Summary for Probability Theory

Jiaxin Hu

July 6, 2020

1 Preliminary

- *DeMorgan's Laws:* Let $\{A_i\}_{i=1}^{\infty}$ be a collection of set. Then, $(\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c$ and $(\bigcap_{i=1}^{\infty} A_i)^c = \bigcup_{i=1}^{\infty} A_i^c$.
- *Some set operations:* Suppose A, B are two sets. (1) $A - B = A \cap B^c$; (2) $\bigcup_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for some } i\}$, $\bigcap_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for all } i\}$.

2 Single variable

2.1 Probability and conditional probability

Definition 1 (*Sample space*)**.** The set S containing all possible outcomes is called the sample space.

Definition 2 (*σ-field*)**.** A collection $\mathcal{F}$ of subsets of a sample space S is called a σ-field (or σ-algebra) if and only if (**iff**) it has the following properties:
(1) The empty set $\varnothing \in \mathcal{F}$;
(2) If $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$;
(3) If $A_i \in \mathcal{F}, i = 1, 2, ...$, then their union $\cup_i A_i \in \mathcal{F}$.
If $A \in \mathcal{F}$, then A is called an *event*.

Definition 3 (*Measure and probability*)**.** A set function v defined on a σ-field $\mathcal{F}$ is called a measure **iff** it has the following properties:
(1) $0 \leq v(A) \leq \infty$ for any $A \in \mathcal{F}$;
(2) $v(\varnothing) = 0$;
(3) If $A_i \in \mathcal{F}, i = 1, 2, ...$ and A_i's are disjoint, i.e. $A_i \cap A_j = \varnothing, \forall i \neq j$, then

$$v(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} v(A_i).$$

If $v(\mathcal{F}) = 1$, then v is a probability defined on $\mathcal{F}$ and we use notation P instead of v.

Theorem 2.1 (Probability)**.** *Let the sample space $S = \{s_1, s_2, ...\}$ and $\mathcal{F}$ be all subsets of S. Let $p_1, p_2, ...$ be non-negative numbers that $\sum_i p_i = 1$. The following defines a probability on $\mathcal{F}$*

$$P(A) = \sum_{i: s_i \in A} p_i, \quad A \in \mathcal{F}.$$

Theorem 2.2 (Properties of probability). *Let P be a probability, A, B be events and $\{A_i\}_{i=1}^{\infty}$ be a collection of event. Let $\{C_i\}_{i=1}^{\infty}$ be a partition of sample space S, i.e. $C_i \cap C_j, \forall i \neq j$ and $\bigcup_{i=1}^{\infty} C_i = S$. Then,*

1.$P(A) \leq 1; P(A^c) = 1 - P(A); P(A) = P(A \cap B) + P(A \cap B^c)$;
2.*If $A \subset B$, then $P(A) \leq P(B)$;*
3.*(General addition formula)*

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

4.$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$;

5.*(Boole's inequality)* $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$;

6.*(Bonferroni's inequality)* $P(\bigcap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$.

Definition 4 (*Conditional Probability*). If A and B are events with $P(B) > 0$, then the conditional probability of A given B is

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

For convenience, we define $P(A|B) = 0$ when $P(B) = 0$.

Theorem 2.3 (Useful formulas for conditional probability). *Let $A, \{A_i\}_{i=1}^{\infty}, B, \{B_i\}_{i=1}^{\infty}, C$ be events. Then we have:*

$$1.P(A|B) = \frac{P(A)P(B|A)}{P(B)};$$
$$2.P(A^c|B) = 1 - P(A|B); P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B);$$
$$3.P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1) \cdots P(A_n| \bigcap_{i=1}^{n-1} A_i);$$
$$4.\text{If } \{B_i\}_{i=1}^{\infty} \text{is a partition of } S, P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i).$$

Theorem 2.4 (Bayes formula). *Let A be an event and $\{B_i\}_{i=1}^{\infty}$ be a partition of S. Then,*

$$P(B_i|A) = \frac{P(A|B_i)P(A)}{\sum_{j=1}^{\infty} P(A|B_i)P(B_i)}.$$

Definition 5 (*Independence*). Two events A, B are independent **iff**

$$P(A \cap B) = P(A)P(B) \quad \text{or} \quad P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B).$$

If A, B are independent, then the following pairs are also independent: A and B^c, A^c and B, A^c and B^c.

Definition 6 (*Mutual and pairwise independence*). A collection of events $A_1, ..., A_n$ are mutually independent **iff** for any sub-collection $A_{i_1}, ..., A_{i_k}$,

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_K}).$$

The events $A_1, ..., A_n$ are pairwise independent **iff** A_i and A_j are independent for all $i \neq j$. Mutual independence is stronger than pairwise independence.

