SDSC2102
Statistical Methods and Data Analysis

Topic 1. Basic Probability and Statistics Theory

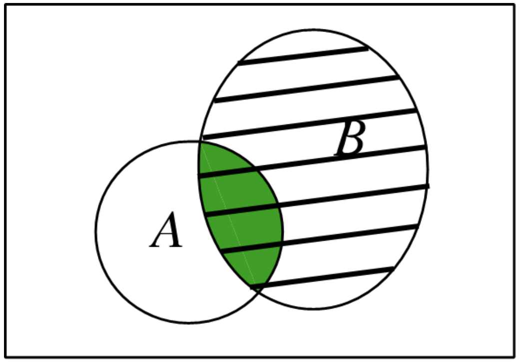
Conditional Probability

- The conditional probability of *A*, given *B*, is the probability that event *A* occurs when it is known that event *B* occurs:

$$P(A|B) = \frac{\#\text{ of outcomes in events } A \text{ and } B}{\#\text{ of outcomes in event } B}$$

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

where $P(B) > 0$



Class Problem

1. Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?

A: At least one die lands on 6
B: Two dice land on different numbers

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

Method 1:

$$P(A|B) = \frac{10}{30}$$

Class Problem

1. Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

$$P(B) = \frac{30}{36}, P(A \cap B) = \frac{10}{36}$$

Method 2:

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

$$= \frac{10/36}{30/36} = \frac{10}{30}$$

Multiplicative Rule

- Derived from the definition of conditional probability:

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

- In general, for events $A_1, A_2, \ldots, A_k$:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k)$$

$$= P(A_1)P(A_2 \mid A_1)P(A_3 | A_1 \cap A_2) \ldots$$

$$P(A_k | A_1 \cap A_2 \cap \ldots \cap A_{k-1})$$

Class Problem

2. An urn contains 6 white balls and 9 black balls. If 4 balls are to be randomly selected without replacement, what is the probability that the first 2 selected are white and the last 2 are black?

E_1: 1st white E_2 : 2nd white E_3 : 3rd black E_4 : 4th black
Find $P(E_1 \cap E_2 \cap E_3 \cap E_4)$

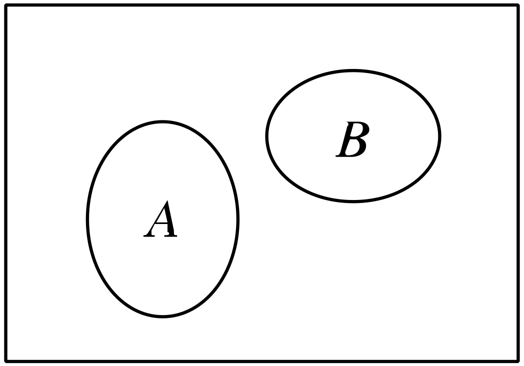
$$P(E_1 \cap E_2 \cap E_3 \cap E_4)$$
$$= P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)P(E_4|E_1 \cap E_2 \cap E_3)$$
$$= \frac{6}{15} \cdot \frac{5}{14} \cdot \frac{9}{13} \cdot \frac{8}{12}$$

Independence

- Two events, *A* and *B*, are independent if and only if the following conditions exist
  - $P(A \mid B) = P(A)$ or
  - $P(B \mid A) = P(B)$
- Given that events *A* and *B* are independent, the multiplicative rule becomes:
  - $P(A \cap B) = P(A)\,P(B)$

Disjoint vs. Independent Events

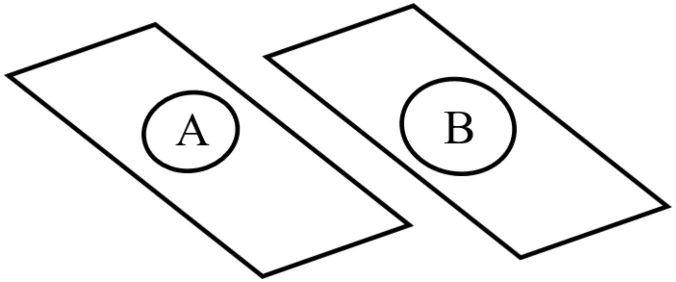
Disjoint



$$A \cap B = \emptyset$$

Two events do not occur at the same time.

Independent



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

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

Two events are unrelated.

Class Problem

3. A couple has 2 children. What is the probability that both are girls if the eldest is a girl? What is the probability that both are girls if we know there is at least one girl out of two children?

