Statistics Inference I

Probability Theory, Lecture 2

Haoming Wang

25 June 2020

This is the Lecture note for the *Mathematical Statistics*. The reference materials is *Statistics Inference second edition, George Casella, Roger L. Berger*. The course covers first 5 chapters of the book: Probability Theory, Transformations and Expectations, Common Families of Distributions, Multiple Random Variables, Properties of a Random Sample.

Content:

Properties for probability measure

Theorem 1. *Suppose $\mathcal{F}$ is a σ algebra on S, if $A_i(i \in I) \in \mathcal{F}$, then $\cap_{i\in I} A_i \in \mathcal{F}$.*

Proof. $A_i(i \in I) \in \mathcal{F} \Rightarrow A_i^c(i \in I) \in \mathcal{F} \Rightarrow \cup_{i\in I} A_i^c \in \mathcal{F} \Rightarrow \left(\cup_{i\in I} A_i^c\right)^c = \cap_{i\in I} A_i \in \mathcal{F}$. □

Definition 1 (Borel field)**.** We call the σ algebra generated by all open intervals on $\mathbb{R}$ the Borel field, denoted as $\mathcal{B}$.

Note 1. $\mathcal{B}$ is the smallest σ algebra that contains all open intervals on $\mathbb{R}$.

As we introduced last lecture, the close interval can be created by the intersection of countable open interval, and the intersection of the elements in a σ algebra is still in it. Thus the close intervals are also contained in $\mathcal{B}$. Actually, any intervals on $\mathbb{R}$ are in $\mathcal{B}$.

The union, intersection, difference of the countable intervals on $\mathbb{R}$ is measurable set on which probability need to be defined. It is the reason why we need $\mathcal{B}$.

Theorem 2. *Given a probability space $(S, \mathcal{F}, \mathbb{P})$, if A is an event ($A \in \mathcal{F}$), then $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.*

Proof. Since $A^c \cap A = \emptyset, A \cup A^c = S$, thus $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(S) = 1$. □

Furthermore, $\mathbb{P}(A^c) \geq 0$ by definition, thus $1 \leq \mathbb{P}(A) = 1 - \mathbb{P}(A^c) \leq 1$.

Theorem 3. *Given probability space $(S, \mathcal{F}, \mathbb{P})$, for $\forall A, B \in \mathcal{F}$:*

1. $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
3. *if $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

Proof.

1. Since $B = (B \cap A) \cup (B \cap A^c)$ and $(B \cap A) \cap (B \cap A^c) = \emptyset$, thus $\mathbb{P}(B) = \mathbb{P}(B \cap A^c) + \mathbb{P}(A \cap B)$;
2. Similarly, $A \cup B = (A \cap B^c) \cup (B \cap A^c) \cup (A \cap B)$ and $(A \cap B^c), (B \cap A^c), (A \cap B)$ are pairwise disjoin. thus

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(B \cap A^c) + \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(B \cap A) + \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).
\end{aligned}$$

3. if $A \subseteq B$, then $B \cap A^c \neq \emptyset$, thus $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A)$.

$\square$

Theorem 4 (Bonferroni's Inequality). *Given probability space $(S, \mathcal{F}, \mathbb{P})$, for $\forall A, B \in \mathcal{F}$: $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$.*

Proof. Since $\mathbb{P}(A \cap B) + \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, we have

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \\ &\geq \mathbb{P}(A) + \mathbb{P}(B) - 1. \end{aligned}$$

When $A \cup B = S$, the equality, $=$, holds. $\square$

Example 1. Given two event A, B in S, where $\mathbb{P}(A) = .8, \mathbb{P}(B) = .9$, then $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1 = .7$. On the other hands, $A \cap B \subseteq A$, thus $.7 \leq \mathbb{P}(A \cap B) \leq .8$.

Furthermore, we can extend the general form of Bonferroni's Inequality: Since $\mathbb{P}\left(\cup_{i=1}^n A_i^c\right) \leq \sum_{i=1}^n \mathbb{P}\left(A_i^c\right) = \sum_{i=1}^n 1 - \mathbb{P}(A_i) = n - \sum_{i=1}^n \mathbb{P}(A_i)$, and $\mathbb{P}\left(\cup_{i=1}^n A^c\right) = 1 - \mathbb{P}\left(\left(\cup_{i=1}^n A_i^c\right)^c\right) = 1 - \mathbb{P}\left(\cap_{i=1}^n A_i\right)$, thus

$$\mathbb{P}\left(\cap_{i=1}^n A_i\right) \geq \sum_{i=1}^n \mathbb{P}\left(A_i^c\right) - (n-1).$$

Bayes' theorem

Theorem 5. *Given a (finite or countable)partition $C_i (i \in I)$ of S, $A \subseteq S$, then $\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap C_i)$.*

Proof. Since $C_i (i \in I)$ is a partition, they are pairwise disjoin, thus $A \cap C_i (i \in I)$ are pairwise disjoin, and $\cup_{i \in I}(A \cap C_i) = A \cap (\cup_{i \in I} C_i) = A \cap S = A$, thus

$$\mathbb{P}(A) = \mathbb{P}(A \cap (\cup_{i \in I} C_i)) = \mathbb{P}(\cup_{i \in I}(A \cap C_i)) = \sum_{i \in I} \mathbb{P}(A \cap C_i).$$

$\square$

Note 2. Distribution holds on countable case: $A \cap (\cup_{i=1}^\infty C_i) = \cup_{i=1}^\infty (A \cap C_i)$, it can be proved by induction.

