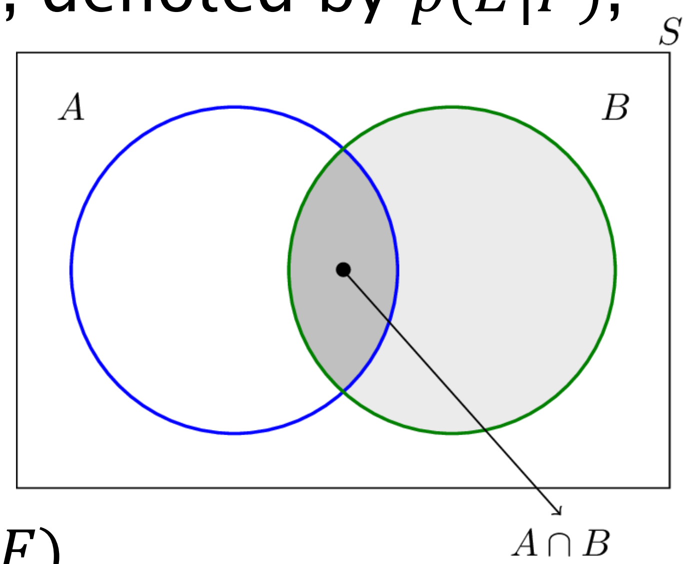
L16: Conditional Probability and Bayes' Theorem

- Reading: Rosen 7.2, 7.3

Conditional Probability

- **Definition**: Let E and F be events with $p(F) > 0$. The conditional probability of E given F, denoted by $p(E|F)$, is defined as:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$



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

- **Corollary 1:**

$$p(E \cap F) = p(F)p(E|F)$$

- **Corollary 2:**

$$p(E_1 \cap E_2 \cap \cdots \cap E_n)$$
$$= p(E_1 \cap E_2 \cap \cdots \cap E_{n-1})p(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}) = \cdots$$
$$= p(E_1)p(E_2|E_1)p(E_3|E_1 \cap E_2) \cdots p(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

Example

- **Question:**
  In a random bit string of length 4 (all 16 possible strings are equally likely), what is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?
- **Solution**:
  E: The event that the bit string contains at least two consecutive 0s
  F: The event that the first bit is a 0.
  - Since $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, $p(E \cap F) = 5/16$.
  - $p(F) = 1/2$.
  - So $p(E|F) = \frac{p(E \cap F)}{p(F)} = 5/8$

Example

- **Example**
  If a student knows 80% of the material in a course, what is the probability that she answers a question correctly on a well-balanced true-false test assuming that she randomly guesses on any question for which she does not know the answer?
- **Solution**
  - Q: The question is in the student's knowledge, $p(Q) = 0.8, p(\bar{Q}) = 0.2$
  - C: The student answers the question correctly, $p(C|Q) = 1, p(C|\bar{Q}) = 0.5$
  - $p(C) = p(C \cap Q) + p(C \cap \bar{Q})$
    $= p(Q)p(C|Q) + p(\bar{Q})p(C|\bar{Q}) = 0.9$

The Birthday Problem

- **Question**
  What is the minimum number of people who need to be in a room so that there must be two people with the same birthday, assuming that there are 366 days in a year?
- **Answer**: 367 (pigeonhole principle)
- **Question**
  What is the minimum number of people who need to be in a room so that the probability that at least two of them have the same birthday is greater than 1/2, assuming that there are 366 days in a year, the birthdays of the people in the room are independent, and each birthday is equally likely?

