Chapter 1

Probability Theory:

Introduction

Definitions – Algebra

Definitions: Semiring

A collection of sets $F$ is called a *semiring* if it satisfies:

- $\emptyset \in F$.
- If $A, B \in F$, then $A \cap B \in F$.
- If $A, B \in F$, then there exists a collection of sets $C_1, C_2, ..., C_n \in F$, such that $A \setminus B = \bigcup_{i \geq 1} C_i$. ($A \setminus B$ all elements of $A$ not in $B$)

Definitions: Algebra

A collection of sets $F$ is called an *algebra* if it satisfies:

- $\emptyset \in F$.
- If $\omega_1 \in F$, then $\omega_1^C \in F$. ($F$ is closed under *complementation*)
- If $\omega_1 \in F$ & $\omega_2 \in F$, then $\omega_1 \cup \omega_2 \in F$. ($F$ is closed under *finite unions*).
Definitions: Sigma-algebra

Definition: Sigma-algebra
A sigma-algebra (σ-algebra or σ-field) \( F \) is a set of subsets \( \omega \) of \( \Omega \) s.t.:

• \( \emptyset \in F \).
• If \( \omega \in F \), then \( \omega^C \in F \). (\( \omega^C \) = complement of \( \omega \))
• If \( \omega_1, \omega_2, \ldots, \omega_n, \ldots \in F \), then \( U_{i = 1}^\infty \omega_i \in F \).

Note that the set \( E = \{ \emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{b,c\}, \{a,c\}, \{a,b,c\} \} \) is an algebra and a σ-algebra.

Sigma algebras are a subset of algebras in the sense that all σ-algebras are algebras, but not vice versa. Algebras only require that they be closed under pairwise unions while σ-algebras must be closed under countably infinite unions.

Sigma-algebra

Theorem:
All σ-algebras are algebras, and all algebras are semi-rings.

Thus, if we require a set to be a semiring, it is sufficient to show instead that it is a σ-algebra or algebra

• Sigma algebras can be generated from arbitrary sets. This will be useful in developing the probability space.

Theorem:
For some set \( X \), the intersection of all σ-algebras, \( \Lambda_i \), containing \( X \)-that is, \( x \in X \Rightarrow x \in \Lambda_i \) for all \( i \)- is itself a σ-algebra, denoted \( \sigma(X) \).

=>This is called the σ-algebra generated by \( X \).
Sample Space, \( \Omega \)

Definition: Sample Space

The *sample space* \( \Omega \) is the set of all possible unique outcomes of the experiment at hand.

Example: If we roll a die, \( \Omega = \{1; 2; 3; 4; 5; 6\} \).

In the probability space, the \( \sigma \)-algebra we use is \( \sigma (\Omega) \), the \( \sigma \)-algebra generated by \( \Omega \). Thus, take the elements of \( \Omega \) and generate the "extended set" consisting of all unions, compliments, compliments of unions, unions of compliments, etc. Include \( \Phi \); with this "extended set" and the result is \( \sigma (\Omega) \), which we denote as \( \Sigma \).

Definition: The \( \sigma \)-algebra generated by \( \Omega \), denoted \( \Sigma \), is the collection of possible events from the experiment at hand.

\((\Omega, \Sigma)\)

We have an experiment with \( \Omega = \{1, 2\} \). Then, \( \Sigma = \{\Phi, \{1\}, \{2\}, \{1, 2\}\} \). Each of the elements of \( \Sigma \) is an event. Think of events as descriptions of experiment outcomes (\( \Phi \): the “nothing occurs” event).

Note that \( \sigma \)-algebras can be defined over the real line as well as over abstract sets. To develop this notion, we need the concept of a topology.

Definition:

A *topology* \( \tau \) on a set \( X \) is a collection of subsets of \( X \) satisfying:

1. \( \Phi, X \in \tau \)
2. \( \tau \) is closed under finite intersections.
3. \( \tau \) is closed under arbitrary unions.

