1 Basic Concepts of Probability

Definitions:

A **statistical experiment** is any procedure that produces data or observations.

The **sample space**, denoted by S, is the set of all possible outcomes of a statistical experiment.

A **sample point** is an outcome (element) in the sample space.

An **event** is a subset of the sample space.

Multiplication Principle: Suppose that r different experiments are to be performed sequentially, and they have $n_1, n_2, \dots, n_r$ possible outcomes respectively. Then there are $n_1 \cdot n_2 \cdot \dots n_r$ possible outcomes for the r experiments.

Addition Principle: Suppose that an experiment can be performed by k different procedures, and the "ways" under different procedures *do not overlap*. Then the total number of ways we can perform the experiment is $n_1 + n_2 + \cdots + n_k$.

Permutation:

$$P^n_r = \frac{n!}{(n-r)!}$$

Combination:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Axioms of Probability:

1. For any event A, $0 \le P(A) \le 1$
2. For the sample space, $P(S) = 1$
3. For any two mutually exclusive events A and B i.e. $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

Propositions:

The probability of the empty set is 0, $P(\emptyset) = 0$.

If $A_1, A_2, \dots, A_n$ are mutually exclusive events, i.e. $A_i \cap A_j = \emptyset$ for any $i \ne j$, then $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$

$P(A') = 1 - P(A)$

$P(A) = P(A \cap B) + P(A \cap B')$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If $A \subset B$, then $P(A) \le P(B)$

Finite Sample Space with Equally Likely Outcomes: For a sample space $S = \{a_1, a_2, \dots, a_k\}$, assume all outcomes are equally likely to occur, i.e. $P(a_1) = P(a_2) = \cdots = P(a_k)$. Then for any event $A \subset S$,

$$P(A) = \frac{\text{number of sample points in A}}{\text{number of sample points in S}}$$

Conditional Probability: For any two events A and B with $P(A) > 0$, the conditional probability of B given that A has occurred is defined by

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

Multiplication Rule: $P(A \cap B) = P(A)P(B|A)$ if $P(A) \ne 0$, or $P(A \cap B) = P(B)P(A|B)$ if $P(B) \ne 0$

Inverse Probability Formula:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Independence: Two events A and B are independent if and only if $P(A \cap B) = P(A)P(B)$. We denote this by $A \perp B$. If A and B are not independent, they are said to be dependent, denoted by $A \not\perp B$. If $P(A) \ne 0, A \perp B$ iff $P(B|A) = P(B)$ (Likewise for $P(B) \ne 0$).

Independent vs Mutually Exclusive:

A, B independent $\Leftrightarrow P(A \cap B) = P(A)P(B)$

A, B mutually exclusive $\Leftrightarrow A \cap B = \emptyset$

Partition: If $A_1, A_2, \dots, A_n$ are mutually exclusive events and $\cup_{i=1}^n A_i = S$, we call $A_1, A_2, \dots, A_n$ a partition of S.

Law of Total Probability: Suppose $A_1, A_2, \dots, A_n$ is a partition of S. Then for any event B, we have

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

Bayes' Theorem: Let $A_1, A_2, \dots, A_n$ be a partition of S, then for any event B and $k = 1, 2, \dots, n$,,

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

When $n = 2$,

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$$

2 Random Variables

Random Variable: Let S be the sample space of an experiment. A function X, which assigns a real number to every $s \in S$ is called a random variable.

Range Space: The **range space** of X is the set of real nmbers

$$R_X = \{x \mid x = X(s), s \in S\}$$

Each possible value x of X corresponds to an event that is a subset or element of the sample space S.

Types of random variables:

Discrete: the number of values in R_X is **finite** or **countable**. That is, we can write $R_X = \{x_1, x_2, \dots\}$

Continuous: R_X is an **interval** or **collection of intervals**

Probability Mass Function: For a discrete RV X, define

$$f(x) = \begin{cases} P(X = x) \text{ for } x \in R_X \\ 0 \text{ for } x \notin R_X \end{cases}$$

Then $f(x)$ is known as the probability function (pf), or probability mass function (pmf) of X. The collection of pairs $(x_i, f(x_i))$, $i = 1, 2, 3, \dots$ is the probability distribution of X.

