Probability and Stochastic Processes
Lecture 4: Probability Spaces

Prof. Washim Uddin Mondal
Department of Electrical Engineering
Indian Institute of Technology Kanpur

Two Philosophies of Probability

Frequentist Approach

(a) Defines the probability of an event as the long-run limit of its relative frequency in infinitely many trials. If $n(A)$ denotes the number of times the event A occurs in n number of trials, then the probability of A is given as:

$$\mathbf{P}(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

(b) Drawbacks: Some experiments are not repeatable. The limit may not exist.

Axiomatic Approach

Probability is defined using a set of axioms.

Probability Spaces

Definition

Let $(\Omega, \mathcal{F})$ be a measurable space. We say that a function $\mathbf{P} : \mathcal{F} \to [0, 1]$ is a probability measure on $(\Omega, \mathcal{F})$ if the following conditions hold.

(a) $\mathbf{P}(\Omega) = 1$, i.e., the probability of the certain event is 1.

(b) $\mathbf{P}$ is countably additive (also called σ-additive), i.e., if $\{A_i\}_{i \in I}$ is a *countable* collection of *disjoint* $\mathcal{F}$-measurable sets, then the following holds.

$$\mathbf{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbf{P}(A_i)$$

- The tuple $(\Omega, \mathcal{F}, \mathbf{P})$ is called a probability space.

Probability Spaces

Examples

- Let $\Omega = \{a, b, c, d\}$ and $\mathcal{F} = \{\phi, \Omega, \{a, b\}, \{c, d\}\}$. If $\mathbf{P} : \mathcal{F} \to [0, 1]$ is such that
  (a) $\mathbf{P}(\phi) = 0,\ \mathbf{P}(\Omega) = 1,\ \mathbf{P}(\{a, b\}) = 0.25,\ \mathbf{P}(\{c, d\}) = 0.75$, then $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space.
  (b) $\mathbf{P}(\phi) = 0,\ \mathbf{P}(\Omega) = 1,\ \mathbf{P}(\{a, b\}) = 0.35,\ \mathbf{P}(\{c, d\}) = 0.75$, then $(\Omega, \mathcal{F}, \mathbf{P})$ is not a probability space.
- Let $\Omega = [0, 1]$ and $\mathcal{F} = \mathcal{B}([0, 1])$. If $\mathbf{P} : \mathcal{F} \to [0, 1]$ is such that
  (a) $\mathbf{P}([a, b]) = b - 2a$, $0 \leq a \leq b \leq 1$, then $(\Omega, \mathcal{F}, \mathbf{P})$ is not a probability space.

Probability Spaces

Exercise

Let $\Omega = [0,2]$, $\mathcal{F} = \mathcal{B}([0,2])$, and $\mathbf{P} : \mathcal{F} \to [0,1]$ be such that $\mathbf{P}([a,b]) = k(b^2 - a^2)$, $0 \le a \le b \le 2$. If $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, then answer the following questions.

(a) Find k.

(b) Find $\mathbf{P}(\{a\})$ and $\mathbf{P}([a,b))$ where $0 \le a < b \le 2$.

Solutions

(a) $\mathbf{P}(\Omega) = k(2^2 - 0^2) = 4k = 1$.

(b) $\mathbf{P}(\{a\}) = k(a^2 - a^2) = 0$.

(c) $\mathbf{P}([a,b)) + \mathbf{P}(\{b\}) = \mathbf{P}([a,b])$.

Properties of Probability Measures

Proposition 4.1

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and A, B be $\mathcal{F}$-measurable. The following holds.

(a) $\mathbf{P}(\phi) = 0$

(b) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$

(c) $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$

(d) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$

(e) $\mathbf{P}(A \cap B) \leq \mathbf{P}(B)$. Particularly, if $A \subset B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$.

Properties of Probability Measures

Proof of Proposition 4.1

(a) Note that $\phi \cap \Omega = \phi$ and $\phi \cup \Omega = \Omega$. Therefore, $\mathbf{P}(\phi) + \mathbf{P}(\Omega) = \mathbf{P}(\Omega) = 1$.

(b) Note that $A \cap A^c = \phi$ and $A \cup A^c = \Omega$. Hence, $\mathbf{P}(A) + \mathbf{P}(A^c) = \mathbf{P}(\Omega) = 1$.

(c) Recall that $(A \cap B) \cap (B \setminus A) = \phi$ and $(A \cap B) \cup (B \setminus A) = B$. Consequently,

$$\mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B) = \mathbf{P}(B)$$

(d) Recall that $A \cap (B \setminus A) = \phi$ and $A \cup (B \setminus A) = A \cup B$. Therefore,

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

(e) Note that $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A \cap B) \geq 0$.