The Birthday Problem: Solution

- Imagine the people entering the room one by one
- D_n: The event that the first n people all have different birthdays.
- Want to find the smallest n such that $p(D_n) < 1/2$.
- B_i: The birthday of the i-th person
- $p(D_1) = 1$
- $p(D_2) = p(D_1) \cdot p(B_2 \neq B_1 \mid D_1) = 1 \times \frac{365}{366}$
- $p(D_3) = p(D_2) \cdot p(B_3 \neq B_1 \land B_3 \neq B_2 \mid D_2) = 1 \times \frac{365}{366} \times \frac{364}{366}$
- $p(D_n) = p(D_{n-1}) \cdot p(\forall i \in \{1, \dots, n-1\}, B_n \neq B_i \mid D_{n-1})$
  $= 1 \times \frac{365}{366} \times \frac{364}{366} \times \cdots \times \frac{367-n}{366}$
- It turns out $p(D_{23}) \approx 0.494$

Independence Revisited

- Recall the definition of two events being independent

$$p(E \cap F) = p(E)p(F)$$

- Since $p(E \cap F) = p(F)p(E|F)$
  - This always holds whether E and F are independent or not
- The independence condition can be rewritten as

$$p(E|F) = p(E)$$

- Symmetrically, the condition can also be

$$p(F|E) = p(F)$$

- This gives an alternative and more intuitive definition of independence.

Independence Example Revisited

- **Example**:
  In a randomly generated bit string of length 4:
  E: it begins with a 1
  F: it contains an even number of 1s.
  Are *E* and *F* independent?
- **Solution**:
  - $p(F) = 1/2$.
  - $p(F|E) = 1/2$
  - Note that whether F happens or not only depends on even/odd just before the last bit and the last bit.

Bayes' Theorem: Example

- Breathalyzers have an error rate of 5%.
  - A sober driver is detected as "drunk" with prob. 5%
  - A drunk driver is not detected with prob. 5%
- 1 in 1000 drivers is driving drunk
- Police officers stops a driver at random
- Breathalyzer detects "drunk"

- Question
  What's the probability that the driver is really drunk?
- Answer
  95%? Wrong! p(drunk | detected) $\neq$ p(detected | drunk)
- Answer
  2%

Bayes' Theorem

- **Theorem**
  Suppose that E and F are events from a sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

- **Proof**

  Just plug in the definition of conditional probability and observe that

$$p(E \cap F) + p(E \cap \bar{F}) = p(E)$$

Breathalyzer Example Explained

- Let
  - D: drunk
  - B: detected by breathalyzer
- We know
  - $p(B|D) = 0.95$
  - $p(B|\overline{D}) = 0.05$
  - $p(D) = 0.001$
  - Want $p(D|B)$

$$p(D|B) = \frac{p(B|D)p(D)}{p(B|D)p(D) + p(B|\overline{D})p(\overline{D})}$$

$$= \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.05 \times 0.999} \approx 0.02$$

Example

- **Example**: We have two boxes.
  - Box 1 contains 2 green balls and 7 red balls.
  - Box 2 contains 4 green balls and 3 red balls.

  Bob first picks one of the boxes at random. Then he selects a ball from that box at random. If he has a red ball, what is the probability that he picked box 1 at first.
- Solution:
  - *E*: Bob has chosen a red ball
  - *F*: Bob has chosen box 1.

  We are given that $p(E|F) = \frac{7}{9}, p(E|\bar{F}) = \frac{3}{7}, p(F) = \frac{1}{2}$

  Applying Bayes' theorem yields $p(F|E) \approx 0.645$.

Example

- **Example**: There is a test for a particular disease.
  - The test's false negative rate is 1%, i.e., it gives a negative result with prob. 1% when given to someone with the disease.
  - The test's false positive rate is 0.5%, i.e., it gives a positive result with prob. 0.5% when given to someone without the disease.
  - On average, one person out of 100,000 has the disease.
- Question:
  Should someone who tests positive be worried?