Properties of Probability Mass Function:

(1) $f(x_i) \ge 0$ for all $x_i \in R_X$

(2) $f(x) = 0$ for all $x \notin R_X$

(3) $\sum_{i=1}^\infty f(x_i) = \sum_{x_i \in R_X} f(x_i) = 1$

For any set $B \subset \mathbb{R}$, $P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i)$

Probability Density Function: The pdf of a continuous random variable X, denoted $f(x)$, is a function that satisfies:
(1) $f(x) \geq 0$ for all $x \in R_X$ and $f(x) = 0$ for $x \notin R_X$
(2) $\int_{R_X} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$
(3) For any a and b s.t. $a \leq b$, $P(a \leq X \leq b) = \int_a^b f(x)dx$
For any specific value x_0, $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$. Hence $P(A) = 0$ but A is not necessarily $\emptyset$
Furthermore, $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx$
Cumulative Distribution Function: For any random variable X, it's cdf is defined by $F(x) = P(X \leq x)$
CDF - Discrete RV: If X is a discrete RV,

$$F(x) = \sum_{t \in R_X;\, t \leq x} f(t) = \sum_{t \in R_X;\, t \leq x} P(X = t)$$

The cdf of a discrete RV is a step function.
For any two numbers $a < b$, we have $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a-)$, where "$a-$" represents the "largest value in R_X that is smaller than a". Mathematically, $F(a-) = \lim_{x \uparrow a} F(x)$
CDF - Continuous RV: If X is a continuous RV,

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

and $f(x) = \frac{dF(x)}{dx}$. Further, $P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$
(i) No matter if X is discrete or continuous, $F(x)$ is non-decreasing. In the sense that for any $x_1 < x_2$, $F(x_1) \leq F(x_2)$.
(ii) The probability function and cdf have a one-to-one correspondence. I.e. for any pdf/pmf, the cdf can be uniquely determined, and vice versa
(iii) The ranges of $F(x)$ and $f(x)$ satisfy: $0 \leq F(x) \leq 1$, for discrete distributions, $0 \leq f(x) \leq 1$, for continuous distributions, $f(x) \geq 0$ but not necessarily that $f(x) < 1$
Expectation - Discrete RV: Let X be a discrete random variable with $R_X = \{x_1, x_2, x_3, \ldots\}$ and probability function $f(x)$. The expectation or mean of X is defined by

$$E(X) = \sum_{x_i \in R_X} x_i f(x_i)$$

By convention, we denote $\mu_X = E(X)$
Expectation - Continuous RV: Let X be a continuous RV with probability function $f(x)$. The expectation or mean of X is defined by

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_{x \in R_X} x f(x)dx$$

Note: the mean of X is not necessarily a possible value of X
Properties of Expectation: Let X be a random variable, and a, b any real numbers. Then $E(aX + b) = aE(X) + b$.
Let X and Y be two random variables. Then $E(X+Y) = E(X)+E(Y)$
Variance: Let X be a random variable. The variance of X is defined as

$$\sigma_x^2 = V(X) = E(X - \mu_X)^2$$

This applies whether X is discrete or continuous. If X is a discrete RV with pmf $f(x)$ and range R_X,

$$V(X) = \sum_{x \in R_X} (x - \mu_X)^2 f(x)$$

If X is a continuous RV with pdf $f(x)$,

$$V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx$$

$V(X) \geq 0$ for any X. Equality holds iff $P(X = E(X)) = 1$, i.e. when X is a constant.
Let a and b be any real numbers, then $V(aX + b) = a^2 V(X)$.
$V(X) = E(X^2) - [E(X)]^2$
Standard deviation, $\sigma_X = \sqrt{V(X)}$