

STAT131 Ch.2 Notes

William Santosa

Winter 2022 Quarter

1 Conditional Probability

Probability is a language for expressing our degrees of belief or uncertainties about events. In light of that, it is only right that we think about the probability of an event given that an event has already occurred. That is, we want to find the *conditional probability* of such an event occurring. As such, conditional probability is an important idea to master as we delve deeper into statistics and its applications in real life.

1.1 Definition and Intuition

If A and B are events with $P(B) > 0$, the *conditional probability* of A given B, denoted as $P(A|B)$, is

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

A is the event whose probability we update, B is the evidence observed (what we know), $P(A)$ is the *prior* probability of A and $P(A|B)$ is the *posterior* probability of A, with it meaning before and after the evidence, respectively.

- **Note:** For any event A, $P(A|A) = \frac{P(A \cap A)}{P(A)} = 1$
- **Note:** The treatment of an event as having been occurred is called *conditioning* (e.g conditioning an event)

1.2 Bayes' Rule and the Law of Total Probability

We explored conditional probability in the previous section and it is fairly simple. However, there are many consequences out of that simple definition. One easily obtainable consequence is seen below,

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

Another theorem, for the probability of the intersection of n events, is as follows. For any events A_1, \dots, A_n with $A_1, A_2, \dots, A_{n-1} > 0$,

$$P(A_1, A_2, \dots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)\dots P(A_n|A_1, \dots, A_{n-1})$$

where the commas denote intersections ($P(A_3|A_1, A_2) = P(A_3|A_1 \cap A_2)$), which is useful to simplify the notation.

The next theorem, *Bayes' Rule*, states that

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

Another way to right Bayes' rule is in terms of probability, the *odds* of an event A are

$$\text{odds}(A) = \frac{P(A)}{P(A^c)}$$

To convert back to probability,

$$P(A) = \frac{\text{odds}(A)}{1 + \text{odds}(A)}$$

The odds form of Bayes' rule is

$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(B|A)}{P(B|A^c)} \frac{P(A)}{P(A^c)}$$

The *laws of total probability* states that when A_1, \dots, A_n is a partition of the sample space S (A_i is disjoint and their union is S), with $P(A_i) > 0$ for all i, then

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

1.3 Conditional Probabilities are Probabilities

Basically, just because we know that an event has happened doesn't mean we can ignore the laws of probabilities that we learned in Chapter 1. Thus, we know that:

1. Conditional probabilities are between 0 and 1, inclusive.
2. $P(S|E) = 1, P(\emptyset|E) = 0$
3. If A_1, A_2, \dots are disjoint, then $P(\cup_{j=1}^{\infty} A_j|E) = \sum_{j=1}^{\infty} P(A_j|E)$
4. $P(A^c|E) = 1 - P(A|E)$
5. Inclusion-exclusion: $P(A \cup B|E) = P(A|E) + P(B|E) - P(A \cap B|E)$

Here's Bayes' rule with extra conditioning. Provided that $P(A \cap E) > 0$ and $P(B \cap E) > 0$ we have

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

Here's the Law of Total Probability with extra conditioning. Let A_1, \dots, A_n be a partition of S. Provided that $P(A_i \cap E) > 0$ for all i , we have

$$P(B|E) = \sum_{i=1}^n P(B|A_i, E)P(A_i|E)$$

1.4 Independence of Events

Since we have talked about the conditioning on one event changing the beliefs about the probability of another event, we will now refer to the situation where events provide no information about each other as *independence*.

Two events A and B are independent if

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

If $P(A) > 0$ and $P(B) > 0$, then this is equivalent to

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

as well as

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

That means two events are independent if we can obtain the probability of their intersection by multiplying their individual probabilities. In word form, they are independent if learning one event occurred does not change the probabilities for the other occurring.

- **Note:** We also know that if A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.

The three events A, B, and C are independent if all of the following are true:

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

$$P(A \cap C) = P(A)P(C)$$

$$P(B \cap C) = P(B)P(C)$$

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

If the first three conditions hold, events A, B, and C are *pairwise independent*. That means learning about one event does not help predict another event, however, learning about two events may be relevant to predict the last event.

1.5 Coherency of Bayes' Rule

One important property of Bayes' rule is that it is *coherent*: receiving multiple pieces of information and updating probabilities with them doesn't matter with the order or simultaneously. That means you can update information as you go or do it all at once.

1.6 Conditioning as a Problem-Solving Tool

Straight from the textbook, "Conditioning is a powerful tool for solving problems because it lets us engage in *wishful thinking*: when we encounter a problem that would be made easier if only we knew whether E happened or not, we can condition on E and then on E^c , consider these possibilities separately, then combine them using LOTP. "

1.7 Pitfalls and Paradoxes

Two more fallacies are the prosecuto's fallacy (Confusion of $P(A|B)$ with $P(B|A)$) and the defense attorney's fallacy where is the failure to condition on all the evidence.