Any element of a topology is known as an open set.
**Borel σ-algebra**

**Definition**: Borel σ-algebra (Emile Borel (1871-1956), France.)

The *Borel σ-algebra* (or, Borel field,) denoted $B$, of the topological space $(X; \tau)$ is the σ-algebra generated by the family $\tau$ of open sets. Its elements are called Borel sets.

**Lemma**: Let $C = \{(a, b); a < b\}$. Then $\sigma(C) = B_\mathbb{R}$ is the Borel field generated by the family of all open intervals $C$.

What do elements of $B_\mathbb{R}$ look like? Take all possible open intervals. Take their compliments. Take arbitrary unions. Include $\Phi$ and $\mathbb{R}$. $B_\mathbb{R}$ contains a wide range of intervals including open, closed, and half-open intervals. It also contains disjoint intervals such as $\{(2; 7) \cup (19; 32)\}$. It contains (nearly) every possible collection of intervals that are imagined.

---

**Measures**

**Definition**: Measurable Space

A pair $(X, \Sigma)$ is a *measurable space* if $X$ is a set and $\Sigma$ is a nonempty σ-algebra of subsets of $X$.

A measurable space allows us to define a function that assigns real-numbered values to the abstract elements of $\Sigma$.

**Definition**: Measure $\mu$

Let $(X, \Sigma)$ be a measurable space. A set function $\mu$ defined on $\Sigma$ is called a *measure* iff it has the following properties.

1. $0 \leq \mu(A) \leq \infty$ for any $A \in \Sigma$.
2. $\mu(\Phi) = 0$.
3. ($\sigma$-additivity). For any sequence of pairwise disjoint sets $\{A_n\} \in \Sigma$ such that $\bigcup_{n=1}^{\infty} A_n \in \Sigma$, we have $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.
Measures

Intuition: A measure on a set, $S$, is a systematic way to assign a number to each suitable subset of that set, intuitively interpreted as its size. In this sense, it generalizes the concepts of length, area, volume.

Examples of measures:
- Counting measure: $\mu(S) =$ number of elements in $S$.
- Lebesgue measure on $\mathbb{R}$: $\mu(S) =$ conventional length of $S$.
That is, if $S = [a,b] \Rightarrow \mu(S) = \lambda[a,b] = b-a$.

Measures & Measure Space

- Note: A measure $\mu$ may take $\infty$ as its value. Rules:
  (1) For any $x \in \mathbb{R}$, $\infty + x = \infty$, $x \times \infty = \infty$ if $x > 0$, $x \times \infty = -\infty$ if $x < 0$, and $0 \times \infty = 0$;
  (2) $\infty + \infty = \infty$;
  (3) $\infty \times a = \infty$ for any $a > 0$;
  (4) $\infty - \infty$ or $\infty/\infty$ are not defined.

Definition: Measure Space
A triplet $(X,\Sigma,\mu)$ is a measure space if $(X,\Sigma)$ is a measurable space and $\mu: \Sigma \rightarrow [0; \infty)$ is a measure.

- If $\mu(X) = 1$, then $\mu$ is a probability measure, which we usually use notation $P$, and the measure space is a probability space.
Lebesgue Measure

- **Lebesgue measure.** There is a unique measure $\lambda$ on $(\mathbb{R}, B_\mathbb{R})$ that satisfies
  $$\lambda([a, b]) = b - a$$
  for every finite interval $[a, b]$, $-\infty < a \leq b < \infty$. This is called the Lebesgue measure.

If we restrict $\lambda$ to the measurable space $([0, 1], B_{[0,1]})$, then $\lambda$ is a probability measure.

**Examples:**
- Any Cartesian product of the intervals $[a,b] \times [c,d]$ is Lebesgue measurable, and its Lebesgue measure is $\lambda = (b-a)(d-c)$.
- $\lambda$([set of rational numbers in an interval of $\mathbb{R}$]) = 0.
**Note:** Not all sets are Lebesgue measurable. See Vitali sets.

Zero Measure

**Definition:** Measure Zero

A ($\mu$-)measurable set $E$ is said to have ($\mu$-)measure zero if $\mu(E) = 0$.