A: The elder is a girl
B: The younger is a girl
C: At least one girl

Find $P(A \cap B|A)$, $P(A \cap B|C)$

Class Problem

$$S = \{gg, gb, bb, bg\}$$

$$P(A \cap B|A) = P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B)$$

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{P(A \cap B)}{P(C)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Class Problem

4. 98% of all babies survive delivery. However 15% of all births involve Cesarean (C) section, and when a C-section is performed the baby survives 96% of the time. If a randomly chosen pregnant woman does not have a C-section, what is the probability that her baby survives?

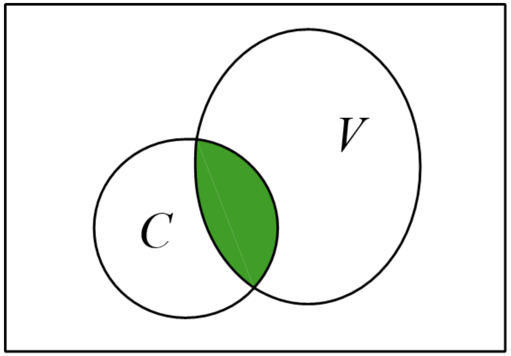
V: A baby survives
C: A mother has a C-section
$P(V) = 0.98, P(C) = 0.15, P(V|C) = 0.96$
Find $P(V|C^c)$

Independence

$$P(V|C^c) = \frac{P(V \cap C^c)}{P(C^c)} = \frac{0.836}{0.85} = 0.9835$$

$$P(C^c) = 1 - P(C) = 1 - 0.15 = 0.85$$

$$\begin{aligned} P(V \cap C^c) &= P(V) - P(V \cap C) \\ &= 0.98 - P(C)P(V|C) \\ &= 0.98 - 0.15 \times 0.96 \\ &= 0.836 \end{aligned}$$



Bayes' Rule

- Bayes' rule allows us to compute a conditional probability knowing the "reverse" conditional probabilities.

  Reason 1, Reason 2, Reason 3

  - Suppose events R_1, R_2, and R_3 partition S.
  - Suppose we know $P(R_1)$, $P(R_2)$, and $P(R_3)$.

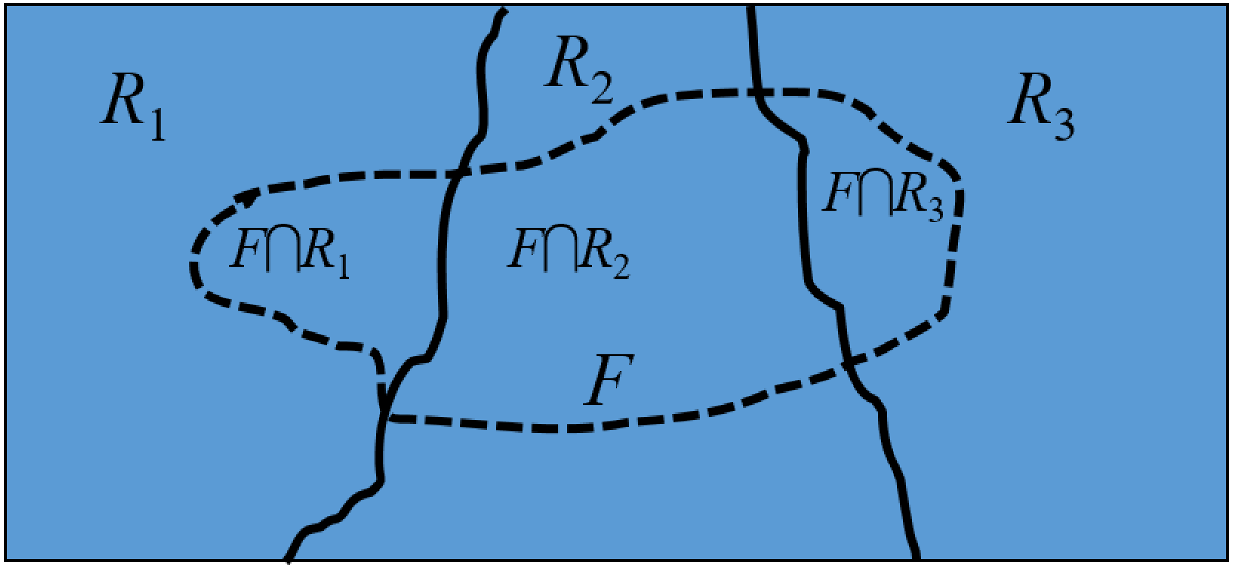
  Fact

  - For a given event F, suppose we also know $P(F \mid R_1)$, $P(F \mid R_2)$, and $P(F \mid R_3)$.

