ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

INDEPENDENCE AND CONDITIONAL PROBABILITY

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Conditional Probability and Independence

- Let B be a set with $\mathbb{P}(B) \neq 0$, then we define **the conditional probability of** A **given** B as

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

- **Independence**: Two events A and B are independent if and only if

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

Example: Independence

Example:

- Suppose that 9% of people are left-handed and 2 people are selected at random from the U.S. population.

- Because 2 is very small relative to the U.S. population, it is reasonable to assume these two people are independent.
  - What is the probability that both are left-handed? By independence $\mathbb{P}(\text{both left handed}) = \mathbb{P}(\text{first left handed}) \times \mathbb{P}(\text{second left handed}) = 0.09^2$.

Example: Conditional Probability

- A data set provides a sample of 6,224 individuals from the year 1721 who were exposed to smallpox in Boston.
- Doctors believed that inoculation (exposing a person to the disease in a controlled form) could reduce the likelihood of death.
- Let's find the probability that a random, non-inoculated person died from small pox.

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

Contingency table for smallpox data set — Openintro text Ch. 3.

		inoculated		
		yes	no	Total
result	lived	0.0382	0.8252	0.8634
	died	0.0010	0.1356	0.1366
	Total	0.0392	0.9608	1.0000

Table of proportions for smallpox data set (entries divided by 6224) — Openintro text Ch. 3.

Example: Conditional Probability

- Let's find the probability that a random, non-inoculated person died from small pox. Define:
  - $A =$ set of people who died .
  - $B =$ set of people who are not inoculated
- We can now calculate

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.1356}{0.9608} \approx 0.1411.$$

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

		inoculated		
		yes	no	Total
result	lived	0.0382	0.8252	0.8634
	died	0.0010	0.1356	0.1366
	Total	0.0392	0.9608	1.0000

Example: Conditional Probability

- Let's find the probability that a random, inoculated person died from small pox. Recall:

  - $A =$ set of people who died .

  - $B =$ set of people who are not inoculated

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

- We now want to calculate

$$\mathbb{P}(A \mid B^c) = \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)} = \frac{6}{244} \approx 0.0246.$$

		inoculated		
		yes	no	Total
result	lived	0.0382	0.8252	0.8634
	died	0.0010	0.1356	0.1366
	Total	0.0392	0.9608	1.0000

Bayes Theorem

- **Bayes Theorem**, stated mathematically, is the expression

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

- It allows us to calculate the probability of (event A conditioned on event B) from (event B conditioned on event A).

- Sometimes $\mathbb{P}(B)$ is only accessible through the expression

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B \mid A_i)\mathbb{P}(A_i) \quad \text{where.} \quad A_i \cap A_j = \varnothing \quad \forall i \neq j \text{ and } \cup_i A_i = \Omega$$

Example: Bayes Theorem

- Recall our smallpox example:
  - $A =$ set of people who died .
  - $B =$ set of people who are not inoculated
- We had calculated. $\mathbb{P}(A \mid B) = 0.1411.$
- Use Bayes' theorem to calculate the probability that a random person was not inoculated, given that they died of small pox.
  - $\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{P(A)} = \dfrac{0.1411 \times 0.9608}{0.1366}$
  - $\mathbb{P}(B \mid A) \approx 0.993.$

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

		inoculated		
		yes	no	Total
result	lived	0.0382	0.8252	0.8634
	died	0.0010	0.1356	0.1366
	Total	0.0392	0.9608	1.0000