**Examples:** The singleton points in $\mathbb{R}^n$, and lines and curves in $\mathbb{R}^n$, $n \geq 2$. By countable additivity, any countable set in $\mathbb{R}^n$ has measure zero.

- A particular property is said to hold *almost everywhere* if the set of points for which the property fails to hold is a set of measure zero.
**Example:** “a function vanishes almost everywhere”; “$f = g$ almost everywhere”.

- **Note:** Integrating anything over a set of measure zero produces zero. Changing a function on a set of measure zero does not affect the value of its integral.
Measure: Properties

- Let \((\Omega, \Sigma, \mu)\) be a measure space. Then, \(\mu\) has the following properties:
  
  (i) (Monotonicity). If \(A \subset B\), then \(\mu(A) \leq \mu(B)\).

  (ii) (Subadditivity). For any sequence \(A_1, A_2, \ldots\),
  \[\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)\]

  (iii) (Continuity). If \(A_1 \subset A_2 \subset A_3 \subset \ldots\) (or \(A_1 \supset A_2 \supset A_3 \supset \ldots\) and \(\mu(A_1) < \infty\)), then
  \[\mu\left(\lim_{n \to \infty} A_n\right) = \lim_{n \to \infty} \mu(A_n),\]
  where
  \[\lim_{n \to \infty} A_n = \bigcup_{i=1}^{\infty} A_i\] (or \(\bigcap_{i=1}^{\infty} A_i\))

Probability Space

A measure space \((\Omega, \Sigma, \mu)\) is called finite if \(\mu(\Omega)\) is a finite real number (not \(\infty\)). A measure \(\mu\) is called \(\sigma\)-finite if \(\Omega\) can be decomposed into a countable union of measurable sets of finite measure.

For example, the real numbers with the Lebesgue measure are \(\sigma\)-finite but not finite.

**Definition:** Probability Space

A measure space is a probability space if \(\mu(\Omega)=1\). In this case, \(\mu\) is a probability measure, which we denote \(P\).

- Let \(P\) be a probability measure. The cumulative distribution function (c.d.f.) of \(P\) is defined as:
  \[F(x) = P\left(\left(-\infty, x\right]\right), x \in \mathbb{R}\]
**Product Space**

**Definition:** Product space

Let $\Gamma_i, i \in I$, be sets, where $I = \{1, ..., k\}$, $k$ is finite or $\infty$.

Define the product space as

$$\prod_{i \in I} \Gamma_i = \Gamma_1 \times \cdots \times \Gamma_k = \{(a_1, ..., a_k) : a_i \in \Gamma_i, i \in I\}$$

**Example:** $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$

- Let $(\Omega_i, \Sigma_i), i \in I$, be measurable spaces
  - $\Pi_i \in I \Sigma_i$ is not necessarily a $\sigma$-field
  - $\sigma(\Pi_i \in I \Sigma_i)$ is called the *product $\sigma$-field* on the product space $\Pi_i \in I \Omega_i$
  - $(\Pi_i \in I \Omega_i, \sigma(\Pi_i \in I \Sigma_i))$ is denoted by $\Pi_i \in I (\Omega_i, \Sigma_i)$

**Example:** $\Pi_{i=1,...,k}(\mathbb{R}, B) = (\mathbb{R}^k, B^k)$.

---

**Product Measure**

**Definition:** Product measure

Consider a rectangle $[a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$. The product measure of the usual area $[a_1, b_1] \times [a_2, b_2]$:

$$(b_1 - a_1)(b_2 - a_2) = \mu([a_1, b_1]) \mu([a_2, b_2])$$

Q: Is $\mu([a_1, b_1]) \mu([a_2, b_2])$ the same as the value of a measure defined on the product $\sigma$-field?

- A measure $\mu$ on $(\Omega, \Sigma)$ is said to be $\sigma$-finite if and only if there exists a sequence $A_1, A_2, ..., \text{ such that } \bigcup A_i = \Omega \text{ and } \mu(A_i) < \infty \text{ for all } i$

- Any finite measure (such as a probability measure) is clearly $\sigma$-finite.