- We wish to find

$$P(R_1 \mid F),\ P(R_2 \mid F), \text{ and } P(R_3 \mid F).$$

Bayes' Rule



- <u>Law of Total Probability</u>

$$P(F) = P(F \cap R_1) + P(F \cap R_2) + P(F \cap R_3)$$

$$= P(F \mid R_1)P(R_1) + P(F \mid R_2)P(R_2) + P(F \mid R_3)P(R_3)$$

Bayes' Rule

- The conditional probability that R_1 (or R_2 or R_3) occurs given that F has occurred:

$$P(R_1|F) = \frac{P(R_1 \cap F)}{P(F)}$$

$$= \frac{P(F|R_1)P(R_1)}{P(F|R_1)P(R_1) + P(F|R_2)P(R_2) + P(F|R_3)P(R_3)}$$

General Formula

- Given events $R_1, R_2, \ldots, R_k$ partition S.
- **Law of Total Probability**

$$P(F) = \sum_{i=1}^{k} P(F|R_i)P(R_i)$$

- **Bayes' Rule**

$$P(R_r|F) = \frac{P(F|R_r)P(R_r)}{\sum_{i=1}^{k} P(F|R_i)P(R_i)}$$

for $r = 1, 2, \ldots, k$

Class Problem

1. Three cooks – A, B, and C – bake a special kind of cake, and with respective probabilities 0.02, 0.03, and 0.05 it fails to rise. In the restaurant where they work A bakes 50% of these cakes, B bakes 30%, and C bakes 20%. What fraction of the cakes fails to rise in the restaurant?

F: Cake fails to rise
A: Cake made by A
B: Cake made by B
C: Cake made by C

Class Problem

Given

$$P(A) = 0.5, P(B) = 0.3, P(C) = 0.2$$

$$P(F|A) = 0.02, P(F|B) = 0.03, P(F|C) = 0.05$$

Find $P(F)$

$$P(F) = P(F|A)P(A) + P(F|B)P(B) + P(F|C)P(C)$$

$$= 0.02 \times 0.5 + 0.03 \times 0.3 + 0.05 \times 0.2$$

$$= 0.029$$

Class Problem

2. The probability that HK wins a football game given it scored less than 30 points is 0.7 and the probability it wins given the score was greater than or equal to 30 points is 0.9. Also, the probability that HK scores less than 30 points in a game is 0.4. What is the probability that HK will win a football game?

W: HK wins

A: HK scores less than 30 points

Class Problem

Given

$$P(W|A) = 0.7, P(W|A^c) = 0.9$$

$$P(A) = 0.4, P(A^c) = 1 - P(A) = 0.6$$

Find $P(W)$

$$P(W) = P(W|A)P(A) + P(W|A^c)P(A^c)$$

$$= 0.7 \times 0.4 + 0.9 \times 0.6$$

$$= 0.82$$

Class Problem

3. Three cooks – A, B, and C – bake a special kind of cake, and with respective probabilities 0.02, 0.03, and 0.05 it fails to rise. In the restaurant where they work A bakes 50% of these cakes, B bakes 30%, and C bakes 20%. What is the probability that a given failed cake was baked by cook A?

F: Cake fails to rise
A: Cake made by A
B: Cake made by B
C: Cake made by C

Class Problem

Given

$$P(A) = 0.5, P(B) = 0.3, P(C) = 0.2$$

$$P(F|A) = 0.02, P(F|B) = 0.03, P(F|C) = 0.05$$

Find $P(A|F)$

From Problem 1 $\quad P(F) = 0.029$

$$P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{P(F|A)P(A)}{P(F)} = \frac{0.02 \times 0.5}{0.029} = 0.3448$$

$$P(B|F) = \frac{P(B \cap F)}{P(F)} = \frac{P(F|B)P(B)}{P(F)} = \frac{0.03 \times 0.3}{0.029} = 0.3103$$

$$P(C|F) = \frac{P(C \cap F)}{P(F)} = \frac{P(F|C)P(C)}{P(F)} = \frac{0.05 \times 0.2}{0.029} = 0.3448$$

Propositions

- $0 \leq P(A|B) \leq 1$
- $P(\varnothing) = 0$
- $P(S|B) = 1$
- If events $A_1, A_2, \ldots, A_k$ are disjoint, then

$$P(A_1 \cup A_2 \cup \ldots \cup A_k|B)$$
$$= P(A_1|B) + P(A_2|B) + \ldots + P(A_k|B)$$

$$\Downarrow$$

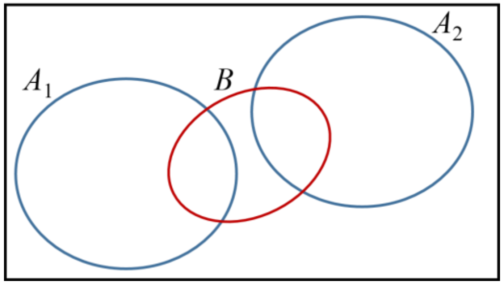
$P(A|B)$ is a probability.

