Chapter 1

Elementray Probability Theory

1.1 Sample Spaces and Events

1.1.1 Sample Spaces

We consider a randon **experiment** whose range of possible outcomes can be described by a set S called the **sample space**.

Examples:

- Coin tossing. $S = H, T$.
- Die rolling. $S = 1, 2, 3, 4, 5, 6$.
- Throwing two coins. $S = (H, H), (H, T), (T, H), (T, T)$.

1.1.2 Events

An event E is any subset of the sample space, $E \subseteq S$, a collection of some possible outcomes.

Examples:

- Coin tossing. $E = H$, $E = T$.
- Throwing two coins. $E =$ **different** $= (H, T), (T, H)$.

The null event $\emptyset$ is an extreme possible event of S.

Sets that contain exactly one element are called **singleton** subsets. These events are called **elementary events** of S.

The smallest events that can occur are the singleton subsets.

Properties of Events

- The outcome of a random experiment is a single element $s^* \in S$.
- For event $E \subseteq S$, the event is occurred iff $s^* \in E$.
- $s^* \notin E \leftrightarrow s^* \in \bar{E}$.

Conclusively,

- The null event $\emptyset$ will not occur.
- The universal event S will always occur because all $s^* \in S$.

1.1.3 Set Operations on Events

The event $\bigcup\limits_i E_i$ will occur iff at least one of the events E_i occurs.
The event $\bigcap\limits_i E_i$ will occur iff all the events E_i occurs.
Events are said to be **mutually exclusive** if $\forall i, i.E_i \cap E_j = \emptyset$ (i.e. disjoint —at most one can occur).

1.2 Axioms of Probability

1.2.1 σ-Algebra of Events

A collection of sets, $\mathcal{F}$, is a **σ-algebra**.
$\mathcal{F}$ must be:

- Nonempty, $S \in \mathcal{F}$.
- closed under complements: $E \in \mathcal{F} \Rightarrow \bar{E} \in \mathcal{F}$. closed under countable union: $E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \bigcup\limits_i E_i \in \mathcal{F}$.

Example: For S, an $\mathcal{F}$ can be $\mathcal{F} = S, \emptyset$ because $\emptyset$ is the negation of S.

Probability Measures

A **probability measure** on the pair $(S, \mathcal{F})$ is a mapping $P : \mathcal{F} \to [0, 1]$ satisfying the axioms for all subsets S which it is defined:

Axiom 1 $\forall E \in \mathcal{F}.0 \Leftarrow P(E) \Leftarrow 1$

Axiom 2 $P(S) = 1$

Axiom 3 Countably addictive: For disjoint subsets $E_1, E_2, \ldots \in \mathcal{F}$,

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

Basic results:

- $P(\bar{E}) = 1 - P(E)$
- $P(\emptyset) = 0$
- For any events E and F:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Propositions:

- If events E and F are independent, then $\bar{E}$ and F are also independent.

1.3 Interpretation of Probability

1.3.1 Classical Interpretation

If S is finite and the elementary events are considered *equally likely*, then for an event $E \in S$:

$$P(E) = \frac{|E|}{|S|}$$

Note that $|E|$ means the cardinality of E, i.e. the time in which event E occurs.

1.3.2 Frequentist Interpretation

If one takes repeated observations in *identical* random situations, in which event E may or may not occur, then the proportion of time in which E occurs tends to a limiting value —the probability of E.

1.4 Independent Events

Two events E and F are said to be **independent** iff $P(E \cap F) = P(E)P(F)$.

This also applies more generally to multiple events, intuitively. Formal definition on Lecture Video at 3:41.

1.4.1 Marginal Events

For the Coin and Die game, where we toss a coin and a die at the same time, consider each of the 12 possible combinations of head/tail and die values.

The **probability table** for this is:

	1	2	3	4	5	6	
H	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
T	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

The sum of rows and columns, $P(H)$, $P(T)$, $P(1)$, $P(2)$, $\ldots$, $P(6)$ are called **marginal events**.

Probabilities such as $P(H,1)$, or generally $P(E \cap F)$, are called **joint probability**.

1.4.2 Dependent Events

A crooked die, namely a *top*, has the same faces on opposite sides, so it has only odd numbers.

If we change the game rule, in the way that we first flip the coin. If it comes up heads, then we roll the normal die, otherwise roll the *top*.

The new probability table is:

	1	2	3	4	5	6	
H	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
T	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{2}$
	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	

Now the two experiments are called **dependent**, because the probabilities of the different outcomes of the die change according to the outcome of the coin toss.

1.5 Conditional Probability

For $E, F \in S$, and $P(F) \neq 0$, the conditional probability of E occurring given that we know F has occurred is defined as:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

If E and F are *independent*, then $P(E|F) = P(E)$.

Informally, conditioning can be thought of a shrinking of the sample space.

