CST 3526 Stochastic Process	Lecture 1 - 09/18/2025

Lecture 1: Review of Probability Theory I

Lecturer: Ziqiao Wang

In this lecture[1], we will review the fundamental concepts of probability theory, including the definitions of sample spaces and events, the axioms of probability, conditional probability, independence of events, Theorem of Total Probability, and random variables along with their CDFs, PMFs and PDFs.

1 Random Experiments

A random experiment is an experiment whose outcome varies in an unpredictable manner when repeated under the same conditions. To properly define a random experiment, one must provide an experimental procedure with a clear and unambiguous description of what is being measured or observed.

Example 1. Specifying random experiments.

Experiment E_1*:* Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and that balls 3 and 4 are white. Note the number and color of the ball you select.

Experiment E_2*:* Toss a coin three times and note the sequence of heads and tails.

Experiment E_3*:* Toss a coin three times and note the number of heads.

Experiments E_4: Pick two numbers at random between zero and one.

Definition 1.1 (Sample Space)**.** The sample space Ω of a random experiment is defined as the set of all possible outcomes.

Note that an outcome (or sample point) ω of a random experiment is an elementary result that cannot be further decomposed into simpler results. In other words, outcomes are *mutually exclusive*—no two outcomes can occur at the same time.

For example, $\Omega_1 = \{\{1, b\}, \{2, b\}, \{3, w\}, \{4, w\}\}$, $\Omega_2 = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}, \mathrm{TTH}, \mathrm{THT}, \mathrm{HTT}, \mathrm{TTT}\}$, $\Omega_3 = \{0, 1, 2, 3\}$ and $\Omega_4 = \{(x, y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\}$.

Definition 1.2 (Discrete Sample Space)**.** Ω is countable.

Definition 1.3 (Continuous Sample Space)**.** Ω is not countable.

Definition 1.4 (Events)**.** An event is a subset of Ω.

Remark 1.1. *The* **certain event** *consists of all outcomes and hence always occurs, and the* **impossible** *or* **null event** $\varnothing$ *contains no outcomes and hence never occurs. An event from a discrete sample space that consists of a single outcome is called an* **elementary event***.*

2 The Axioms of Probability

Probabilities are numerical values assigned to events that quantify how "likely" it is about their occurrence when a random experiment is performed. A **probability law** for a random experiment E is a rule that assigns probabilities to the events associated with E. Formally, a probability law is a function $P : \mathcal{F} \to [0, 1]$, where $\mathcal{F}$ is a collection of events (subsets of the sample space Ω). For any event $A \in \mathcal{F}$, the value $P(A)$ is called the probability of A.

[1] *Reading:* Chapter 2-3 of Leon-Garcia.

Any probability assignment must satisfy:

Axiom I:	$P(A) \geq 0$	(probability is a nonnegative measure)
Axiom II:	$P(\Omega) = 1$	(probability is a finite measure)
Axiom III:	If $A \cap B = \varnothing$ then $P(A \cup B) = P(A) + P(B)$	(additivity property)

Remark 2.1. *The probability of an event is an attribute similar to physical mass. Axiom I states that the probability (mass) is nonnegative, and Axiom II states that there is a fixed total amount of probability (mass), namely 1 unit. Axiom III states that the total probability (mass) in two disjoint objects is the sum of the individual probabilities (masses).*

Below we list some properties developed from the axioms.

Corollary 2.1.
1. $P(A^c) = 1 - P(A)$.
2. $P(A) \leq 1$.
3. $P(\varnothing) = 0$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *and* $P(A \cup B) \leq P(A) + P(B)$.
5. *If* $A \subset B$, *then* $P(A) \leq P(B)$.
6. *Let* $A_1, A_2, \ldots, A_n$ *be a finite collection of pairwise disjoint events. Then* $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Initial probability assignment. Once the sample space Ω is specified, an initial probability assignment must be introduced to define the probabilities of events, subject to the axioms of probability. If Ω is discrete, it is sufficient to assign probabilities to the elementary outcomes. If Ω is continuous, it is sufficient to specify the probabilities of basic sets such as intervals on the real line or regions in the plane. The probabilities of more complex events can then be derived from this initial assignment using the probability axioms and their corollaries.

3 Conditional Probability and Independence of Events

Definition 3.1 (Conditional Probability)**.** For two events A and B, the conditional probability $P(A|B)$ of event A given B (has occurred) is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ for } P(B) > 0.$$

Computing $P(A|B)$ can be interpreted as restricting the sample space to B. Within this reduced sample space, the event A occurs if and only if the outcome ω lies in $A \cap B$. Thus, the conditional probability $P(A|B)$is obtained by renormalizing the probabilities of events relative to B, so that the total probability within B equals 1.