The Lebesgue measure on $\mathbb{R}$ is $\sigma$-finite, since $\mathbb{R} = \bigcup A_n$ with $A_n = (-n, n)$, $n = 1, 2, ...$$

- The counting measure in is $\sigma$-finite if and only if $\Omega$ is countable.
Product Measure & Joint CDF

**Theorem:** Product measure theorem

Let \((\Omega_i, \Sigma_i, \mu_i), i = 1, ..., k\), be measure spaces with \(\sigma\)-finite measures, where \(k \geq 2\) is an integer. Then there exists a unique \(\sigma\)-finite measure on the product \(\Sigma = \sigma(A_1 \times ... \times A_k)\), called the **product measure** and denoted by

\[
\mu_1 \times ... \times \mu_k(A_1 \times ... \times A_k) = \mu(A_1) \ldots \mu(A_k)
\]

for all \(A_i \in \Sigma_i, i = 1, ..., k\).

- Let \(P\) be a probability measure on \((\mathbb{R}^k, B^k)\). The c.d.f. (or joint c.d.f.) of \(P\) is defined by

\[
F(x_1, ..., x_k) = P((-\infty, x_1] \times ... \times (-\infty, x_k]), \quad x_i \in \mathbb{R}
\]

- There is a one-to-one correspondence between probability measures and joint c.d.f.'s on \(\mathbb{R}^k\).

Product Measure & Marginal CDF

- If \(F(x_1, ..., x_k)\) is a joint c.d.f., then

\[
F_j(x) = \lim_{x_i \to \infty, i=1,...,j-1,j+1,...,k} F(x_1, ..., x_k)
\]

is a c.d.f and is called the \(j^{th}\) marginal c.d.f.

- Marginal c.d.f's are determined by their joint c.d.f. But, a joint c.d.f cannot be determined by \(k\) marginal c.d.f's.

- If \(F(x_1, ..., x_k) = F(x_j) \ldots F(x_k)\), then, the probability measure corresponding to \(F\) is the product measure \(P_1 \times ... \times P_k\) with \(P_i\) being the probability measure corresponding to \(F_i\).
**Measurable Function**

**Definition:** Inverse function

Let $f$ be a function from $\Omega$ to $\Lambda$ (often $\Lambda = \mathbb{R}^k$)
Let the inverse image of $B \subseteq \Lambda$ under $f$

$$f^{-1}(B) = \{f \in B\} = \{\omega \in \Omega : f(\omega) \in B\}.$$

- **Useful properties:**
  - $f^{-1}(B^c) = (f^{-1}(B))^c$ for any $B \subseteq \Lambda$;
  - $f^{-1}(\bigcup B_i) = \bigcup f^{-1}(B_i)$ for any $B_i \subseteq \Lambda, i = 1, 2, ...$

**Note:** The inverse function $f^{-1}$ need not exist for $f^{-1}(B)$ to be defined.

**Definition:** Let $(\Omega, \Sigma)$ and $(\Lambda, G)$ be measurable spaces and $f$ a function from $\Omega$ to $\Lambda$. The function $f$ is called a measurable function from $(\Omega, \Sigma)$ to $(\Lambda, G)$ if and only if $f^{-1}(G) \subseteq \Sigma$.

**Measurable Function**

- If $f$ is measurable from $(\Omega, \Sigma)$ to $(\Lambda, G)$ then $f^{-1}(G)$ is a sub-$\sigma$-field of $\Sigma$. It is called the $\sigma$-field generated by $f$ and is denoted by $\sigma(f)$.

- If $f$ is measurable from $(\Omega, \Sigma)$ to $(\mathbb{R}, B)$, it is called a Borel function or a random variable (RV).

- A random variable is a convenient way to express the elements of $\Omega$ as numbers rather than abstract elements of sets.
RS – Chapter 1 – Random Variables

Measurable Function

Example: Indicator function for $A \subset \Omega$.

