STSCI 5080
Probability Models and Inference
Lecture 2: Probability and Conditional Probability

August 28, 2018

Announcement

- Homework 1 will be posted tonight. Due is 9/6 (Th) in class.
- If you have not received enrollment PINs yet, please contact me (kk976@cornell.edu) immediately.

Probability

Definition (Probability measure)

A probability measure P on a sample space Ω is a function from subsets of Ω to real numbers that satisfies:

Axiom 1 (Normalization) $P(\Omega) = 1$.

Axiom 2 (Nonnegativity) $P(A) \geq 0$ for any event $A \subset \Omega$.

Axiom 3 (Additivity) If A and B are disjoint events (i.e., $A \cap B = \emptyset$), then

$$P(A \cup B) = P(A) + P(B).$$

Example 2.1

Example

Suppose that we know that $P(A) = 1/4, P(B) = 3/8$, and $A \subset B$. What is $P(B \cap A^c)$?

Example 2.1

Example

Suppose that we know that $P(A) = 1/4, P(B) = 3/8$, and $A \subset B$. What is $P(B \cap A^c)$?

Recall that

$$B = (B \cap A) \cup (B \cap A^c)$$

where $B \cap A$ and $B \cap A^c$ are disjoint. So,

$$P(B) = P(B \cap A) + P(B \cap A^c).$$

Since $A \subset B$, we have $B \cap A = A$, and so $P(B \cap A) = P(A)$. Hence

$$P(B \cap A^c) = P(B) - P(A) = 3/8 - 1/4 = 1/8.$$

Some properties

Property A

$$P(A^c) = 1 - P(A).$$

Proof.

A and A^c are disjoint, and $A \cup A^c = \Omega$. So

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c),$$

and hence $P(A^c) = 1 - P(A)$. $\square$

Property B

$$P(\emptyset) = 0.$$

Proof.

Since $\emptyset^c = \Omega$,

$$1 = P(\Omega) = P(\emptyset^c) = 1 - P(\emptyset),$$

which implies $P(\emptyset) = 0$. $\square$

Property C

If $A \subset B$, then $P(A) \leq P(B)$.

Proof.

We can write B as

$$B = \underbrace{(B \cap A)}_{=A} \cup (B \cap A^c) = A \cup (B \cap A^c).$$

Since A and $B \cap A^c$ are disjoint, Axiom 3 implies that

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

□

In particular, for any event A, since $\emptyset \subset A \subset \Omega$, we have

$$0 \leq P(A) \leq 1.$$

Some properties (cont.)

Property D

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof.

In $P(A) + P(B)$, we double-count the probability of $A \cap B$. So by subtracting $P(A \cap B)$ once, we get $P(A \cup B)$. $\square$

Example 2.2

Example

In a certain town, two newspapers X and Y are published. Suppose

- 45% of families in the town subscribe to newspaper X;
- 35% to Y;
- 10% to both X and Y.

What percentage of families in the city subscribe to at least one of the two newspapers?

Let

$$A = \text{a randomly selected family subscribes to newspaper X,}$$
$$B = \text{a randomly selected family subscribes to newspaper Y.}$$

We know that $P(A) = 0.45, P(B) = 0.35$, and $P(A \cap B) = 0.1$. We want to compute $P(A \cup B)$. Property D implies that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.45 + 0.35 - 0.1 = 0.7.$$

Generalization of Property D

Problem 1.8.6

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

(Hint). Let $D = B \cup C$, and apply Property D to $A \cup D$.

$$P(A \cup D) = P(A) + P(D) - P(A \cap D).$$

Apply again Property D to $D = B \cup C$ and to

$$A \cap D = A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Google "Inclusion-exclusion formula".

Finite sample space

- Suppose that the sample space Ω is finite, i.e.,

$$\Omega = \{\omega_1, \dots, \omega_N\}.$$

  E.g. coin tossing, dice rolling etc.

- In this case, defining a probability measure on Ω is very simple: it is enough to assign a probability to each ω_i.

Finite sample space (cont.)

Theorem

Suppose that Ω is finite, i.e., $\Omega = \{\omega_1, \ldots, \omega_N\}$. Let $p_1, \ldots, p_N$ be such that

- $p_i \geq 0$ *for all* $i = 1, \ldots, n$;
- $p_1 + \cdots + p_N = 1$.

Now, let $P(\{\omega_1\}) = p_1, \ldots, P(\{\omega_N\}) = p_N$, and more generally,

$$P(A) = \sum_{i\colon\ \omega_i \in A} p_i \quad \text{for any event } A \subset \Omega.$$

Then so defined P is a probability measure on Ω.

Special case

Theorem

Suppose that Ω is finite, i.e., $\Omega = \{\omega_1, \ldots, \omega_N\}$, and all the outcomes occur equally likely, i.e., $P(\{\omega_1\}) = \cdots = P(\{\omega_N\}) = 1/N$. Then

$$P(A) = \frac{\textit{number of elements of } A}{N}.$$

Example 2.3

Example

Suppose that we toss a coin three times and all the outcomes occur equally likely.

$$P(\{hhh\}) = P(\{hht\}) = \cdots = P(\{ttt\}) = \frac{1}{8}.$$

Now, what is the probability of the following event?

$$A = \text{exactly two heads appear.}$$

Conditional probability

- Conditional probability provides a way to reason about the outcome of an experiment based on partial information.

