Chapter 1. Review of Probability Theory and Statistics

1 Probability Space and Rules of Probability

To any experiment we assign its **sample space**, denoted by S, consisting of all its possible outcomes (called **elementary events**, denoted by $e_i,\ i \in \mathbb{N}$).
An **event** is a subset of S (events are denoted by capital letters, $A, B, A_i,\ i \in \mathbb{N}$).

Since events are defined as sets, we use set theory in describing them.

- two special events associated with every experiment:
  - the **impossible** event, denoted by $\emptyset$ ("never happens");
  - the **sure (certain)** event, denoted by S ("surely happens").
- for events, we have the usual operations of sets:
  - **complementary** event, $\overline{A}$,
  - **union** of A and B, $A \cup B = \{e \in S \mid e \in A \text{ or } e \in B\}$, the event that occurs if either A or B or both occur;
  - **intersection** of A and B, $A \cap B = \{e \in S \mid e \in A \text{ and } e \in B\}$, the event that occurs if both A and B occur;
  - **difference** of A and B, $A \setminus B = \{e \in S \mid e \in A \text{ and } e \notin B\} = A \cap \overline{B}$, the event that occurs if A occurs and B does not;
  - A **implies (induces)** B, $A \subseteq B$, if every element of A is also an element of B, or in other words, if the occurrence of A induces (implies) the occurrence of B; A and B are equal, $A = B$, if A implies B and B implies A;
- two events A and B are **mutually exclusive (disjoint, incompatible)** if A and B cannot occur at the same time, i.e. $A \cap B = \emptyset$;
- three or more events are mutually exclusive if **any two of them are**, i.e.

$$A_i \cap A_j = \emptyset,\ \forall i \neq j;$$

- events $\{A_i\}_{i \in I}$ are **collectively exhaustive** if $\bigcup_{i \in I} A_i = S$;

- events $\{A_i\}_{i\in I}$ form a **partition** of S if the events are collectively exhaustive and mutually exclusive, i.e.
$$\bigcup_{i\in I} A_i = S, \text{ and } A_i \cap A_j = \emptyset, \forall i,j \in I,\ i \neq j.$$
- we consider all events relating to an experiment to belong to a **σ-field**, $\mathcal{K}$, a collection of events from from S, an algebraic structure that allows all the usual set operations (mentioned above) within itself (e.g. the power set $\mathcal{P}(S) = \{S' | S' \subseteq S\}$).

Definition 1.1. *Let $\mathcal{K}$ be a σ-field over S. A mapping $P : \mathcal{K} \to \mathbb{R}$ is called* ***probability*** *if it satisfies the following conditions:*

(i) $P(S) = 1$;

(ii) $P(A) \geq 0$, *for all* $A \in \mathcal{K}$;

(iii) *for any sequence* $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{K}$ *of mutually exclusive events,*
$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n). \tag{1.1}$$

The triplet $(S, \mathcal{K}, P)$ is called a ***probability space****.*

Theorem 1.2. *(*Rules of Probability*)*
Let $(S, \mathcal{K}, P)$ be a probability space, and let $A, B \in \mathcal{K}$. Then the following properties hold:

a) $P(\overline{A}) = 1 - P(A)$.

b) $0 \leq P(A) \leq 1$.

c) $P(\emptyset) = 0$.

d) $P(A \setminus B) = P(A) - P(A \cap B)$.

e) *If $A \subseteq B$, then $P(A) \leq P(B)$, i.e. P is monotonically increasing.*

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

g) *more generally,*
$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i<j\leq n} P(A_i \cap A_j) + \sum_{1\leq i<j<k\leq n} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right), \text{ for all } n \in \mathbb{N}.$$

Definition 1.3. *Let* $(S, \mathcal{K}, P)$ *be a probability space and let* $B \in \mathcal{K}$ *be an event with* $P(B) > 0$. *Then for every* $A \in \mathcal{K}$*, the* ***conditional probability of*** A ***given*** B *(or the* ***probability of*** A ***conditioned by*** B*) is defined by*

$$P(A|B) \quad = \quad \frac{P(A \cap B)}{P(B)}. \tag{1.2}$$

Theorem 1.4. (Rules of Probability − Continued)

h) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

i) *Multiplication Rule*
$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1) \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1})$.

j) *Total Probability Rule*

- $P(A) = P(B)P(A|B) + P\left(\overline{B}\right) P\left(A|\overline{B}\right)$.
- *in general, if* $\{A_i\}_{i \in I}$ *is a partition of* S*,*

$$P(A) \quad = \quad \sum_{i \in I} P\left(A_i\right) P\left(A|A_i\right). \tag{1.3}$$

Definition 1.5. *Two events* $A, B \in \mathcal{K}$ *are* ***independent*** *if*

$$P(A \cap B) = P(A)P(B). \tag{1.4}$$

- A, B independent $<=> P(A|B) = P(A) <=> P(B|A) = P(B)$.
- $A = \emptyset$ or $A = S$ and $B \in \mathcal{K}$, then A, B independent.
- A, B independent $<=> \overline{A}, B$ independent $<=> \overline{A}, \overline{B}$ independent.

