Mathematical Statistics I

Derek Li

Contents

1 Probability and Distributions

1.1 Sets

Theorem 1.1 (Distributive Laws)**.** For any sets A, B, and C,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Theorem 1.2 (DeMorgan's Laws)**.** For any sets A and B,

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$

1.2 Probability Set Function

Definition 1.1 (Probability Set Function)**.** Let $\mathcal{S}$ be a sample space, let $\mathcal{B}$ be the set of events, P be a real-valued function defined on $\mathcal{B}$. Then P is a ***probability set function*** if P satisfies the following three conditions:

1. $P(A) \geqslant 0, \forall A \in \mathcal{B}$.
2. $P(\mathcal{S}) = 1$.
3. If $\{A_n\}$ is a sequence of events in $\mathcal{B}$ and $A_m \cap A_n = \varnothing, \forall m \neq n$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Definition 1.2. A collection of events whose members are pairwise disjoint is said to be a ***mutually exclusive collection*** and its union is refereed to as a disjoint union. The collection is said to be ***exhaustive*** if the union of its events is the sample space. A mutually exclusive and exhaustive collection of events forms a partition of $\mathcal{S}$.

Theorem 1.3. For each event $A \in \mathcal{B}, P(A) = 1 - P(A^c)$.

Proof. We have $\mathcal{S} = A \cup A^c$ and $A \cap A^c =$. Thus, $P(A) + P(A^c) = 1$. □

Theorem 1.4. The probability of the null set is zero, i.e., $P(\varnothing) = 0$.

Proof. We have $\varnothing^c = \mathcal{S}$. Accordingly, $P(\varnothing) = 1 - P(\mathcal{S}) = 1 - 1 = 0$. □

Theorem 1.5. If A and B are events s.t. $A \subset B$, then $P(A) \leqslant P(B)$.

Proof. We have $B = A \cup (A^c \cap B)$ and $A \cap (A^c \cap B) = \varnothing$. Hence, $P(B) = P(A) + P(A^c \cap B)$. From the definition, $P(A^c \cap B) \geqslant 0$, and thus $P(B) \geqslant P(A)$. □

Theorem 1.6. For each $A \in \mathcal{B}, 0 \leqslant P(A) \leqslant 1$.

Proof. Since $\varnothing \subset A \subset \mathcal{S}$, we have $P(\varnothing) \leqslant P(A) \leqslant P(\mathcal{S})$ or $0 \leqslant P(A) \leqslant 1$. □

Theorem 1.7. If A and B are events in $\mathcal{S}$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. We can represent $A \cup B$ and B as a union of non-intersecting sets: $A \cup B = A \cup (A^c \cap B)$ and $B = (A \cap B) \cup (A^c \cap B)$. Hence, $P(A \cup B) = P(A) + P(A^c \cap B)$ and $P(B) = P(A \cap B) + P(A^c \cap B)$. Therefore, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. □

Definition 1.3 (Equiprobability)**.** Let $\mathcal{S} = \{x_1, \cdots, x_m\}$ be a finite sample space. Let $p_i = \dfrac{1}{m}$ for all $i = 1, \cdots, m$. For all subsets A of $\mathcal{S}$ define

$$P(A) = \sum_{x_i \in A} \frac{1}{m} = \frac{\#(A)}{m}$$

where $\#(A)$ denotes the number of elements in A. Then P is a probability on $\mathcal{S}$ and it is refereed to as the equilikely case.

1.2.1 Counting Rules

Definition 1.4 (Permutation)**.** The number of k ***permutations*** taken from a set of n elements is

$$P_k^n = \frac{n!}{(n-k)!}$$

Definition 1.5 (Combination)**.**

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

which is also referred to a ***binomial coefficient***.

1.2.2 Additional Properties of Probability

Theorem 1.8. Let $\{C_n\}$ be a non-decreasing sequence of events, then

$$\lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = P\left(\bigcup_{n=1}^{\infty} C_n\right)$$

Let $\{C_n\}$ be a decreasing sequence of events, then

$$\lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = P\left(\bigcap_{n=1}^{\infty} C_n\right)$$

Theorem 1.9 (Boole's Inequality)**.** Let $\{C_n\}$ be an arbitrary sequence of events, then

$$P\left(\bigcup_{n=1}^{\infty} C_n\right) \leqslant \sum_{n=1}^{\infty} P(C_n)$$

1.3 Conditional Probability and Independence

Definition 1.6 (Conditional Probability)**.** Let B and A be events with $P(A) > 0$, then we defined the conditional probability of B given A as

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

Property 1.1. We have:

1. $P(B|A) \geqslant 0$.
2. $P(A|A) = 1$.
3. $$P\left(\bigcup_{n=1}^{\infty} B_n|A\right) = \sum_{n=1}^{\infty} P(B_n|A)$$

provided that $B_1, \ldots, B_n$ are mutually exclusive events.

