Lecture Notes 1: Probability

Professor: Zhihua Zhang

1 Probability Theory Basics

1.1 Sample Space and Events

Definition 1.1 *The sample space Ω is the set of possible outcomes of an experiment, $\omega \in \Omega$ are called sample outcomes, realizations or elements. The subsets of Ω are called events.*

Definition 1.2 *Given an event, $A \subset \Omega$, let $A^c = \{\omega \in \Omega, \omega \notin A\}$ denote the complement of A.*

Definition 1.3 *A sequence of sets $A_1, A_2, \cdots$ is* monotone increasing*, if $A_1 \subset A_2 \subset \cdots$, we define $\lim\limits_{n\to\infty} A_n = \bigcup\limits_{i=1}^{\infty} A_i$.*

Definition 1.4 *A sequence of sets $A_1, A_2, \cdots$ is* monotone decreasing*, if $A_1 \supset A_2 \supset \cdots$, we define $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{i=1}^{\infty} A_i$.*

Example 1.1 *Let $\Omega = \mathbf{R}$ and $A_i = [0, 1/i)$ for $i = 1, 2 \cdots$,then*
$\bigcup\limits_{i=1}^{\infty} A_i = [0,1), \bigcap\limits_{i=1}^{\infty} A_i = \{0\}$. If $A_i = (0, 1/i)$, then $\bigcup\limits_{i=1}^{\infty} A_i = (0,1), \bigcap\limits_{i=1}^{\infty} A_i = \emptyset$.

1.2 σ-field and Measures

Definition 1.5 *Let $\mathcal{A}$ be a collection of subsets of a sample space Ω. $\mathcal{A}$ is called σ-field (or σ-algebra). iff*

1. *The empty set $\emptyset \in \mathcal{A}$.*
2. *If $A \in \mathcal{A}$, $A^c \in \mathcal{A}$.*
3. *If $A_i \in \mathcal{A}$, $i \in \{1, 2, ..., k\}$, then $\bigcup\limits_{i=1}^{k} A_i \in \mathcal{A}$.*

Definition 1.6 *A pair $(\Omega, \mathcal{A})$ is called a measurable space.*

Example 1.2 *Let A be a nonempty proper subset of Ω, i.e. $A \neq \emptyset$, $A \neq \Omega$, the smallest $\mathcal{A} = \{\emptyset, \Omega, A, A^c\}$.*

Example 1.3 $\Omega = \mathbb{R}$. *The smallest σ-field that contains all the finite open sets of $\mathbb{R}$ is called Borel σ-field.*

Definition 1.7 *Let $(\Omega, \mathcal{A})$ be a measurable space. A set function ν defined on $\mathcal{A}$ is called a measure iff*

1. $0 \leq \nu(A) \leq \infty$ *for any* $A \in \mathcal{A}$.
2. $\nu(\emptyset) = 0$.
3. *If $A \in \mathcal{A}$, and A_i are disjoint, i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$, then* $\nu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \nu(A_i)$.

Definition 1.8 *Tripe $(\Omega, \mathcal{A}, \nu)$ is called a measure space.*

If $\nu(\Omega) = 1$, then ν is called a probability measure and denote it by P. $(\Omega, \mathcal{A}, P)$ is called a probability space.

Example 1.4 *Let Ω be a sample space, $\mathcal{A}$ is a collection of all subsets, and $\nu(A)$ is the number of elements in A.*

Lemma 1.1 *For any two events A and B. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Theorem 1.1 (Continuity of Probability) *If $A_n \to A$, then*

$$P(A_n) \to P(A) \text{ as } n \to \infty.$$

Proof: We first consider the case where A_n is monotone increasing.
Recall that $A_1 \subset A_2 \ldots$ and let $A = \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$.
Define $B_1 = A_1$, $B_2 = \{\omega \in \Omega : \omega \in A_2, \omega \notin A_1\}$, $B_3 = \{\omega \in \Omega : \omega \in A_3, \omega \notin A_2\} \ldots$. Then for each n, we have $A_n = \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$.
Thus, $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$. So that,

$$P(A_n) = \sum_{i=1}^{n} P(B_i)$$

Hence, we have,

$$\begin{aligned}
\lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i) \\
&= P(\bigcup_{i=1}^{\infty} B_i) = P(A)
\end{aligned}$$

For arbitrary sequence $\{A_i\}$, we can define $\{C_i\}$ to construct a monotone increasing sequence. Specifically, $C_1 = A_1 \cap A$, $C_2 = (A_1 \cup A_2) \cap A$, $C_3 = (A_1 \cup A_2 \cup A_3) \cap A, \ldots$

1.3 Independent Events

Definition 1.9 *Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.*

We write $A \perp\!\!\!\perp B$ to denote independence. For a set of events $\{A_i, i \in I\}$ A, it is independent if $P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$, for every finite subset J of I.

1.4 Conditional Probability

Definition 1.10 *If $P(B) > 0$, the conditional probability of A given B is*

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

Lemma 1.2 *If A and B are independent events, then $P(A|B) = P(A)$. Also, for any events A, B*

$$P(AB) = P(A|B)P(B) = P(B|A)P(A).$$

1.5 Bayes Theorem

Theorem 1.2 (The Law of Total Probability) *Let $A_1, A_2, \ldots, A_k$ be partition of Ω. Then for any event B,*

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

Proof: Define $C_j = B \cap A_j$ for $j = 1, \ldots, k$. Then we have $C_j \cap C_i = \emptyset$ and $B = \bigcup_{i=1}^{k} C_i$. Thus,

$$P(B) = \sum P(C_j) = \sum P(B \cap A_j) = \sum P(B|A_j)P(A_j)$$

Theorem 1.3 (Bayes Theorem) *Let $A_1, \ldots, A_k$ be a partition of Ω, such that $P(A_i) > 0$ for each i. If $P(B) > 0$, then for each $i = 1, \ldots, k$*

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

Remarks: We usually call those probabilities as

- $P(A_i)$ - prior probability of A_i
- $P(A_i|B)$ - posterior probability of A_i
- $P(B|A_i)$ - likelihood