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c \end{cases}$$

For any $B \subset R$

$$I_A^{-1}(B) = \begin{cases} \emptyset & \text{if } 0 \text{ not in } B, 1 \text{ not in } B \\ A & \text{if } 0 \text{ not in } B, 1 \in B \\ A^c & \text{if } 0 \in B, 1 \text{ not in } B \\ \Omega & \text{if } 0 \in B, 1 \in B \end{cases}$$

Then, $\sigma(I_A) = \{\emptyset, A, A^c, \Omega\}$ and $I_A$ is Borel if and only if $A \in \Sigma$.

$\sigma(f)$ is much simpler than $\Sigma$.

• Note: We express the elements of $\Omega$ as numbers rather than abstract elements of sets.

Measurable Function - Properties

Theorems: Let $(\Omega, \Sigma)$ be a measurable space.

(i) $f$ is a RV if and only if $f^{-1}(a, \infty) \in \Sigma$ for all $a \in R$.

(ii) If $f$ and $g$ are RVs, then so are $fg$ and $af + bg$, where $a$ and $b \in R$; also, $f/g$ is a RV provided $g(\omega) \neq 0$ for any $\omega \in \Omega$.

(iii) If $f_1, f_2, \ldots$ are RVs, then so are $\sup_n f_n, \inf_n f_n, \limsup_n f_n,$ and $\liminf_n f_n$. Furthermore, the set $A = \{\omega \in \Omega: \lim_{n \to \infty} f_n(\omega) \text{ exists}\}$ is an event and the function

$$b(\omega) = \begin{cases} \lim_{n \to \infty} f_n(\omega) & \omega \in A \\ f_1(\omega) & \omega \in A^c \end{cases}$$

is a RV.
Measurable Function - Properties

Theorems: Let $(\Omega, \Sigma)$ be a measurable space.

(iv) (Closed under composition) Suppose that $f$ is measurable from $(\Omega, \Sigma)$ to $(\Lambda, G)$ and $g$ is measurable from $(\Lambda, G)$ to $(\Delta, H)$. Then, the composite function $g \circ f$ is measurable from $(\Omega, \Sigma)$ to $(\Delta, H)$.

(v) Let $\Omega$ be a Borel set in $\mathbb{R}^p$. If $f$ is a continuous function from $\Omega$ to $\mathbb{R}^p$, then $f$ is measurable.

Distribution

Definition
Let $(\Omega, \Sigma, \mu)$ be a measure space and $f$ be a measurable function from $(\Omega, \Sigma)$ to $(\Lambda, G)$. The induced measure by $f$, denoted by $\mu \circ f^{-1}$, is a measure on $G$ defined as

$$\mu \circ f^{-1}(B) = \mu(f^{-1}(B)), \quad B \in G$$

If $\mu = P$ is a probability measure and $X$ is a random variable or a random vector, then $P \circ X^{-1}$ is called the distribution (or the law) of $X$ and is denoted by $P_X$.

• The cdf of $P_X$ is also called the cdf (or joint cdf) of $X$ and is denoted by $F_X$. 

**Probability Space – Definition and Axioms**

**Kolmogorov's axioms**

Kolmogorov defined a list of axioms for a probability measure. Let \( P: E \rightarrow [0; 1] \) be our probability measure and \( E \) be some \( \sigma \)-algebra (events) generated by \( X \).

- **Axiom 1**: \( P[A] \leq 1 \) for all \( A \in E \)
- **Axiom 2**: \( P[X] = 1 \)
- **Axiom 3**: \( P[A_1 U A_2 U ... U A_n] = P[A_1] + P[A_2] + ... + P[A_n], \) where \( \{A_1; A_2; ...; A_n\} \) are disjoint sets in \( E \).

**Probability Space – Properties of P**

The three Kolmogorov’s basic axioms imply the following results:

- Theorem: \( P[\emptyset] = 0 \).
- Theorem: \( P[A] \in [0,1] \).
- Theorem: \( P[B \cap A] = P[B] - P[A \cap B] \).
- Theorem: \( A \) is in \( B \) \( \Rightarrow \) \( P[A] \leq P[B] \).
- Theorem: \( A = B \) \( \Rightarrow \) \( P[A] = P[B] \).
- Theorem: \( P[A] = \sum_{i=1} P[A \cap C_i] \) where \( \{C_1; C_2; ...\} \) forms a partition of \( E \).
- Theorem (Boole's Inequality, aka "Countable Subadditivity"):
  \[ P[U_{i=1} A_i] \leq \sum_{i=1} P[A_i] \] for any set of sets \( \{A_1; A_2; ...\} \).
Probability Space - \((\Omega, \Sigma, P)\)

Now, we have all the tools required to establish that \((\Omega, \Sigma, P)\) is a probability space.