Definition 2 (Conditional probability). Given two events A, B of S, and $\mathbb{P}(B) > 0$. The condition probability $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ is the probability of A occurs under the condition that B occurs.

Note 3. The multiplication principle: $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$.

Theorem 6 (Bayes' theorem). *Given a (finite or countable) partition $A_i (i \in I)$ of S, $B \subseteq S$ and $\mathbb{P}(B) > 0$, then*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)}{\sum_{j \in I} \mathbb{P}\left(B|A_j\right) \cdot \mathbb{P}\left(A_j\right)}.$$

Proof. Trivial. $\square$

Theorem 7. *Given a sample space S with a σ algebra $\mathcal{F}$, suppose B is an event of S, and $\mathbb{P}(B) > 0$, then $\mathbb{P}(\cdot|B)$ is a probability on $\mathcal{F}$.*

Proof. 1. for $\forall A \in \mathcal{F}$, $\mathbb{P}(A|B) = \frac{A\cap B}{\mathbb{P}(B)} \geq 0$;

2. $\mathbb{P}(S|B) = \frac{\mathbb{P}(S\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$;
3. for countable pairwise disjoin events $A_i (i \in I)$, we have

$$\begin{aligned}
\mathbb{P}(\cup_{i\in I} A_i|B) &= \frac{\mathbb{P}(B \cap (\cup_{i\in I} A_i))}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{i\in I}(A_i \cap B))}{\mathbb{P}(B)} \\
&= \frac{\sum_{i\in I} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
&= \sum_{i\in I} \mathbb{P}(A_i|B).
\end{aligned}$$

□

Definition 3 (Independence)**.** Given two events A, B of S, we say A, B are independent events if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Otherwise, we say A, B are dependent events.

Note 4. If A, B are dependent, and $\mathbb{P}(B) > 0$, then $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Theorem 8. *Given two independent events A, B of S, we have*

1. *A and B^c are independent event;*
2. *A^c and B are independent event;*
3. *A^c and B^c are independent event;*

Proof. Since $A = (A \cap B) \cup (A \cap B^c)$, we have

$$\begin{aligned}
\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) \\
&= \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A) \cdot \mathbb{P}(B^c).
\end{aligned}$$

The others propositions are trivial. □

Definition 4 (Mutually independent)**.** We say events $A_1, \cdots, A_n$ on S are mutually independent if

$$\begin{aligned}
&\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j), \\
&\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j) \cdot \mathbb{P}(A_k), \\
&\qquad\qquad\vdots \\
&\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).
\end{aligned}$$

for any $i \neq j \neq k \neq \cdots$, $i, j, k, \cdots \in \{1, \cdots, n\}$.

Random variable

Intuitively, random variable X is a real function on sample space $S \xrightarrow{X} \mathbb{R}$. For example, flip a coin twice, we can define X is the number of heads, that is $X(H,H) = 2, X(H,T) = 1, X(T,H) = 1, X(T,T) = 0$.

Definition 5 (Random variable)**.** Random variable is a measurable real map from sample space S to $\mathbb{R}$.

Notice that the probability $\mathbb{P}$ is defined on the σ algebra $\mathcal{F}$ of the sample space S, instead of random variable X. So when we talk about the probability of the value of the random variable, what we mean is the probability of the **pre-image** of the map X, which is the element of $\mathcal{F}$.

The pre-image of X is denoted by X^{-1}, means

$$X^{-1}(x) = \{s \in S | X(s) = x\},$$

Note 5. Here X^{-1} is not the inverse of map X, since X would not be a bijection.

of course $X^{-1}(x) \subseteq S$ and $X^{-1}(x) \in \mathcal{F}$. And furthermore:

$$\mathbb{P}\left(X \in A\right) = \mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}\left(\{s \in S | X(s) \in A\}\right).$$

So when we flip a fair coin twice, and define X is the number of heads, then

$$\begin{aligned}\mathbb{P}\left(X = 1\right) &= \mathbb{P}\left(X^{-1}(1)\right) \\ &= \mathbb{P}\left(\{s \in S | X(s) = 1\}\right) \\ &= \mathbb{P}\left(\{(H,T), (T,H)\}\right) \\ &= \frac{1}{4}.\end{aligned}$$

Notice that