Definition 1.6. *Consider an experiment whose outcomes are finite and equally likely. Then the* ***probability*** *of the event* A *is given by*

$$P(A) \quad = \quad \frac{\text{number of favorable outcomes for the occurrence of } A}{\text{total number of possible outcomes of the experiment}} \quad \overset{not}{=} \quad \frac{N_f}{N_t}. \tag{1.5}$$

Remark 1.7. This notion is closely related to that of *relative frequency* of an event A: repeat an experiment a number of times N and count the number of times event A occurs, N_A. Then the relative frequency of the event A is

$$f_A = \frac{N_A}{N}.$$

Such a number is often used as an approximation to the probability of A. This is justified by the fact that

$$f_A \overset{N\to\infty}{\longrightarrow} P(A).$$

The relative frequency is used in computer simulations of random phenomena.

2 Probabilistic Models

Binomial Model

This model is used when the trials of an experiment satisfy three conditions, namely

(i) they are independent,

(ii) each trial has only two possible outcomes, which we refer to as "success" (A) and "failure" ($\overline{A}$) (i.e. the sample space for each trial is $S = A \cup \overline{A}$),

(iii) the probability of success $p = P(A)$ is the same for each trial (we denote by $q = 1-p = P(\overline{A})$ the probability of failure).

Trials of an experiment satisfying (i) $-$ (iii) are known as **Bernoulli trials**.
Model: Given n Bernoulli trials with probability of success p, find the probability $P(n;k)$ of exactly k ($0 \le k \le n$) successes occurring.
We have

$$\begin{aligned} P(n;k) &= C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}, \ \ k = 0,1,\ldots,n \ \text{ and} \qquad (2.1)\\ \sum_{k=0}^{n} P(n;k) &= 1. \end{aligned}$$

Pascal (Negative Binomial) Model

Model: Consider an infinite sequence of Bernoulli trials with probability of success p (and probability of failure $q = 1-p$) in each trial. Find the probability $P(n,k)$ of the nth success occurring

after k failures ($n \in \mathbb{N},\ k \in \mathbb{N} \cup \{0\}$).
We have

$$\begin{aligned} P(n,k) &= C_{n+k-1}^{k} p^n q^k, \quad k = 0, 1, \dots \quad \text{and} \\ \sum_{k=0}^{\infty} P(n;k) &= 1. \end{aligned} \tag{2.2}$$

Geometric Model

Although a particular case for the Pascal Model (case $n = 1$), the Geometric model comes up in many applications and deserves a place of its own.
Model: Consider an infinite sequence of Bernoulli trials with probability of success p (and probability of failure $q = 1 - p$) in each trial. Find the probability p_k that the first success occurs after k failures ($k \in \mathbb{N} \cup \{0\}$).
Here, we have

$$\begin{aligned} p_k &= pq^k, \quad k = 0, 1, \dots \quad \text{and} \\ \sum_{k=0}^{\infty} p_k &= 1. \end{aligned} \tag{2.3}$$

3 Random Variables

3.1 Random Variables, PDF and CDF

Random variables, variables whose observed values are determined by chance, give a more comprehensive quantitative overlook of random phenomena. Random variables are the fundamentals of modern Statistics.

Definition 3.1. *Let $(S, \mathcal{K}, P)$ be a probability space. A **random variable** is a function $X : S \to \mathbb{R}$ satisfying the property that for every $x \in \mathbb{R}$, the event*

$$(X \le x) \ := \ \{e \in S \mid X(e) \le x\} \in \mathcal{K}. \tag{3.1}$$

- *if the set of values that it takes, $X(S)$, is at most countable in $\mathbb{R}$, then X is a **discrete random variable** (quantities that are counted);*
- *if $X(S)$ is a continuous subset of $\mathbb{R}$ (an interval), then X is a **continuous random variable** (quantities that are measured).*

For each random variable, discrete or continuous, there are two important functions associated with it:

- **PDF (probability distribution/density function)**
  - if X is discrete, then the pdf is an array

$$X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}, \tag{3.2}$$

    where $x_i \in \mathbb{R},\ i \in I$, are the values that X takes and $p_i = P(X = x_i)$
  - if X is continuous, then the pdf is a function $f : \mathbb{R} \to \mathbb{R}$;

- **CDF (cumulative distribution function)** $F = F_X : \mathbb{R} \to \mathbb{R}$, defined by

$$F(x) \quad = \quad P(X \leq x). \tag{3.3}$$

  - if X is discrete, then

$$F(x) \quad = \quad \sum_{x_i \leq x} p_i. \tag{3.4}$$

  - if X is continuous, then

$$F(x) \quad = \quad \int_{-\infty}^{x} f(t)\, dt. \tag{3.5}$$

The pdf has the following properties:

- all values $x_i, i \in I$, are distinct and listed in increasing order;
- all probabilities $p_i > 0, i \in I$ and $f(x) \geq 0$, for all $x \in \mathbb{R}$;
- $\sum\limits_{i \in I} p_i = 1$ and $\int\limits_{\mathbb{R}} f(t)dt = 1$.

