Probability Theory

Abhijit Amrendra Kumar

August 2023

Chapter 1

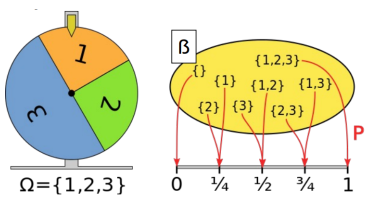
Introduction

1.1 Axioms

- **Set**: An unordered collection of unique objects
- **Experiment**: An empirical procedure
- **Statistical/Random Experiment**: An empirical procedure (aka. experiment) with an uncertain outcome
- **Sample Space** Ω: Set of all possible outcomes of a random experiment.
- **Event**: A subset of a sample space
  - Operations on events include union $A \cup B$, intersection $A \cap B$, complement $\bar{A}$, and subtraction $A - B$
- **Event space** β: A set of all possible events that we want to model
  - It is a subset of the **powerset** of the sample space Ω
  - Includes the empty set ϕ and the sample space Ω
  - It must be closed under countable unions
  - It must be closed under complementation
  - It is also referred to as a σ-algebra
- **Probability Function/Measure** P: A function which gives the probability/chance of the occurence of an event.
  - $P : \beta \to [0, 1]$
  - $P(\phi) = 0$
  - $P(\Omega) = 1$
  - For pairwise disjoint sets/events $A_1, A_2, ..., A_n$,

$$P(A_1 \cup A_2 \cup ...A_n) = P(A_1) + P(A_2) + ...P(A_n)$$

  - Operations of probability functions include complement $\bar{P}$, union $A \cup B$, and difference $A - B$
- **Probability Space** (Ω, β, P): A probability space is a triplet (Ω, β, P), where Ω is the sample space, β is the probability space, and P is the probability function.



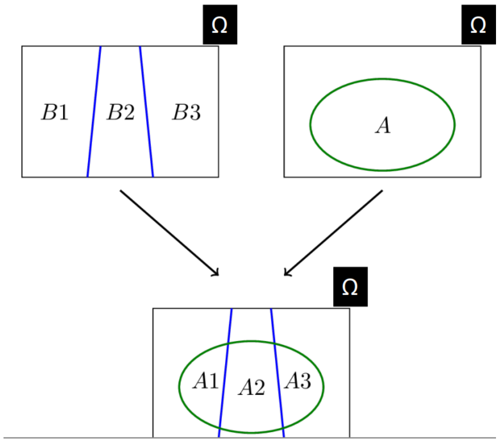
  - Probability Space for 2 Different Kinds of Experiments: $\Omega' = \Omega_1 \times \Omega_2$
  - Probability Space for Repeated Experiments: $\Omega' = \Omega \times \Omega$

1.2 Definitions

- **Population**: A set of similar items or events which is of interest for some question or experiment
- **Sample**: A set of objects collected/selected from a statistical population by a defined procedure
- **Joint Probability**: $P(A \text{ and } B) := P(A, B) := P(A \cap B)$
- **Conditional Probability**: Given an event B, the conditional probability of another event A is given by $P(A|B) := P(A \cap B)/P(B)$
- **Partitions**: A set of events $\{B_1, B_2, ..., B_n\}$ is a partition of a sample space Ω if
  - the set is mutually exclusive (i.e. $B_i \cap B_j = \phi \forall i \neq j$)
  - the set is exhaustive (i.e. $\bigcup_i B_i = \Omega$)
- **Total Probability**: Given a partition $B = \{B_1, B_2, ...B_n\}$, the total probability is given by

$$P(A) := \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$$

  - The partition B induces a partition over the event A



- **Independent Events**: Two events A and B are independent iff $P(A, B) = P(A)P(B)$
- **Conditional Independence**: Given event C, events A, B are conditionally independent iff

$$P(A, B|C) := P(A|C)P(B|C)$$

Chapter 2

Random Variable

2.1 Definition

A **random variable** X is a function defined on the probability space (Ω, β, P) that maps each element in the sample space Ω to a real number $X : \Omega \to \mathbb{R}$. Random variables are used when we are more interested in the value associated with an outcome instead of the outcome itself.

Discrete RV: $X : \Omega \to S$, where $S \subset \mathbb{R}$ and the cardinality of the set $|S|$ is countably infinite $|S| \leq |\mathbb{N}|$.
Continuous RV: $X : \Omega \to S$, where $S \subset \mathbb{R}$ and the cardinality of the set $|S|$ is uncountably infinite $|S| \leq |\mathbb{R}|$.

2.2 Events via Random Variables

Notation: Upper case letters (eg. X) will be used to denote a random variable, while lowercase letters (eg. x) will be used to denote a specific value that can be taken by a random variable.

Examples:

$$\{X = a\} = \{s \in \Omega : X(s) = a\}$$
$$\{X < a\} = \{s \in \Omega : X(s) < a\}$$
$$\{a < X < b\} = \{s \in \Omega : a < X(s) < b\}$$

Using this notation, we can also define event probabilities

$$P_X(\{X = a\}) = P(\{s \in \Omega : X(s) = a\})$$
$$P_X(\{X < a\}) = P(\{s \in \Omega : X(s) < a\})$$
$$P_X(\{a < X < b\}) = P(\{s \in \Omega : a < X(s) < b\})$$

2.3 Distribution Functions via Random Variables

2.3.1 Cumulative Distribution Function (CDF)

For a real-valued random variable X, the CDF is defined as $f_X(x) := P_X(X \leq x)$.

Properties of CDF:

- f is monotonically non-decreasing.
- f is right-continuous, i.e. $\lim_{\epsilon \to 0^+} f(x + \epsilon) = f(x), \forall\ x \in \mathbb{R}$
- $\lim_{x \to -\infty} f(x) = 0$
- $\lim_{x \to +\infty} f(x) = 1$

Theorem: Let X be a random variable with CDF f_X. Then,

$$P(a < X \leq b) = f_X(b) - f_X(a)$$

Proof: We know that

$$\{-\infty < X \leq b\} = \{-\infty < X \leq a\} \cup \{a < X \leq b\}$$

The two sets on the right side are disjoint, which implies

$$\begin{aligned}
P_X(-\infty < X \leq b) &= P_X(X \in \{-\infty < X \leq a\} \cup \{a < X \leq b\}) \\
&= P_X(-\infty < X \leq a) + P_X(a < X \leq b) \\
\implies P_X(a < X \leq b) &= P_X(X \leq b) - P_X(X \leq a) \\
&= f_X(b) - f_X(a)
\end{aligned}$$

Theorem: Let X be a random variable with CDF f_X. Then,

$$P(X = c) = f_X(c) - f_X(c^-)$$

Proof: For all $x \in \mathbb{R}$, we have

$$\{x\} = \bigcap_{n=1}^{\infty} \big(x - 1/n, x\big]$$

that is, $\{x\}$ is the limit of a decreasing sequence of sets. Thus we can write

$$\begin{aligned}
P_X(X = x) &= P_X\Big[\bigcap_{n=1}^{\infty} \{x - 1/n < X \leq x\}\Big] \\
&= \lim_{n\to\infty} P_X\big[x - 1/n < X \leq x\big] \\
&= \lim_{n\to\infty} \big[F_X(x) - F_X(x - 1/n)\big] \\
&= F_X(x) - F_X(x^-)
\end{aligned}$$