Example: Solution

- D: the person has the disease
- E: this person tests positive.
- We need to compute $p(D|E)$

$$p(D) = 1/100,000 = 0.00001 \quad p(\overline{D}) = 1 - 0.00001 = 0.99999$$

$$p(E|D) = .99 \quad p(\overline{E}|D) = .01 \quad p(E|\overline{D}) = .005 \quad p(\overline{E}|\overline{D}) = .995$$

$$p(D|E) = \frac{p(E|D)p(D)}{p(E|D)p(D) + p(E|\overline{D})p(\overline{D})}$$

$$= \frac{(0.99)(0.00001)}{(0.99)(0.00001) + (0.005)(0.99999)}$$

$$\approx 0.002$$

Example: Solution (cnt'd)

- What if the result is negative?

$$p(\overline{D}|\overline{E}) = \frac{p(\overline{E}|\overline{D})p(\overline{D})}{p(\overline{E}|\overline{D})p(\overline{D}) + p(\overline{E}|D)p(D)}$$

$$= \frac{(0.995)(0.99999)}{(0.995)(0.99999) + (0.01)(0.00001)}$$

$$\approx 0.9999999$$

$$\begin{aligned} p(D|\overline{E}) & \\ &\approx 1 - 0.9999999 \\ &= 0.0000001. \end{aligned}$$

Bayesian Spam Filter

- Given:
  - A set B of spam messages
  - A set G of non-spam messages
- Goal: Compute the probability that a new email message is spam
- We look at a particular word w, and count the number of messages in which it occurs in B and in G; $n_B(w)$ and $n_G(w)$.
  - (Estimated) prob. that a spam email contains w: $p(w|B) = n_B(w)/|B|$
  - Prob. that a good email contains w: $p(w|G) = n_G(w)/|G|$
  - The prob. that a new email is spam is $p(B) = \frac{|B|}{|B|+|G|}$.

Bayesian Spam Filter: Example

- Given:
  - $|B| = 2000$
  - $|G| = 1000$
  - The word "Rolex" occurs in 250 spam messages and 5 good messages
  - $p(w|B) = 250/2000$
  - $p(w|G) = 5/1000$
  - $p(B) = \frac{2000}{1000+2000}$
- Suppose the new email contains "Rolex"
- The prob. that it is spam is

$$p(B|w) = \frac{p(w|B)p(B)}{p(w|B)p(B) + p(w|G)p(G)} \approx 0.98$$

Generalized Bayes' Theorem

- **Theorem**:
  Suppose that E is an event from a sample space S and that $F_1, F_2, \dots, F_n$ are mutually exclusive events such that $\cup_{i=1}^{n} F_i = S$. Assume that $p(E) \neq 0$ and $p(F_i) \neq 0$ for $i = 1, 2, \dots, n$. Then
  $$p(F_j|E) = \frac{p(E|F_j)p(F_j)}{\sum_{i=1}^{n} p(E|F_i)p(F_i)}.$$

- **Remark**:
  This degenerates into the standard Bayes' theorem when $n = 2$.
- **Proof:**
  Similar, just using $E = \cup_{i=1}^{n} (E \cap F_i)$

Example

- The market share of a particular device is as follows:
  - Company A: 80%
  - Company B: 15%
  - Company C: 5%
- The defect rate of devices by these companies:
  - Company A: 4%
  - Company B: 6%
  - Company C: 9%
- Questions
  - Given a random device from the market, what are the probabilities that it is made by these 3 companies, respectively?
  - How do the results change if the random device is found to be defective?

Example: Solution

- Solution to the first question: Just the market shares
- Solution to the second question
  - A, B, C: The events that the device is made by these 3 companies, respectively
    $p(A) = 0.8, p(B) = 0.15, p(C) = 0.05$
  - D: The device is defective
    $p(D|A) = 0.04, p(D|B) = 0.06, p(D|C) = 0.09$

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|B)p(B) + p(D|C)p(C)} \approx 0.7$$

$$p(B|D) = \frac{p(D|B)p(B)}{p(D|A)p(A) + p(D|B)p(B) + p(D|C)p(C)} \approx 0.2$$

$$p(C|D) = \frac{p(D|C)p(C)}{p(D|A)p(A) + p(D|B)p(B) + p(D|C)p(C)} \approx 0.1$$

Monty Hall Problem

- Behind one door is a prize; behind the others, goats.
- You pick a door, say door 1
- The host looks behind the 3 doors.
  - If your pick is incorrect, the host opens the other door with a goat
  - If your pick is correct, the host randomly chooses another door and open it
- Now he asks: Do you want to change your mind?
- Answer: Change!
- Explanation 1: Your first pick is correct with probability 1/3. This doesn't change after the host opens a door. The 3rd door then must have probability 2/3 to have the prize.