Proof

$$1.\ 0 \le P(A|B) = \frac{P(A \cap B)}{P(B)} \le \frac{P(B)}{P(B)} = 1$$

$$2.\ P(\emptyset|B) = \frac{P(\emptyset \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$$

$$3.\ P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$



$$4.\ P(A_1 \cup A_2|B) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)}$$

$$= \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)}$$

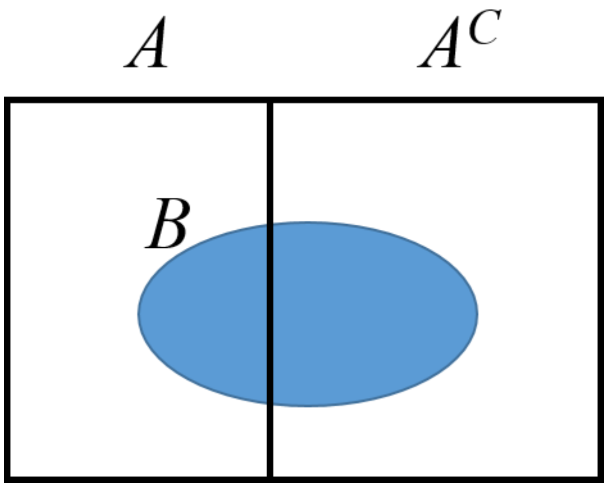
$$= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = P(A_1|B) + P(A_2|B)$$

Rules of Conditional Probability

➢ Extensions of Axioms and Rules:

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

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



Class Problem

4. Suppose that there was a cancer diagnostic test that was 95% accurate on those that do have cancer and 96% accurate on those who don’t. If 0.4% of the population has cancer, compute the probability that a tested person has cancer, given that his or her test results indicate so.

F: Test indicates cancer

F^C: Test does not indicate cancer

A: A person has cancer

A^C: A person does not have cancer

Class Problem

Given

$$P(F|A) = 0.95, P\left(F^C \middle| A^C\right) = 0.96$$
$$P(A) = 0.004$$

Find $P(A|F)$

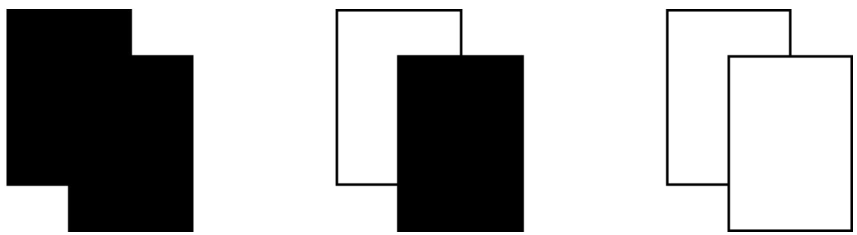
$$P(A|F) = \frac{P(F|A)P(A)}{P(F)} = \frac{P(F|A)P(A)}{P(F|A)P(A) + P\left(F\middle|A^C\right)P(A^C)}$$

$$= \frac{P(F|A)P(A)}{P(F|A)P(A) + (1 - P\left(F^C\middle|A^C\right))(1 - P(A))}$$

$$= \frac{0.95 \times 0.004}{0.95 \times 0.004 + 0.04 \times 0.996} = 0.0871$$

Three-Card Problem

Problem : Suppose there are 3 cards in a hat. One card is black on both sides, one card is white on one side and black on the other, and the last card is white on both sides. Close your eyes and mix up the cards nicely. Draw one card and drop it on the floor and put away the hat (with the other 2 cards). Now open your eyes. If you see the card on the floor with a black face up, what is the probability that the other side is also black?



C_1 C_2 C_3

B: Black appears

W: White appears

Find $P(C_1|B)$

Three-Card Problem

Given

$$P(C_1) = P(C_2) = P(C_3) = 1/3$$

$$P(B|C_1) = 1, P(B|C_2) = 0.5, P(B|C_3) = 0$$

Find $P(C_1|B)$

$$P(C_1|B) = \frac{P(B|C_1)P(C_1)}{P(B|C_1)P(C_1) + P(B|C_2)P(C_2) + P(B|C_3)P(C_3)}$$

$$= \frac{1 \times 1/3}{1 \times 1/3 + 0.5 \times 1/3 + 0 \times 1/3}$$

$$= \frac{1 \times 1/3}{1 \times 1/3 + 0.5 \times 1/3 + 0 \times 1/3} = \frac{2}{3}$$