Handout[1] 2

Sept 9, 2022

[Topics]:
Probability
Conditional Probability
Bayes Rule

1 Probability

Definition 1: A *probability function* $\mathbb{P}$ assigns a numerical value to events and satisfies the following *axioms*:

1. $\mathbb{P}(A) \geq 0$.
2. $\mathbb{P}(S) = 1$.
3. If $A_1, A_2, \ldots$ are disjoint, then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} A_i$.

Axiom 1 says that for any event, its probability cannot be negative. It makes no sense with a negative probability. By Axiom 2 we know that $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(S) = 0$. That is, if an event will never happen, then it has probability 0. However, the inverse may not hold: An event with probability 0 dose not mean it is impossible. Think about a ruler, you pick a point at exact 10cm. It has probability 0 since a "dot" on a "line" has a proportion 0. But it is not an impossible event.

Example 1: The roll of a six-sided die has possible outcomes {1,2,3,4,5,6}. By Axiom 1, we have: $\mathbb{P}(1) = \frac{1}{6}$, $\mathbb{P}(7) = 0$. By Axiom 2, we know: $\mathbb{P}$(1 or 2 or 3 or 4 or 5 or 6)=1. By Axiom 3, we know: $\mathbb{P}$(2 or 3)=$\mathbb{P}(2) + \mathbb{P}(3)$.

Interpretation 1: There are mainly two ways to view probability. First is to treat probability as the frequency of outcomes. As in the example 1, the probability of the outcome is 1 can be considered as you roll the die again and again and again and again and ..., the frequency that the number 1 appears. Another view is to treat probability as a subjective belief. For example, you bet with you friend on whether Hongkai can swim. And you hold the belief that there is a 80% probability that I can. (In fact I can't swim, so you loss)

Property 1: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
Proof: A and A^c are disjoint and $A \cup A^c = S$. By Axiom 2 and 3, we have $\mathbb{P}(A \cup A^c) = \mathbb{P}(S) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1$.

Property 2: $\mathbb{P}(\emptyset) = 0$.
Proof: By $\emptyset = S^c$, and apply property 1, we have $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(S) = 0$.

Property 3: $\mathbb{P}(A) \leq 1$.
Proof: By Axiom 1, we have $\mathbb{P}(A^c) \geq 0$. Then $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \leq 1$.

Property 4: $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.
Proof: See property 5.

Property 5: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
Proof: By partition theroem, $A = (A \cap B) \cup (A \cap B^c)$, where $A \cap B$ and $A \cap B^c$ are disjoint. Also, $A \cup B = (A \cap B^c) \cup B$, where $A \cap B^c$ and B are disjoint. So we have:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cap B^c) + \mathbb{P}(B).$$

So we have:

$$\mathbb{P}(A \cup B) - \mathbb{P}(A) = \mathbb{P}(A \cap B^c) + \mathbb{P}(B) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B^c),$$

[1] This handout is made by Hongkai Wang for Econ 329 Economic Statistics. It is adapted from Professor Wiseman's lectures, however, all errors are mine.

which implies $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Interpretation 2: Look at the Venn diagram:

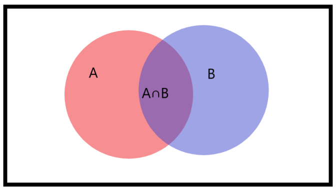

Figure 1: Venn Diagram for A and B

Observe that:

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(onlyA) + \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) + \mathbb{P}(onlyB).$$

While

$$\mathbb{P}(A \cup B) = \mathbb{P}(onlyA) + \mathbb{P}(onlyB) + \mathbb{P}(A \cap B).$$

So it is easy to see that for $\mathbb{P}(A \cup B)$, we need to eliminate once the middle part $\mathbb{P}(A \cap B)$ from $\mathbb{P}(A) + \mathbb{P}(B)$.

Property 6: If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
Proof: By partition theorem, we have $B = (B \cap A) \cup (B \cap A^c)$. By $A \subset B$, we have $A \cap B = A$. So $B = A \cup (B \cap A^c)$, where A and $B \cap A^c$ are disjoint. Then:

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A).$$

Example 2: For students in the Econ 329 calssroom, suppose 80% of them are taller than 5'. Then if you randomly pick a student, the probability that she is taller than 5' is 0.8. What is the possible probability that you randomly pick a student in the Econ 329 classroom, she is taller than 6'? It can be 0.75, 0.65, 0.5, 0.2, etc, but can not be larger than 0.8!

2 Conditional Probability

Definition 1: If $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is:

$$\mathbb{P}(A \mid B) = \tfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

$A \mid B$ denotes "A given B", it represents that suppose B already happened, what is the probability that A happens.