1.5.1 Conditional Independence

For events E_1, E_2 and F, the event pair E_1 and E_2 are **conditionally independent** iff:

$$P(E_1 \cap E_2|F) = P(E_1|F)P(E_2|F)$$

1.5.2 Bayes Theorem

$$P(E|F) = \frac{P(E)P(F|E)}{P(F)}$$

Informally, Bayes Theorem introduces a way to determine conditional events 'reversely' (with information on E).

1.5.3 Partition Rule

Partition Rule (also known as the Law of Total Probability) states that:

$$P(E) = \sum_i P(E|F_i)P(F_i) = \sum_i P(E \cap F_i)$$

Partition Rule indicates the probability of E can be 'split' into partitions of smaller events based on some conditions.

Special Case

For any event S, $E, \bar{E}$ forms a partition of S, so by the Partition Rule,

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap \bar{F}) \\ &= P(E|F)P(F) + P(E|\bar{F})P(\bar{F}) \end{aligned}$$

Chapter 2

Random Variables

2.1 Probability Spaces

A **probability spaces** $(S, \mathcal{F}, P)$ that models random experiment by means of probability measure $P(E)$ defined on subsets $E \subseteq S$ of the sample space S belonging to the sigma algebra $\mathcal{F}$.

2.1.1 Random Variables

A **random variable** is a mapping from the sample space to $\mathbb{R}$.

Discrete Random Variables Countable Random variables are called **discrete**.

Simple Random Variables Discrete Random Variables with a finite set of possible outcomes are called **simple**.

Continuous Random Variables

Every $s \in S$ has corresponding $X(s)$. Multiple s can be mapped to the same random variable.
Informally, it defines how to 'assign' an event with a number, so we can use the number (the random variable) to represent the event.
For example, we can map dice faces to variables X: $X(\text{"Dice Face is 1"}) = 1$, $X(\text{"Dice Face is 2"}) = 2$, $\ldots$. Then we have $P_X(1 \leq X \leq 2) = P_X(\text{"Dice Face is 1"}, \text{"Dice Face is 2"}) = \frac{2}{6} = \frac{1}{3}$

2.1.2 Induced Probability

For each $x \in \mathbb{R}$, let $S_x \subseteq S$, such that $S_x = s \in S | X(s) \leq x$, then we write:

$$P_X(X \leq X) \equiv P(S_x)$$

P_X is the **induced probability** on the random variable X in $\mathbb{R}$.

Informally, Induced Probability allows us to model any problem in an interval. For example, event $E = X \Leftarrow b$ can be written as $E = F \cup G$ where $F = (-\infty, a)$ and $G = (b]$.

Support

The *image* of S under X is called the **support** of X:

$$\text{supp}(X) \equiv X(S) = x \in \mathbb{R} | \exists s \in S.X(s) = x$$

So supp(X) contains all the possible outcomes for the random variable X. So $P_X(X \leq x)$ is defined for all $x \in \text{supp}(X)$.

2.1.3 Example Random Variable Problem

For the game of tossing a fair coin, suppose we win 1 if we get heads, or we lose 1 otherwise. Then we have:

$$X(T) = -1$$
$$X(H) = 1$$

$$S_x = \begin{cases} \emptyset & \text{if} x < -1; \\ T & \text{if} -1 \leq x < 1; \\ H, T & \text{if} 1 < x. \end{cases}$$

$$P_X(X \leq x) = P(S_x) = \begin{cases} P(\emptyset) = 0 & \text{if} x < -1' \\ P(T) = \frac{1}{2} & \text{if} -1 \leq x < 1; \\ P(H, T) = 1 & \text{if} 1 < x. \end{cases}$$

2.1.4 Cumulative Distribution Function

$$F_X(x) = P_X(X \leq x)$$

Properties of CDF

Monotonicity $\forall x_1, x_2 \in \mathbb{R}.x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$

$F_X(-\infty) = 0, F_X(\infty) = 1$

F_X is right-continuous.

$0 \leq F_X(x) \leq 1$.

$$P_X(a < X \leq b) = F_X(b) - F_X(a) \tag{2.1}$$

2.2 Discrete Random Variables

A random variable is **discrete** if it can take only a countable number of possible values:

$$X \text{ is discrete} \iff \operatorname{supp}(X) \text{ is countable}$$

Each sample space element $s \in S$ is mapped by X to one of the values $\mathcal{X} = \operatorname{supp}(X) = x_1, x_2, \ldots$. The probability of a discrete random variable x_1 can be written as:

$$P_X(X = x_i) = P(E_i) = F_X(x_i) - F_X(x_{i-1})$$

2.2.1 CDF for Discrete Random Variables

CDF for Discrete Random Variables is a monotonic increasing step function which jumps at points in $\mathcal{X}$, continuous on the right.