- E.g. Suppose that you toss a coin three times. What is the probability that you have more heads than tails given that the first toss is a head?

- E.g. How likely is it that a person can live until age 90 given that he/she is alive at age 75?

Definition

Definition

Let Ω be a sample space and let P a probability measure on Ω. For two events A, B with $P(B) > 0$, the conditional probability of A given B is defined as

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

Example 2.4

Example

Suppose you toss a coin three times and all the outcomes occur equally likely. What is the probability that you have more heads than tails given that the first toss is a head?

Example 2.4

Example

Suppose you toss a coin three times and all the outcomes occur equally likely. What is the probability that you have more heads than tails given that the first toss is a head?

The sample space is $\Omega = \{hhh, hht, hth, thh, htt, tht, tth, ttt\}$, and

$$A = \text{more heads than tails} = \{hhh, hht, hth, thh\},$$
$$B = \text{the first toss is a head} = \{hhh, hht, hth, htt\}.$$

We want to compute $P(A \mid B)$. We know that $P(B) = 4/8 = 1/2$. In addition,

$$P(A \cap B) = P(\{hhh, hht, hth\}) = \frac{3}{8}.$$

So,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{1/2} = \frac{3}{4} = 0.75.$$

Example 2.5

Example

In a certain country, 40% percent of the population can live until age 75, and 10% can live until 90. What is the probability that you can live until age 90 given that you are alive at age 75?

Example 2.5

Example

In a certain country, 40% percent of the population can live until age 75, and 10% can live until 90. What is the probability that you can live until age 90 given that you are alive at age 75?

Let

$$A = \text{a randomly selected person can live until age 90,}$$
$$B = \text{a randomly selected person can live until age 75.}$$

We know that $P(A) = 0.4$ and $P(B) = 0.1$. In addition, $A \subset B$ (why?), and so $A \cap B = A$. Hence,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{0.1}{0.4} = 0.25.$$

Conditional probabilities specify a probability measure

Theorem

Fix any event B with $P(B) > 0$. Then $P(A \mid B)$ as a function of $A \subset \Omega$ is a new probability measure on Ω.

- The conditional probability is a probability measure and so all properties of the probability measure remain valid. E.g.,

$$P(A^c \mid B) = 1 - P(A \mid B).$$

- Since

$$P(B \mid B) = P(B \cap B)/P(B) = P(B)/P(B) = 1,$$

  the conditional probability is concentrated on B. We may discard all possible outcomes outside B and treat the conditional probability as a probability measure on the new universe B.

Multiplication law

Theorem

If $P(B) > 0$, then $P(A \cap B) = P(A \mid B)P(B)$.

Message: Sometimes we know $P(A \mid B)$ and $P(B)$. In such cases, we can calculate $P(A \cap B)$ by the multiplication law.

Example 2.6

Example

Two cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). What is the probability that none of the two cards is a heart?

Let

$$A = \text{the second card is not a heart,}$$
$$B = \text{the first card is not a heart.}$$

We want to calculate $P(A \cap B)$. By the multiplication law,

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

We know that $P(B) = 39/52$. Next, given that the first card is drawn, we are left with 51 cards 38 of which are not hearts, and so

$$P(A \mid B) = \frac{38}{51}.$$

Hence,

$$P(A \cap B) = \frac{38}{51} \cdot \frac{39}{52}.$$

Generalization

Theorem

Assuming that all of the conditioning events have positive probabilities, we have

$$P\left(\cap_{i=1}^{n} A_i\right) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P\left(A_n \mid \cap_{i=1}^{n-1} A_i\right).$$

For example,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2).$$

Law of total probability and Bayes rule

Theorem (Law of total probability)

Let B *be an event such that* $0 < P(B) < 1$. *Then for any event* A,

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c).$$

Theorem (Bayes rule)

Let A *and* B *be events such that* $P(A) > 0$ *and* $P(B) > 0$. *Then*

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

Example 2.7

Example

Suppose that there are two urns X and Y. Urn X contains 3 red balls and 7 blue balls, and urn Y contains 6 red balls and 4 blue balls. Now:

1. You first choose urn X or urn Y with probability $1/2$.
2. Next, you draw a ball from the chosen urn.

Then (a) what is the probability that your ball is red? (b) Given that your ball is red, what is the probability that your ball was drawn from urn Y?

Let

$$\begin{aligned} A &= \text{you draw a red ball}, \\ B &= \text{you choose urn X}. \end{aligned}$$

We know that $P(B) = 0.5,\ P(A \mid B) = 3/10 = 0.3$, and $P(A \mid B^c) = 6/10 = 0.6$. Hence, the law of total probability implies that

$$\begin{aligned} P(A) &= P(A \mid B)P(B) + P(A \mid B^c)P(B^c) \\ &= 0.3 \times 0.5 + 0.6 \times 0.5 = 0.45. \end{aligned}$$

For (b), we want to calculate $P(B^c \mid A)$. The Bayes rule implies that

$$P(B^c \mid A) = \frac{P(A \mid B^c)P(B^c)}{P(A)} = \frac{0.6 \times 0.5}{0.45} = \frac{2}{3}.$$