Let $B_1, B_2, \ldots, B_n$ be mutually exclusive events whose union is the sample space Ω. The collection $\{B_1, B_2, \ldots, B_n\}$ is called a **partition** of Ω. Any event A can then be expressed as the union of mutually exclusive components:

$$A = A \cap \Omega = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

Hence, by Corollary 2.1, we have

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n).$$

This result may be rewritten as

$$\boxed{P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n),}$$

which is known as the **Theorem of Total Probability**. Consider now, assuming $P(A) > 0$,

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)}.$$

Applying the Theorem of Total Probability to the denominator of this, we find that

$$\boxed{P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_j)P(B_j)},}$$

which is **Bayes' rule** and is a basic result relating different conditional probabilities.

For example, the Monty Hall problem.

Independence of Events If the occurrence of an event B does not change the probability of another event A, we say that A is independent of B. Namely, this means $P(A|B) = P(A)$, $P(B) > 0$.

Definition 3.2 (Independence of Events)**.** Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.

Remark 3.1 (Independence v.s. Mutual Exclusiveness)**.** *If $A \cap B = \varnothing$, then A and B are said to be mutually exclusive. In general,* **if two events have nonzero probability and are mutually exclusive, then they cannot be independent***. Conversely, independence does not imply mutual exclusivity. In fact, if two events A and B are both independent and mutually exclusive, then at least one of them must have probability zero, since both $A \cap B = \varnothing$ and $P(A \cap B) = P(A)P(B) = 0$ hold.*

Remark 3.2 (Mutually Independence v.s. Pairwise Independence)**.** *Three events A, B, and C are said to be* ***mutually independent*** *if the probability of the intersection of any subset of these events equals the product of the probabilities of the individual events. That is,*

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C), \quad P(A \cap B \cap C) = P(A)P(B)P(C).$$

If only the first three conditions hold, the events are called ***pairwise independent****. Importantly,* **pairwise independence does *not* in general imply mutual independence**.

Example 2 (Pairwise but not Mutual Independence)**.** Consider two numbers x and y are selected at random from the unit interval (i.e. $[0, 1]$) (i.e. the random experiment E_4). Let the events A, B, and C be defined as follows:

$$A = \left\{y > \tfrac{1}{2}\right\}, \quad B = \left\{x < \tfrac{1}{2}\right\},$$
$$C = \left\{x < \tfrac{1}{2},\, y < \tfrac{1}{2}\right\} \cup \left\{x > \tfrac{1}{2},\, y > \tfrac{1}{2}\right\}.$$

It can be easily verified that any pair of these events is independent:

$$P(A \cap B]) = \tfrac{1}{4} = P(A)P(B),$$

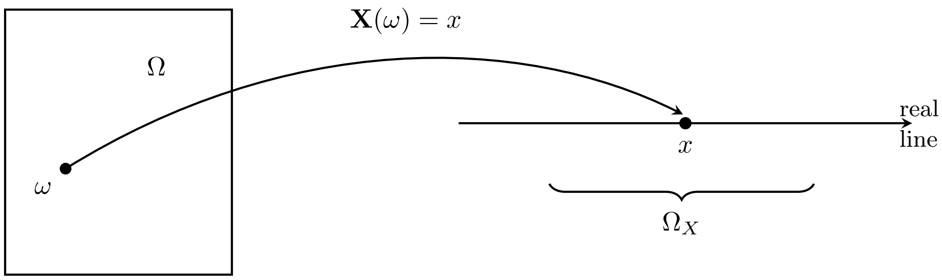
$$P(A \cap C) = \tfrac{1}{4} = P(A)P(C),$$
$$P(B \cap C) = \tfrac{1}{4} = P(B)P(C).$$

However, the three events are not independent, since $A \cap B \cap C = \varnothing$, so

$$P(A \cap B \cap C) = P(\varnothing) = 0 \neq P(A)P(B)P(C) = \tfrac{1}{8}.$$

4 Random Variables

Definition 4.1 (Random Variable)**.** A random variable X is a function that assigns a real number, $X(\omega)$,to each outcome ω in the sample space Ω of a random experiment.



The sample space Ω is the domain of a random variable, and the set

$$\Omega_X = \{X(\omega) : \omega \in \Omega\}$$

of all possible values taken by X is called the *range* (or support) of the random variable. Clearly, $\Omega_X \subseteq \mathbb{R}$. Throughout, we adopt the convention that capital letters (e.g., X, Y) denote random variables, while lowercase letters (e.g., x, y) denote particular values (i.e. realizations) of these random variables.

Remark 4.1. *The mapping $X : \Omega \to \mathbb{R}$ that assigns a value to each outcome is fixed and deterministic. Thus, the randomness observed in X is entirely induced by the randomness inherent in the underlying experiment.*