Example 1: Suppose you roll a die, the outcomes are in {1,2,3,4,5,6}. Let A denotes "the outcome is 2 or 3", and B denotes "the outcome is even". Then $\mathbb{P}(B) = \mathbb{P}(2 \quad or \quad 4 \quad or \quad 6) = \frac{3}{6} = \frac{1}{2}$. Suppose now given the outcome is even, what is the probability that the outcome is 2 or 3? i.e, find out $\mathbb{P}(A \mid B)$.

$$\begin{aligned}\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(2)}{\mathbb{P}(B)} \\ &= \frac{\frac{1}{6}}{\frac{1}{2}} \\ &= \frac{1}{3}\end{aligned}$$

Interpretation 1: Look at the venn diagram:

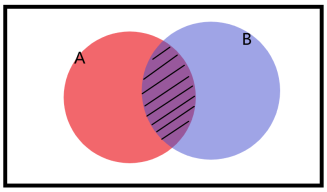

Figure 2: Venn Diagram for conditional probability

Given B already happened, then we now focus on the area of B instead of the whole sample space. Inside the B area, the event for A happens is exactly the shaded area, i,e. $A \cap B$. The conditional probability $\mathbb{P}(A \mid B)$ can be seen as the fraction of the shaded area over the area of event B.

Property 1: $\mathbb{P}(A \mid A) = 1$.
Proof: It follows directly by $A \cap A = A$. Given the event A happedned, the probability that event A happen is 1.

Property 2: If $B \subset A$, then $\mathbb{P}(A \mid B) = 1$.
Proof: $B \subset A$ implies $B \cap A = B$, hence $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.

Property 3: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$.
Proof: It follows by the definition of conditional probability.

Definition 2: Event A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Example 2: Hongkai will go to H.E.B after class and Hongkai's roommate will go to Target after calss. Suppose Hongkai has a probability of 0.8 to buy a bouquet of roses and Hongkai's roommate has a probability of 0.3 to buy a bouquet of lilies. Assume they don't know each other's plan after calss (hence the two events are independent). What is the probability that Hongkai's home will have flowers?
Solution: Let A denotes the event "Hongkai will buy the rose", and B denotes "Hongkai's roommate will buy the lily". Then the home will have flowers means either one of them buy fmowers or both of them buy the flowers, i,e. $A \cup B$. So

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= 0.8 + 0.3 - 0.8 * 0.3 \\ &= 0.86 \end{aligned}$$

Property 4: If A and B are independent with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then:

$$\mathbb{P}(A \mid B) = \mathbb{P}(A), \text{ and } \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

If A and B are independent, then given B happened, it has no influence on whether A will happen or not.

Property 5: For $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, if A and B are disjoint, then they can not be independent.
Proof: A and B disjoint implies $A \cap B = \emptyset$. So $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)$.

3 Bayes Rule

Theorem 1 (Law of Total Probability): Suppose $\{B_i\}_{i=1}^{\infty}$ is a partition of the sample space S, then

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Proof: Recall the partition theorem: if $\{B_i\}_{i=1}^{\infty}$ is a partition of S, then $A = \bigcup_{i=1}^{\infty}(A \cap B_i)$, and $(A \cap B_i)$ are disjoint. Then by Axiom 3, we have $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i)$. Now apply property 3 in secton 2.

Example 1: You roll a die. Let A denotes the event "the outcome is 2 or 3", let B_1 denotes the event "the outcome is odd" and B_2 denotes "the outcome is even". Then given B_1, i,e. we know the outcome is odd. The conditional probability of A is $\frac{1}{3}$, which comes from the number 3 is one of {1,3,5}. Similar for others. Now

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\mathbb{P}(B_2) \\
&= \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \\
&= \frac{1}{3}
\end{aligned}$$

Theorem 2 (Bayes Rule): Let $\{B_i\}_{i=1}^{k}$ is a partition of S and $\mathbb{P}(B_i) > 0$, then

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

Proof: Numerator follows from property 3 in section 2. Denominatior follows from law of total probability.

A more simple case is:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)}.$$

Here, B and B^c form a partition of S.

Example 2: There is a game between Texas and Alabama, but Hongkai doesn't have the ticket to watch it. Suppose the probability that Texas wins is 0.9. UT students have a probability of 0.95 to cheer if Texas wins. UT students love their team, there is a probability of 0.5 that they will cheer if Texas loses. Hongkai is waiting outside the DKR stadium, suppose he sees people are cheering. What is the probability that Texas wins?

Solution: Let A denotes the event "Texas wins" and B denotes "people cheer". Then

$$\begin{aligned}
&\mathbb{P}(A) = 0.9, \\
&\mathbb{P}(B \mid A) = 0.95, \\
&\mathbb{P}(B \mid A^c) = 0.5.
\end{aligned}$$

We want to calculate $\mathbb{P}(A \mid B)$.

$$\begin{aligned}
\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c)} \\
&= \frac{0.95 \cdot 0.9}{0.95 \cdot 0.9 + 0.5 \cdot 0.1} \\
&\approx 0.9448.
\end{aligned}$$

Good luck on your midterm!