**Theorem:**
Let \(\Omega\) be the sample space of outcomes of an experiment, \(\Sigma\) be the \(\sigma\)-algebra of events generated from \(\Omega\), and \(P:\Sigma \to [0, \infty)\) be a probability measure that assigns a nonnegative real number to each event in \(\Sigma\). The space \((\Omega, \Sigma, P)\) satisfies the definition of a probability space.

**Remark:** The sample space is the list of all possible outcomes. Events are groupings of these outcomes. The \(\sigma\)-algebra \(\Sigma\) is the collection of all possible events. To each of these possible events, we assign some "size" using the probability measure \(P\).

**Probability Space**

**Example:** Consider the tossing of two fair coins. The sample space is \{HH; HT; TH; TT\}.
Possible events:
- The coins have different sides showing: \{HT; TH\}.
- At least one head: \{HH; HT; TH\}.
- First coin shows heads or second coin shows tails: \{HH, HT, TT\}

The sigma algebra generated from the sample space is the collection of all possible such events: \(\emptyset, \{H\}, \{H,T\}, \{T\}, \{H,T\}, \{HH,HT\}, \{HH,TH\}, \{HT,TH\}, \{HT,TT\}, \{HH,HT,TH\}, \{HH,HT,TT\}, \{HH,TH,TT\}, \{HT,TH,TT\}, \{HH,HT,TH,TT\}\}

The probability measure \(P\) assigns a number from 0 to 1 to each of those events in the sigma algebra.
Probability Space

Example (continuation):
The probability measure $P$ assigns a number from 0 to 1 to each of those events in the sigma algebra. With a fair coin, we can assign the following probabilities to the events in the above sigma algebra:
\{0; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{2}; \frac{1}{2}; \frac{1}{2}; \frac{1}{2}; \frac{1}{2}; \frac{1}{2}; \frac{3}{4}; \frac{3}{4}; \frac{3}{4}; \frac{3}{4}; 1\}

But, we do not have to use these fair values. We may believe the coin is biased so that heads appears $\frac{3}{4}$ of the time. Then the following values for $P$ would be appropriate:
\{0; \frac{9}{16}; \frac{3}{16}; \frac{3}{16}; \frac{1}{16}; \frac{3}{4}; \frac{3}{4}; \frac{5}{8}; \frac{3}{8}; \frac{1}{4}; \frac{1}{4}; \frac{15}{16}; \frac{13}{16}; \frac{13}{16}; \frac{7}{16}; 1\}

Probability Space

As long as the values of the probability measure are consistent with Kolmogorov's axioms and the consequences of those axioms, then we consider the probabilities to be mathematically acceptable, even if they are not reasonable for the given experiment.

Philosophical comment: Can the probability values assigned be considered reasonable as long as they're mathematically acceptable?
Random Variables Revisited

A random variable is a convenient way to express the elements of $\Omega$ as numbers rather than abstract elements of sets.

**Definition:** Measurable function
Let $A_X; A_Y$ be nonempty families of subsets of $X$ and $Y$, respectively. A function $f: X \to Y$ is $(A_X; A_Y)$-measurable if $f^{-1}(A) \in A_X$ for all $A \in A_Y$.

**Definition:** Random Variable
A *random variable* $X$ is a measurable function from the probability space $(\Omega, \Sigma, P)$ into the probability space $(\chi, A_X, P_X)$, where $\chi$ in $\mathbb{R}$ is the range of $X$ (which is a subset of the real line) $A_X$ is a Borel field of $X$, and $P_X$ is the probability measure on $\chi$ induced by $X$.

Specifically, $X: \Omega \to \chi$. 