Lecture 01: Basics of Probability

Mathematical Statistics I, MATH 60061/70061

Tuesday August 31, 2021

Independence of three events

Events A, B, and C are said to be independent if *all* of the following equations hold:

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

If the first three conditions hold, we say that A, B, and C are *pairwise independent*.

Pairwise independence does not imply independence

Consider two fair, independent coin tosses, and define the events:

- A: The first is Head, $P(A) = 1/2$
- B: The second is Head, $P(B) = 1/2$
- C: Both tosses have the same result, $P(C) = 1/2$

Pairwise independence does not imply independence

Consider two fair, independent coin tosses, and define the events:

- A: The first is Head, $P(A) = 1/2$
- B: The second is Head, $P(B) = 1/2$
- C: Both tosses have the same result, $P(C) = 1/2$

$A \cap B$, $A \cap C$, $B \cap C$, and $A \cap B \cap C$: Both tosses are Heads.

- A, B, and C are pairwise independent.

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$$

- But, A, B, and C are NOT independent.

$$P(A \cap B \cap C) = 1/4 \neq P(A)P(B)P(C)$$

Example: two children problem

Martin Gardner, *Scientific American* 1959

Does it matter whether we learn the older child's gender, as opposed to just learning one child's gender?

- *Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?*
- *Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?*

Example: two children problem

Martin Gardner, *Scientific American* 1959

Does it matter whether we learn the older child's gender, as opposed to just learning one child's gender?

- *Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?*
- *Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?*

Assumptions:

- A child's gender is binary (i.e., either a boy or a girl).
- $P(\text{boy}) = P(\text{girl})$
- The genders of the two children are *independent*.

Example: two children problem

Martin Gardner, *Scientific American* 1959

- Probability that both children are girls, given the older one is a girl, $P(\text{both girls} \mid \text{elder is a girl})$?

- Probability that both children are boys, given that at least one of them is a boy, $P(\text{both boys} \mid \text{at least one boy})$?

Example: two children problem

Martin Gardner, *Scientific American* 1959

- Probability that both children are girls, given the older one is a girl, $P(\text{both girls} \mid \text{elder is a girl})$?

$$\frac{P(\text{both girls}, \text{elder is a girl})}{P(\text{elder is a girl})} = \frac{1/4}{1/2} = \frac{1}{2}$$

- Probability that both children are boys, given that at least one of them is a boy, $P(\text{both boys} \mid \text{at least one boy})$?

Example: two children problem

Martin Gardner, *Scientific American* 1959

- Probability that both children are girls, given the older one is a girl, $P(\text{both girls} \mid \text{elder is a girl})$?

$$\frac{P(\text{both girls}, \text{elder is a girl})}{P(\text{elder is a girl})} = \frac{1/4}{1/2} = \frac{1}{2}$$

- Probability that both children are boys, given that at least one of them is a boy, $P(\text{both boys} \mid \text{at least one boy})$?

$$\frac{P(\text{both boys}, \text{at least one boy})}{P(\text{at least one boy})} = \frac{1/4}{3/4} = \frac{1}{3}$$

Example: testing for rare disease

Suppose a rare disease affects 1% of the population, and a test for the disease is 95% accurate. A patient named Fred is tested for disease, and the test result is positive. Given the evidence provided by the test result, what's the conditional probability that Fred has the disease?

Example: testing for rare disease

Suppose a rare disease affects 1% of the population, and a test for the disease is 95% accurate. A patient named Fred is tested for disease, and the test result is positive. Given the evidence provided by the test result, what's the conditional probability that Fred has the disease?

Let D be the event that Fred has the disease and T be the event that he tests positive.

The test is "95% accurate" means

- $P(T \mid D) = 0.95$ (sensitivity or true positive rate)
- $P(T^c \mid D^c) = 0.95$ (specificity or true negative rate)

Example: testing for rare disease

Suppose a rare disease affects 1% of the population, and a test for the disease is 95% accurate. A patient named Fred is tested for disease, and the test result is positive. Given the evidence provided by the test result, what's the conditional probability that Fred has the disease?

$$
\begin{aligned}
P(D \mid T) &= \frac{P(T \mid D)P(D)}{P(T)} \\
&= \frac{P(T \mid D)P(D)}{P(T \mid D)P(D) + P(T \mid D^c)P(D^c)} \\
&= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \\
&\approx 0.16
\end{aligned}
$$

Example: testing for rare disease, continued

With a test of 95% accuracy, Fred was tested positive for a rare disease that affects 1% of the population. He decides to get tested a second time. The new test is *independent* of the original test (given his disease status) and has the same sensitivity and specificity. Unfortunately, he tests positive a second time. Given the evidence, what's the probability that Fred has the disease?