Monty Hall Problem (cnt'd)

- Explanation using Bayes' theorem
- W: the door with the prize
- M: the door Monty (the host) opens
- We know:

$$p(W = 1) = p(W = 2) = p(W = 3) = 1/3$$

- Suppose we pick door 1 first. Then Monty opens door M:

$$p(M = 2|W = 1) = 1/2, p(M = 3|W = 1) = 1/2$$
$$p(M = 3|W = 2) = 1$$
$$p(M = 2|W = 3) = 1$$

- Suppose Monty opens door 2. Then

$$p(W = 1|M = 2)$$
$$= \frac{p(M = 2|W = 1)p(W = 1)}{p(M = 2|W = 1)p(W = 1) + p(M = 2|W = 2)p(W = 2) + p(M = 2|W = 3)p(W = 3)}$$
$$= \frac{1}{3}$$

- $p(W = 3|M = 2) = \frac{2}{3}$
- Switch

Monty Hall Problem (cnt'd)

- Now suppose:

$$p(W = 1) = 0.5, \qquad p(W = 2) = 0.3, \qquad p(W = 3) = 0.2$$

- Suppose we pick door 1 first. Then Monty opens door 2:

$$p(M = 2|W = 1) = 1/2$$
$$p(M = 2|W = 2) = 0$$
$$p(M = 2|W = 3) = 1$$

- $p(W = 1|M = 2) = \frac{p(M = 2|W = 1)p(W=1)}{p(M=2)}$

$$= \frac{p(M = 2|W = 1)p(W=1)}{p(M = 2|W = 1)p(W=1)+p(M = 2|W = 2)p(W=2)+p(M = 2|W = 3)p(W=3)}$$

$$= \frac{\frac{1}{2}(0.5)}{\frac{1}{2}(0.5)+0(0.3)+1(0.2)} = \frac{5}{9}$$

- $p(W = 3|M = 2) = \frac{4}{9}$

The probability that our initial pick containing a prize has increased after door 2 is opened. So should not switch to door 3.

Monty Hall Problem (cnt'd)

- Suppose we pick door 1 first. Then Monty opens door 3:

$$p(M = 3|W = 1) = 1/2$$
$$p(M = 3|W = 2) = 1$$
$$p(M = 3|W = 3) = 0$$

- $p(W = 1|M = 3) = \frac{p(M = 3|W = 1)p(W=1)}{p(M=3)}$

$$= \frac{p(M = 3|W = 1)p(W=1)}{p(M = 3|W = 1)p(W=1)+p(M = 3|W = 2)p(W=2)+p(M = 3|W = 3)p(W=3)}$$

$$= \frac{\frac{1}{2}(0.5)}{\frac{1}{2}(0.5)+1(0.3)+0(0.2)} = \frac{5}{11}$$

- $p(W = 2|M = 3) = \frac{6}{11}$
- The probability that our initial pick containing a prize has decreased after door 3 is opened. So should switch to door 2.

Monty Hall Problem (cnt'd)

- Overall winning probability if we adopt the strategy:

$$p(W = 1 ; M = 2) + p(W = 2 ; M = 3)$$
$$= p(W = 1)p(M = 2|W = 1) + p(W = 2)p(M = 3|W = 2)$$
$$= 0.5 \cdot 0.5 + 0.3 \cdot 1 = 0.55$$

- Alternatively,

$$p(W = 1 ; M = 2) + p(W = 2 ; M = 3)$$
$$= p(M = 2)p(W = 1|M = 2) + p(M = 3)p(W = 2|M = 3)$$
$$= 0.45 \cdot (\frac{5}{9}) + 0.55 \cdot (\frac{6}{11}) = 0.55$$

- Is there a better strategy? See 2019 Spring Final Exam Q11