribution function corresponding to $\mu * \nu$ is $(F * G)(y) = \int_{-\infty}^{\infty} G(y - x)dF(x)$. If F and G have densities f and g, then $(F * g)(y) = \int_{-\infty}^{\infty} g(y - x)dF(x)$ and $(f * g)(y) = \int_{-\infty}^{\infty} g(y - x)f(x)dx$.

9.2 Empirical CDF's

Definition Let $\{X_n\}$ be independent and identically distributed with a CDF F. The *empirical CDF* based on a sample $X_1, ..., X_n$ is given by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x)$$

F_n is an estimator of $F(x) = P(X_1 \leq x)$.

Since $E(F(X_1 \leq x) - F(x)) = 0$ and $E(|I(X_1 \leq x)|) < \infty$, we may apply a law of large numbers to find that $F_n(x) \to F(x)$ almost surely for each x.

Theorem 9.1 Glivenko-Cantelli Theorem. $\sup_{x \in R} |F_n(x) - F(x)| \to 0$.

10 Inequalities

Theorem 10.1 *If $X \geq 0$, for any $\alpha > 0$, $P(X \geq \alpha) \leq \frac{1}{\alpha} E(X)$.*

Theorem 10.2 Markov's Inequality. $P(|X| \geq \alpha) \leq \frac{1}{\alpha^k} E(|X|^k)$.

Theorem 10.3 Chebyshev's Inequality. $P(|X - E(X)| \geq \alpha) \leq \frac{1}{\alpha^2} Var(X)$.

Theorem 10.4 Chebyshev's Inequality (Generalized). *Let f be a strictly positive and increasing function on $(0, \infty)$ with $f(u) = f(-u)$. Let X be a random variable with $E(f(X)) < \infty$. Then, for every $u > 0$, $P(|X| \geq u) \leq E(f(X))/f(u)$.*

Definition A function $f : R \to R$ is *convex* if for every $\lambda_1, ..., \lambda_n \geq 0$ with $\sum_{i=1}^n \lambda_i = 1$, and every $x_1, ..., x_n \in R$, $f(\sum_{i=1}^n \lambda_i x_i) \geq \sum_{i=1}^n \lambda_i f(x_i)$.

Theorem 10.5 Jensen's Inequality. *If $\phi : R \to R$ is convex on the range of X, then $\phi(E(X)) \leq E(\phi(X))$.*

Theorem 10.6 Holder's Inequality. *Let $(\Omega, \mathcal{F}, p)$ be a probability space and X, Y random variables on Ω. If $\frac{1}{p} + \frac{1}{q} = 1$ with $p, q > 1$, then $E(|XY|) \leq E(|X|^p)^{1/p} E(|Y|^q)^{1/q}$.*

Theorem 10.7 (Cauchy-)Schwarz Inequality. $E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$.

Theorem 10.8 Lyapounov's Inequality *If $0 < \alpha \leq \beta$, then $E(|X|^\alpha)^{1/\alpha} \leq E(|X|^\beta)^{1/\beta}$.*