Theory of Statistical Practice

Derek Li

Contents

1 Probability Review

1.1 Basic Definition

Definition 1.1. ***Random experiment*** is a mechanism producing an outcome (result) perceived as random or uncertain.

Definition 1.2. ***Sample space*** is a set of all possible outcomes of the experiment:

$$\mathcal{S} = \{\omega_1, \omega_2, \cdots\}.$$

Example 1.1. Waiting time until the next bus arrives: $\mathcal{S} = \{t : t \geqslant 0\}$.

1.2 Probability Function/Measure

Definition 1.3. Given a sample space $\mathcal{S}$, define $\mathcal{A}$ to be a collection of subsets (events) of $\mathcal{S}$ satisfying the following conditions:

1. $\mathcal{S} \in \mathcal{A}$;
2. $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$;
3. $A_1, A_2, \cdots \in \mathcal{A} \Rightarrow A_1 \cup A_2 \cup \cdots \in \mathcal{A}$.

If $\mathcal{S}$ is finite or countably infinite, then $\mathcal{A}$ could consist of all subsets of $\mathcal{S}$ including $\varnothing$.

Definition 1.4. The ***probability function*** (***measure***) P on $\mathcal{A}$ satisfies the following conditions:

1. $P(A) \geqslant 0, \forall A \in \mathcal{A}$;
2. $P(\varnothing) = 0$ and $P(\mathcal{S}) = 1$;
3. If $A_1, A_2, \cdots$ are disjoint (mutually exclusive) events, i.e., $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Property 1.1. $P(A^C) = 1 - P(A)$.

Proof. $1 = P(\mathcal{S}) = P(A \cup A^C) = P(A) + P(A^C)$. □

Property 1.2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. $P(A) = P(A \cap B) + P(A \cap B^C)$ and $P(A \cup B) = P(B) + P(A \cap B^C)$. □

Corollary 1.1. $P(A \cup B) \leqslant P(A) + P(B)$.

Property 1.3. In general,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots - (-1)^n P(A_1 \cap \cdots \cap A_n).$$

Property 1.4 (Bonferroni's Inequality)**.** In general,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} P(A_i).$$

1.3 Conditional Probability

Definition 1.5. The probability of A ***conditional*** on B is defined by

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

if $P(B) > 0$. Note that if $P(B) = 0$, we can still define $P(A|B)$ but we need to be more careful mathematically.

Theorem 1.1 (Bayes Theorem)**.** If $B_1, \cdots, B_k$ are disjoint events with $B_1 \cup \cdots \cup B_k = \mathcal{S}$, then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

1.4 Independence

Definition 1.6. Two events A and B are ***independent*** if

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

When $P(A), P(B) > 0$, we can also say

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B).$$

Events $A_1, \cdots, A_k$ are independent if

$$P\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i).$$

1.5 Interpretation of Probability

- Long-Run frequencies: If we repeat the experiment many times, then $P(A)$ is the proportion of times the event A occurs.
- Degrees of belief (subjective probability): If $P(A) > P(B)$, then we believe that A is more likely to occur than B.
- Frequentist versus Bayesian statistical methods:
  - Frequentists: Pretend that an experiment is at least conceptually repeatable.
  - Bayesians: Use subjective probability to describe uncertainty in parameters and data.

1.6 Random Variable

Definition 1.7. ***Random variable*** is a real-valued function defined on a sample space $\mathcal{S}, X : \mathcal{S} \to \mathbb{R}$. In other words, for each outcome $\omega \in \mathcal{S}, X(\omega)$ is a real number.

Definition 1.8. The ***probability distribution*** of X depends on the probabilities assigned to the outcomes in $\mathcal{S}$.

Definition 1.9. The ***cumulative distribution function*** (CDF) of X is

$$F(x) = P(X \leqslant x) = P(\omega \in \mathcal{S} : X(\omega) \leqslant x).$$

We denote it $X \sim F$.

Property 1.5. CDF satisfies:
1. If $x_1 \leqslant x_2$, then $F(x_1) \leqslant F(x_2)$;
2. $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$;
3. F is right-continuous with left-hand limits:

$$\lim_{y \to x^+} F(y) = F(x), \lim_{y \to x^-} F(y) = F(x-) = P(X < x);$$

4. $P(X = x) = F(x) - F(x-)$.

Definition 1.10. If $X \sim F$ where F is a continuous function, then X is a ***continuous r.v.***, and we can typically find a ***probability density function*** (PDF) f s.t.

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

Definition 1.11. If X takes only a finite or countably infinite number of possible values, then X is a ***discrete r.v.***, and F is a step function. We can define its ***probability mass function*** (PMF) by

$$f(x) = F(x) - F(x-) = P(X = x).$$

1.7 Expected Value

Definition 1.12. Suppose X with PDF $f(x)$ and Y with PMF $f(y)$. We can define the ***expected value*** of $h(X)$ and $h(Y)$ by

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)\mathrm{d}x \text{ and } \mathbb{E}[h(Y)] = \sum_{y} h(y)f(y).$$

We can also write $h(x) = h^+(x) - h^-(x)$ where $h^+(x) = \max\{h(x), 0\}$ and $h^-(x) = \max\{-h(x), 0\}$, then $\mathbb{E}[h(X)] = \mathbb{E}[h^+(X)] - \mathbb{E}[h^-(X)]$:
1. If $\mathbb{E}[h^+(X)]$ and $\mathbb{E}[h^-(X)]$ are finite, then $\mathbb{E}[h(X)]$ is well defined.
2. If $\mathbb{E}[h^+(X)] = \infty$ and $\mathbb{E}[h^-(X)]$ is finite, then $\mathbb{E}[h(X)] = \infty$.
3. If $\mathbb{E}[h^+(X)]$ is finite and $\mathbb{E}[h^-(X)] = \infty$, then $\mathbb{E}[h(X)] = -\infty$.
4. If $\mathbb{E}[h^+(X)]$ and $\mathbb{E}[h^-(X)]$ are infinite, then $\mathbb{E}[h(X)]$ does not exist.

Example 1.2 (Expected Values of Cauchy Distribution)**.** X is a continuous r.v. with

$$f(x) = \frac{1}{\pi(1 + x^2)}, -\infty < x < \infty.$$

We have

$$\mathbb{E}[X^+] = \mathbb{E}[X^-] = \int_0^{\infty} \frac{x}{\pi(1 + x^2)}\mathrm{d}x = \lim_{x \to \infty} \frac{1}{2\pi} \ln(1 + x^2) = +\infty.$$

Thus, $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$ does not exist.

1.8 Independent Random Variable

Definition 1.13. R.v.s. $X_1, X_2, \cdots$ are independent if the events $[X_1 \in A_1], [X_2 \in A_2], \cdots$ are independent events for any $A_1, A_2, \cdots$.

If $X_1, \cdots, X_n$ are independent r.v.s. with PDF or PMF $f_1, \cdots, f_n$, then the joint PDF or PMF of $(X_1, \cdots, X_n)$ is

$$f(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_i(x_i).$$

Suppose $X_1, \cdots, X_n$ are independent r.v.s. with mean $\mu_1, \cdots, \mu_n$ and variance $\sigma_1^2, \cdots, \sigma_n^2$. Define $S = X_1 + \cdots + X_n$, then $\mathbb{E}[S] = \mu_1 + \cdots + \mu_n$ (which is true even if $X_1, \cdots, X_n$ are not independent) and $\mathrm{Var}[S] = \sigma_1^2 + \cdots + \sigma_n^2$.