Definition 7 (*Conditional independence*). Events A and B are conditionally independent given event C **iff**

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

Let A, B, C are events. Independence does not imply conditional independence:

$$P(A \cap B) = P(A)P(B) \nRightarrow P(A \cap B|C) = P(A|C)P(B|C).$$

Conditional independence does not imply independence:

$$P(A \cap B|C) = P(A|C)P(B|C) \nRightarrow P(A \cap B) = P(A)P(B).$$

Mutual independence implies conditional independence:

$$A, B, C \text{ mutually independent} \nRightarrow P(A \cap B|C) = P(A|C)P(B|C).$$

2.2 Random variable and distribution

Definition 8 (*(Random variable and distribution)*). A random variable X is a function from S to $\mathbb{R}$ such that, for any Borel set $\mathcal{B} \subset \mathbb{R}$,

$$\{X \in \mathcal{B}\} = \{\omega \in S : X(\omega) \in \mathcal{B}\}.$$

The induced probability of X is

$$P_X(\mathcal{B}) = P(X \in \mathcal{B}) = P(\omega \in \{\omega \in S : X(\omega) \in \mathcal{B}\}).$$

The probability P_X is called the distribution of X.

Definition 9 (*Cumulative distribution function(cdf)*). The cdf of a random variable X, denoted by $F_X(x)$, is defined as

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Theorem 2.5 (Cdf). *The function $F(x)$ is a cdf **iff***

$$1. \lim_{x \to -\infty} F(x) = 0 \textit{ and } \lim_{x \to \infty} F(x) = 1;$$
$$2. F(x) \textit{ is non-decreasing in } x;$$
$$3. F(x) \textit{ is right-continuous}: \lim_{\epsilon > 0, \epsilon \to 0} F(x + \epsilon) = F(x), \quad \forall x \in \mathbb{R}.$$

Definition 10 (*Continuity of random variable*). A random variable X is continuous if $F_X(x)$ is continuous in x. A random variable x is discrete if $F_X(x)$ is a step function of x.

Note that the continuity of a random variable depends on the cdf rather than pdf or pmf. There are random variable's that are mixtures of these two types.

Definition 11 (*Probability mass function(pmf)*)**.** The pmf of a discrete random variable X is

$$f_X(x) = P(X = x), \quad x \in \mathbb{R}.$$

The cdf of X, $F_X(x) = P(X \le x) = \sum_{k \le x} f_X(k)$.

Definition 12 (*Probability density function(pdf)*)**.** The pdf of a continuous random variable X is the function $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad x \in \mathbb{R},$$

if $f_X(x)$ exists. The continuous random variable X has a pdf **iff** F_X is absolutely continuous. If f is a pdf, the set $\{x : f(x) > 0\}$ is called its support.

If F_X is differentiable, then $f_X(x) = \frac{d}{dx}F_X(x)$.

Theorem 2.6 (Pdf)**.** *A function $f(x)$ is a pdf* ***iff:***

$$1. f(x) \ge 0, \quad \forall x \in \mathbb{R};$$
$$2. \int_{-\infty}^{\infty} f(x)dx = 1.$$

How to find pdf given cdf? (1) $f_X(x) = F'_X(x)$ for x at which F_X is differentiable; (2) $f_X(x)$ can be any $c \ge 0$ for x at which F_X is not differentiable.

2.3 Transformation

Let X be a random variable and $Y = g(X)$, where g is function $\mathbb{R} \mapsto \mathcal{Y}$ and $\mathcal{Y}$ is the domain of Y. For any $A \in \mathcal{Y}$,

$$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A)), \quad \text{where } g^{-1}(A) = \{x : g(x) \in A\}.$$

Given F_X or f_X, we want to obtain $f_Y(y)$. If X is discrete, then

$$f_Y(y) = \sum_{x \in g^{-1}(\{y\})} P(X = x)$$

If f_X is continuous and g is a continuously differentiable monotone function, then

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| & y \in \mathcal{Y} \\ 0 & \text{otherwise} \end{cases}$$

Theorem 2.7 (Transformation for continuous random variable)**.** *Let X be a continuous random variable with pdf f_X. Suppose there are disjoint $\{A_i\}_{i=1}^k$ such that $P(X \in \bigcup_{i=1}^k A_i) = 1$, and f_X is continuous on each $A_i, i \in [k]$. There are functions $g_1(x), ..., g_k(x)$ defined on $A_i, i \in [k]$ respectively, satisfying*
1. $g(x) = g_i(x), \forall x \in A_i$;
2. $g_i(x)$ is strictly monotone on A_i;
3. The set $\mathcal{Y} = \{y : y = g_i(x)$ for some $x \in A_i\}$ is the same for each i;
4. $g_i^{-t}(y)$ has a continuous derivative on $\mathcal{Y}$ for each i.
Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y))|\frac{d}{dy}g_i^{-1}(y)| & y \in \mathcal{Y} \\ 0 & \textit{otherwise} \end{cases}$$

Example 1. Suppose a random variable $X \sim N(0,1)$. Obtain the distribution of $Y = |X|$.