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

Theorem 1.10 (Bayes' Theorem)**.** Let $A_1, \cdots, A_k$ be events s.t. $P(A_i) > 0, i = 1, \cdots, k$. Assume that $A_1, \cdots, A_k$ form a partition of the sample space $\mathcal{S}$. Let B be any event. Then

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum\limits_{i=1}^{k} P(A_i)P(B|A_i)}$$

Definition 1.7 (Independence)**.** We say A and B are ***independent*** if when $P(A) > 0, P(B|A) = P(B)$, i.e., the occurrence of A does not change the probability of B; or when $P(A \cap B) = P(A)P(B)$.

Property 1.2. Suppose A and B are independent, then the following three pairs are independent: A^c and B, A and B^c, and A^c and B^c.

Proof. We have

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = [1 - P(A)]P(B) = P(A^c)P(B)$$

□

Definition 1.8 (Mutually Independence)**.** The events are ***mutually independent*** iff they are pairwise independent.

Example 1.1. We say A_1, A_2, and A_3 are mutually independent iff

$$\begin{aligned} P(A_1 \cap A_3) &= P(A_1)P(A_3) \\ P(A_1 \cap A_2) &= P(A_1)P(A_2) \\ P(A_2 \cap A_3) &= P(A_2)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3) \end{aligned}$$

1.4 Random Variables

Definition 1.9. Consider a random experiment with a sample space $\mathcal{S}$. A function X, which assigns to each element $s \in \mathcal{S}$ one and only one number $X(s) = x$, is called a ***random variable***. The ***space/range*** of X is the set of real numbers $\mathcal{D} = \{x : x = X(s), s \in \mathcal{S}\}$.

Definition 1.10 (Cumulative Distribution Function (CDF))**.** Let X be a r.v., then its ***cumulative distribution function*** (CDF) is defined as

$$F_X(x) = P_X((-\infty, x]) = P(\{s \in \mathcal{S} : X(s) \leqslant x\}) = P(X \leqslant x)$$

Definition 1.11 (Equal in Distribution)**.** Let X and Y be two r.v.s., then X and Y are ***equal in distribution*** iff $F_X(x) = F_Y(x), \forall x \in \mathbb{R}$, denoted $X \overset{D}{=} Y$.

Note. While X and Y may be equal in distribution, they may be quite different.

Theorem 1.11. Let X be a r.v. with CDF $F(x)$. Then

1. $\forall a, b$, if $a < b$, then $F(a) \leqslant F(b)$.
2. $\lim\limits_{x \to -\infty} F(x) = 0$.
3. $\lim\limits_{x \to \infty} F(x) = 1$.
4. F is right continuous: $\lim\limits_{x \downarrow x_0} F(x) = F(x_0)$.

Theorem 1.12. Let X be a r.v. with CDF F_X. Then for $a < b$, $P(a < X \leqslant b) = F_X(b) - F_X(a)$.

Theorem 1.13. For any r.v., $P(X = x) = F_X(x) - F_X(x-), \forall x \in \mathbb{R}$, where $F_X(x-) = \lim_{z \uparrow x} F_X(z)$.

Proof. For any $x \in \mathbb{R}$, we have

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right]$$

i.e., $\{x\}$ is the limit of a decreasing sequence of sets. Hence, by theorem,

$$\begin{aligned} P(X = x) &= P\left(\bigcap_{n=1}^{\infty} \left\{x - \frac{1}{n} < X \leqslant x\right\}\right) \\ &= \lim_{n \to \infty} P\left(x - \frac{1}{n} < X \leqslant x\right) \\ &= \lim_{n \to \infty} [F_X(x) - F_X(x - 1/n)] \\ &= F_X(x) - F_X(x-) \end{aligned}$$

□

1.5 Discrete Random Variable

Definition 1.12 (Discrete Random Variable)**.** We say a r.v. is a ***discrete random variable*** is its space is either finite or countable.

Definition 1.13 (Probability Mass Function (PMF))**.** Let X be a discrete r.v. with space $\mathcal{D}$. The ***probability mass function*** (PMF) of X is

$$p_X(x) = P(X = x), x \in \mathcal{D}$$

which satisfies the two properties:

1. $0 \leqslant p_X(x) \leqslant 1, x \in \mathcal{D}$.
2. $\sum_{x \in \mathcal{D}} p_X(x) = 1$.

1.5.1 Transformation

Assume X is discrete with space $\mathcal{D}_X$ and $Y = g(X)$, then the space of Y is $\mathcal{D}_Y = \{g(x) : x \in \mathcal{D}_X\}$. If g is one-to-one, then the PMF of Y is

$$p_Y(y) = P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y)) = p_X(g^{-1}(y))$$

1.6 Continuous Random Variable