Properties of Probability Measures

Inclusion-Exclusion Principle

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and A_i be $\mathcal{F}$-measurable $\forall i \in \{1, \cdots, n\}$. We have:

$$\begin{aligned}
\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{1 \le i < j \le n} \mathbf{P}(A_i A_j) + \cdots + (-1)^{n+1} \mathbf{P}(A_1 A_2 \cdots A_n) \\
&= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbf{P}(A_{i_1} \cdots A_{i_k})
\end{aligned}$$

Note

- $A_i A_j$ is a shorthand notation for $A_i \cap A_j$.

Properties of Probability Measures

Exercise

A student excels in Math and English with probabilities 0.25 and 0.3 respectively. The probability that she excels in at least one of these subjects is 0.4. What is the probability that she excels in both?

Solution

Let A, B be the events that the student excels in Math and English, respectively. Clearly, $\mathbf{P}(A) = 0.25$, $\mathbf{P}(B) = 0.3$, and $\mathbf{P}(A \cup B) = 0.4$. Using the inclusion-exclusion principle:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cup B) = 0.15$$

Properties of Probability Measures

Union Bound

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and A_i be $\mathcal{F}$-measurable, $\forall i \in \{1, \cdots, n\}$. We have:

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbf{P}(A_i)$$

Proof via Induction

- $\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2) \leq \mathbf{P}(A_1) + \mathbf{P}(A_2)$.
- $\mathbf{P}(\cup_{i=1}^{n} A_i) \leq \mathbf{P}(\cup_{i=1}^{n-1} A_i) + \mathbf{P}(A_n) \leq \sum_{i=0}^{n-1} \mathbf{P}(A_i) + \mathbf{P}(A_n)$.

Properties of Probability Measures

Example

Consider a communication scheme where each transmitted symbol s represents two bits i.e., $s \in \{00, 01, 11, 10\}$. The probabilities that the first and second bits are corrupted during transmission are 0.1 and 0.3, respectively. We say the communication is erroneous if either bit is corrupted. What is the least error-free communication probability that one can expect out of this scheme?

Solution

Let A_i be the event that the ith bit gets corrupted, $i \in \{1, 2\}$. Therefore, $\mathbf{P}(A_1) = 0.1$, and $\mathbf{P}(A_2) = 0.3$. Using union bound, we get $\mathbf{P}(A_1 \cup A_2) \leq 0.4$. Error-free communication probability is, therefore, $1 - \mathbf{P}(A_1 \cup A_2) \geq 0.6$.

Necessity of Measurable Spaces

Motivation

- We defined the σ-algebra as a collection of relevant or interesting events. However, there is more to the story!
- Say we are modeling the *uniform* probability on the unit interval $\Omega = (0, 1]$. Let the associated σ-algebra be the power set 2^{Ω}.
- For any $A \subset (0, 1]$, and $x \in (0, 1]$, define $A \oplus x$ as the following translated set.

$$A \oplus x = \{a + x | a \in A, a + x \leq 1\} \cup \{a + x - 1 | a \in A, a + x > 1\}$$

- If $\mathbf{P} : 2^{\Omega} \to [0, 1]$ is such that $\mathbf{P}(A) = \mathbf{P}(A \oplus x)$ for every $A \subset (0, 1]$ and $x \in (0, 1]$, then $\mathbf{P}$ is called a translation-invariant or a uniform probability measure.

Necessity of Measurable Spaces

Theorem

There does not exist any translation-invariant probability measure associated with the measurable space $(\Omega, 2^\Omega)$, where $\Omega = (0, 1]$.

Implications

- The σ-algebra 2^Ω is too large to be compatible with the notion of uniform probability.
- To define uniform probability, we need to abandon 2^Ω as the event space and settle for a much smaller σ-algebra, such as the Borel σ-algebra $\mathcal{B}((0, 1])$.
- The idea of a measurable space is not just for convenience; it is a necessity.

Extension Theorem

Motivation

- Given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, it is not easy to characterize $\mathbf{P} : \mathcal{F} \to [0, 1]$ by specifying $\mathbf{P}(B)$ for every $B \in \mathcal{F}$, especially if $\mathcal{F}$ is large or infinite.
- Does there exist a small $\mathcal{P} \subset \mathcal{F}$ such that specifying $\mathbf{P}(B)$ for every $B \in \mathcal{P}$ is enough to uniquely characterize $\mathbf{P}$ over the entire σ-algebra $\mathcal{F}$?

π-System

A non-empty set $\mathcal{P} \subset 2^{\Omega}$ is called a π-system of Ω if it is closed under intersection, i.e., $A \cap B \in \mathcal{P}$ whenever $A, B \in \mathcal{P}$.

Extension Theorem

Examples

- The set $\mathcal{P} = \{\{a\}, \{a, b, c\}\}$ is a π-system of $\Omega = \{a, b, c\}$.
- The collection of subsets of $\mathbb{R}$ of the following forms constitutes a π-system of $\mathbb{R}$.
  - $(-\infty, a]$, $[a, \infty)$, $(-\infty, a)$, (a, ∞) where $a \in \mathbb{R}$.
  - (a, b), $[a, b)$, $(a, b]$, $[a, b]$, where $a, b \in \mathbb{R}$, and $a < b$.