Proof. Let $A_1 = (-\infty, 0)$, $A_2 = (0, +\infty)$ and $\mathcal{Y} = (0, +\infty)$. On A_1, $g_1(x) = x$ and $g_1^{-1}(x) = x$. On A_2, $g_2(x) = -x$ and $g_2^{-1}(x) = -x$.
By theorem 2.7,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \sqrt{\frac{2}{\pi}} e^{-x^2/2}.$$

□

2.4 Expectation

Definition 13 (*Expectation*)**.** The expected value or mean of a random variable $g(x)$ is

$$\mathbb{E}[g(x)] = \begin{cases} \int_{-\infty}^{+\infty} g(x) f_X(x) dx \xlongequal{Y=g(x)} \int_{-\infty}^{+\infty} y f_Y(y) dx & \text{if } X \text{has pdf } f_X \\ \sum_x g(x) f_X(x) \xlongequal{Y=g(x)} \sum_x y f_Y(y) & \text{if } X \text{has pmf } f_X \end{cases}$$

provided that $\mathbb{E}[|g(x)|] < \infty$. Otherwise, the expected value of $g(X)$ does not exist.

Theorem 2.8 (Properties of expectation)**.** *Let X, Y be random variables whose expectations exist. Let a, b, c be constants.*
1. $\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c$;
2. If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

Theorem 2.9 (Relationship between expectation and cdf)**.** *Let F be the cdf of a random variable X. If X has a pdf or pmf, then,*

$$\mathbb{E}[|X|] = \int_0^{+\infty} [1 - F(x)] dx + \int_{-\infty}^0 F(x) dx,$$

*and $\mathbb{E}[|X|] < +\infty$ **iff** both integrals are finite. In case where $\mathbb{E}[|X|] < +\infty$,*

$$\mathbb{E}[X] = \int_0^{+\infty} [1 - F(x)] dx - \int_{-\infty}^0 F(x) dx.$$

Proof. Without the loss of generality, suppose random variable X has a pdf. Let f be the pdf of X. We have

$$\begin{aligned} \mathbb{E}[|X|] &= \int_0^{+\infty} x f(x) dx - \int_{-\infty}^0 x f(x) dx \\ &= \int_0^{+\infty} \int_0^x f(x) dt dx + \int_{-\infty}^0 \int_x^0 f(x) dt dx \\ &= \int_0^{+\infty} \int_t^{+\infty} f(x) dx dt + \int_{-\infty}^0 \int_{-\infty}^t f(x) dx dt \\ &= \int_0^{+\infty} [1 - F(t)] dt + \int_{-\infty}^0 F(t) dt. \end{aligned}$$

Therefore, $\mathbb{E}[|X|] < +\infty$ **iff** the two integrals are finite. Simiarly,

$$\begin{aligned} \mathbb{E}[X] &= \int_0^{+\infty} x f(x) dx + \int_{-\infty}^0 x f(x) dx \\ &= \int_0^{+\infty} [1 - F(t)] dt - \int_{-\infty}^0 F(t) dt. \end{aligned}$$

□

Corollary 1 (Relationship between expectation and cdf)**.** Let F be the cdf of a random variable X. If X has a pdf or pmf, then,

$$\mathbb{E}[|X|] = \int_0^{+\infty} P(X > x) + P(-X \leq x)dx,$$

and

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq \mathbb{E}[|X|] \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

Proof. By the proof of theorem 2.9,

$$\begin{aligned}
\mathbb{E}[|X|] &= \int_0^{+\infty} [1 - F(t)]dt + \int_{-\infty}^{0} F(t)dt \\
&= \int_0^{+\infty} [1 - F(t)]dt + \int_0^{+\infty} F(-t)dt \\
&= \int_0^{+\infty} P(X > t) + P(-X \leq t)dt
\end{aligned}$$

To show $\sum_{n=1}^{\infty} P(|X| \geq n) \leq \mathbb{E}[|X|]$, we have

$$\begin{aligned}
\mathbb{E}[|X|] &= \int_0^{+\infty} P(X > t) + P(-X \leq t)dt \\
&\geq \int_0^{+\infty} P(|X| > t)dt \\
&= \sum_{n=0}^{+\infty} \int_n^{(n+1)} P(|X| > t)dt \\
&\geq \sum_{n=0}^{+\infty} \int_n^{(n+1)} P(|X| \geq (n+1))dt \\
&= \sum_{n=1}^{+\infty} P(|X| \geq n)
\end{aligned}$$

To show $\mathbb{E}[|X|] \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n)$, we have

$$\begin{aligned}
\mathbb{E}[|X|] &= \int_0^{+\infty} P(X > t) + P(-X \leq t)dt \\
&\leq \int_0^{+\infty} P(|X| \geq t)dt \\
&= \sum_{n=0}^{+\infty} \int_n^{(n+1)} P(|X| \geq t)dt \\
&\leq \sum_{n=0}^{+\infty} \int_n^{(n+1)} P(|X| \geq n)dt \\
&\leq 1 + \sum_{n=1}^{+\infty} P(|X| \geq n).
\end{aligned}$$

$\square$