The cdf has the following properties:

- if $a < b$ are real numbers, then $P(a < X \leq b) = F(b) - F(a)$;
- $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to \infty} F(x) = 1$;

- if X is discrete, then $P(X < x) = F(x-0) = \lim\limits_{y \nearrow x} F(y)$ and $P(X = x) = F(x) - F(x-0)$;
- if X is continuous, then $P(X = x) = 0$, $P(X < x) = P(X \le x) = F(x)$ and
$$P(a < X \le b) = P(a < X \le b) = P(a < X < b) = P(a \le X \le b) = \int_a^b f(t)\,dt;$$
- if X is continuous, then $F'(x) = f(x)$, for all $x \in \mathbb{R}$.

3.2 Numerical Characteristics of Random Variables

The **expectation (expected value, mean value)** of a random variable X is a real number $E(X)$ defined by

- if X is a discrete random variable with pdf $\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$,
$$E(X) = \sum_{i \in I} x_i P(X = x_i) = \sum_{i \in I} x_i p_i, \tag{3.6}$$
if it exists;
- if X is a continuous random variable with pdf $f : \mathbb{R} \to \mathbb{R}$,
$$E(X) = \int_{\mathbb{R}} x f(x) dx, \tag{3.7}$$
if it exists.

The **variance (dispersion)** of a random variable X is the number
$$V(X) = E\Big(X - E(X)\Big)^2, \tag{3.8}$$
if it exists.
The **standard deviation** of a random variable X is the number
$$\sigma(X) = \text{Std}(X) = \sqrt{V(X)}. \tag{3.9}$$

Properties:

- $E(aX + b) = aE(X) + b$, for all $a, b \in \mathbb{R}$;

- $E(X+Y) = E(X) + E(Y)$;
- If X and Y are independent, then $E(X \cdot Y) = E(X)E(Y)$;
- If $X(e) \leq Y(e)$ for all $e \in S$, then $E(X) \leq E(Y)$;
- $V(X) = E(X^2) - E(X)^2$.
- If X and Y are independent, then $V(X+Y) = V(X) + V(Y)$.

Let X be a random variable with cdf $F : \mathbb{R} \to \mathbb{R}$ and $\alpha \in (0,1)$. A **quantile of order $\boldsymbol{\alpha}$** is a number q_α satisfying the condition $P(X < q_\alpha) \;\leq\; \alpha \;\leq\; P(X \leq q_\alpha)$, or, equivalently,

$$F(q_\alpha - 0) \;\leq\; \alpha \;\leq\; F(q_\alpha). \tag{3.10}$$

If X is continuous, then for each $\alpha \in (0,1)$, there is a *unique* quantile q_α , given by $F(q_\alpha) = \alpha$, or equivalently, $q_\alpha = F^{-1}(\alpha)$. It is the number with the property that the area to its left, under the graph of the pdf is equal to α (see Figure 1).

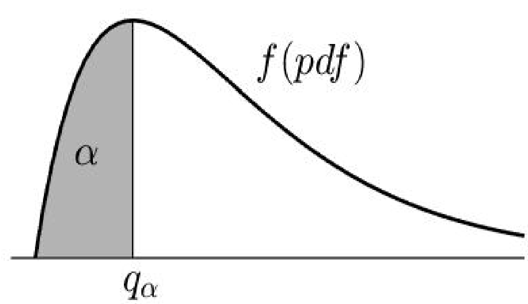

Fig. 1: Quantile q_α

Quantiles are oftenly used in various statistical procedures, such as confidence intervals, rejection regions, etc. (see Figure 2).

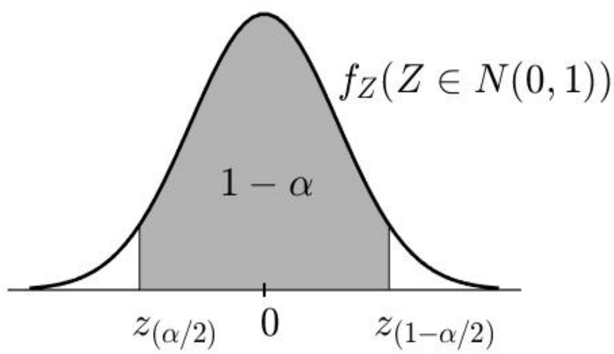

Fig. 2: Quantiles for the Normal distribution