Note

- Every σ-algebra of Ω is a π-system of Ω, but not conversely. For example, $\mathcal{P} = \{(-\infty, a] \mid a \in \mathbb{R}\}$ is a π-system but not a σ-algebra of $\mathbb{R}$.

Extension Theorem

Dynkin's Extension Theorem

Let $\mathcal{P}$ be a π-system of Ω. If $(\Omega, \mathcal{F}, \mathbf{P}_1)$ and $(\Omega, \mathcal{F}, \mathbf{P}_2)$ are two probability spaces such that $\mathcal{F} = \sigma(\mathcal{P})$, and $\mathbf{P}_1(A) = \mathbf{P}_2(A)$, $\forall A \in \mathcal{P}$, then $\mathbf{P}_1(A) = \mathbf{P}_2(A)$, $\forall A \in \mathcal{F}$. In simple words, two probability measures agreeing on a π-system must agree on the σ-algebra generated by the π-system.

Corollary 4.2

Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two probability measures defined on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. If $\mathbf{P}_1((-\infty, a]) = \mathbf{P}_2((-\infty, a])$, $\forall a \in \mathbb{R}$, then $\mathbf{P}_1(A) = \mathbf{P}_2(A)$, $\forall A \in \mathcal{B}(\mathbb{R})$.

Extension Theorem

Proof of Corollary 4.2

- Note that $\mathcal{P} = \{(-\infty, a] \mid a \in \mathbb{R}\}$ is a π-system of $\mathbb{R}$. Moreover, $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$.

Notes

- Corollary 4.2 is important in the context of random variables.
- Similar results can be obtained for other π-systems that generate $\mathcal{B}(\mathbb{R})$. For example, if $\mathbf{P}_1$ and $\mathbf{P}_2$ are two probability measures associated with $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbf{P}_1((-\infty, a)) = \mathbf{P}_2((-\infty, a)), \forall a \in \mathbb{R}$, then $\mathbf{P}_1(A) = \mathbf{P}_2(A), \forall A \in \mathcal{B}(\mathbb{R})$.

Continuity of Probability Measures

Proposition 4.3

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability measure, and $A_i \in \mathcal{F}$, $\forall i \in \mathbb{N}$.

(a) If $(A_i)_{i\in\mathbb{N}}$ is an increasing sequence of events, i.e., $A_i \subset A_{i+1}$, $\forall i \in \mathbb{N}$, then

$$\mathbf{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \lim_{n\to\infty} \mathbf{P}(A_n)$$

(b) If $(A_i)_{i\in\mathbb{N}}$ is a decreasing sequence of events, i.e., $A_i \supset A_{i+1}$, $\forall i \in \mathbb{N}$, then

$$\mathbf{P}\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{n\to\infty} \mathbf{P}(A_n)$$

Continuity of Probability Measures

Example

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, where $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, and $\mathbf{P} : \mathcal{F} \to [0,1]$ be such that $\mathbf{P}((a,b)) = b - a$, $0 \le a < b \le 1$. Find $\mathbf{P}(\{a\})$ where $a \in [0,1]$.

Solution

- Define $A_n = (a - \frac{1}{n}, a + \frac{1}{n}) \cap [0,1]$, $\forall n \in \mathbb{N}$.
- Clearly, $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of $\mathcal{F}$-measurable sets. We have

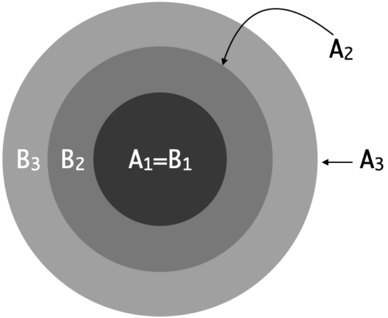
$$\mathbf{P}(\{a\}) = \mathbf{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \frac{2}{n} = 0$$

Continuity of Probability Measures

Proof of Proposition 4.3(a)

Define a sequence of $\mathcal{F}$-measurable sets $(B_i)_{i\in\mathbb{N}}$ such that $B_1 = A_1$ and $B_i = A_i \setminus A_{i-1}$, $\forall i \geq 2$. It is evident that B_i's are disjoint and the following holds $\forall n \in \mathbb{N}$.

$$\bigcup_{i=1}^{n} B_i = A_n \quad \text{and} \quad \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$



Continuity of Probability Measures

Proof of Proposition 4.3(a) (Continued)

Note the following chain of equalities.

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}(B_i) = \lim_{n\to\infty} \mathbf{P}\left(\bigcup_{i=1}^{n} B_i\right)$$

Proof of Proposition 4.3(b)

Note that $(A_i^c)_{i\in\mathbb{N}}$ is an increasing sequence of events. Using De Morgan's Law,

$$\mathbf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i^c\right) = 1 - \lim_{n\to\infty} \mathbf{P}(A_n^c) = \lim_{n\to\infty} \mathbf{P}(A_n)$$