Discrete Probability

Chapter 7

With Question/Answer Animations

Chapter Summary

- Introduction to Discrete Probability
- Probability Theory
- Bayes' Theorem

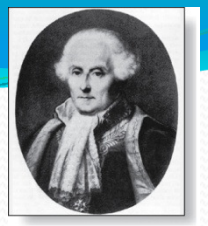
An Introduction to Discrete Probability

Section 7.1

Section Summary

- Finite Probability
- Probabilities of Complements and Unions of Events

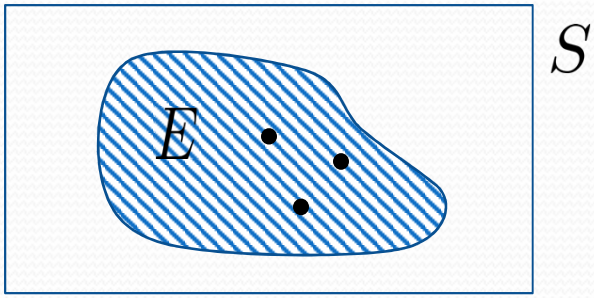
Probability of an Event



Pierre-Simon Laplace
(1749-1827)

We first study Pierre-Simon Laplace's classical theory of probability, which he introduced in the 18th century, when he analyzed games of chance.

- We first define these key terms:
  - An *experiment* is a procedure that yields one of a given set of possible outcomes.
  - The *sample space* of the experiment is the set of possible outcomes.
  - An *event* is a subset of the sample space.



Definition: If S is a finite sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the *probability* of E is

$$p(E) = |E| / |S|.$$

- For every event E, we have $0 \leq p(E) \leq 1$. This follows directly from the definition because $0 \leq p(E) = |E|/|S| \leq |S|/|S| \leq 1$, since $0 \leq |E| \leq |S|$.

Applying Laplace's Definition

Example: An urn contains four blue balls and five red balls. What is the probability that a ball chosen from the urn is blue?

Solution: The probability that the ball is chosen is 4/9 since there are nine possible outcomes, and four of these produce a blue ball.

Example: What is the probability that when two dice are rolled, the sum of the numbers on the two dice is 7?

Solution: By the product rule there are $6^2 = 36$ possible outcomes. Six of these sum to 7. Hence, the probability of obtaining a 7 is $6/36 = 1/6$.

Applying Laplace's Definition

Example: In a lottery, a player wins a large prize when they pick four digits that match, in correct order, four digits selected by a random mechanical process. What is the probability that a player wins the prize?

Solution: By the product rule there are $10^4 = 10{,}000$ ways to pick four digits.

- Since there is only 1 way to pick the correct digits, the probability of winning the large prize is $1/10{,}000 = 0.0001$.

A smaller prize is won if only three digits are matched. What is the probability that a player wins the small prize?

Solution: If exactly three digits are matched, one of the four digits must be incorrect and the other three digits must be correct. For the digit that is incorrect, there are 9 possible choices (all except the correct one). Hence, by the sum rule, there a total of 36 possible ways to choose four digits that match exactly three of the winning four digits. The probability of winning the small price is $36/10{,}000 = 9/2500 = 0.0036$.

Applying Laplace's Definition

Example: There are many lotteries that award prizes to people who correctly choose a set of six numbers out of the first n positive integers, where n is usually between 30 and 60.

What is the probability that a person picks the correct six numbers out of 40?

Solution: The number of ways to choose six numbers out of 40 is

$$C(40,6) = 40!/(34!6!) = 3,838,380.$$

Hence, the probability of picking a winning combination is

$$1/\ 3,838,380 \approx 0.00000026.$$

Applying Laplace's Definition

Example: What is the probability that the numbers 11, 4, 17, 39, and 23 are drawn in that order from a bin with 50 balls labeled with the numbers 1,2, …, 50 if

a) The ball selected is not returned to the bin.

b) The ball selected is returned to the bin before the next ball is selected.

Solution: Use the product rule in each case.

a) *Sampling without replacement*: The probability is $1/254{,}251{,}200$ since there are $P(50,5) = 50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 = 254{,}251{,}200$ ways to choose the five balls.

b) *Sampling with replacement*: The probability is $1/50^5 = 1/312{,}500{,}000$ since $50^5 = 312{,}500{,}000$.

The Probability of Complements and Unions of Events

Theorem 1: Let E be an event in sample space S. The probability of the event $\overline{E} = S - E$, the complementary event of E, is given by

$$p(\overline{E}) = 1 - p(E).$$

Proof: Using the fact that $|\overline{E}| = |S| - |E|$,

$$p(\overline{E}) = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E). \quad \blacktriangleleft$$

The Probability of Complements and Unions of Events

Example: A sequence of 10 bits is chosen randomly. What is the probability that at least one of these bits is 0?

Solution: Let E be the event that at least one of the 10 bits is 0. Then $\overline{E}$ is the event that all of the bits are 1s. The size of the sample space S is 2^{10}. Hence,

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{2^{10}} = 1 - \frac{1}{1024} = \frac{1023}{1024}.$$

The Probability of Complements and Unions of Events

Theorem 2: Let E_1 and E_2 be events in the sample space S. Then

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

Proof: Given the inclusion-exclusion formula from Section 2.2, $|A \cup B| = |A| + |B| - |A \cap B|$.

It follows that

$$p(E_1 \cup E_2) = \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|}$$

$$= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$= p(E_1) + p(E_2) - p(E_1 \cap E_2). \quad \blacktriangleleft$$

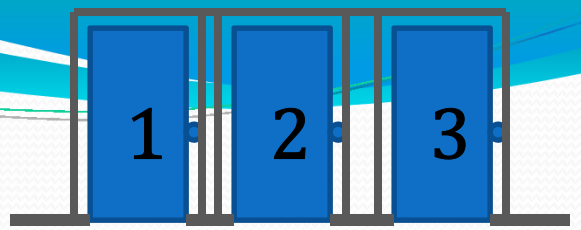
The Probability of Complements and Unions of Events

Example: What is the probability that a positive integer selected at random from the set of positive integers not exceeding 100 is divisible by either 2 or 5?

Solution: Let E_2 be the event that the integer is divisible by 2 and E_5 be the event that it is divisible 5? Then the event that the integer is divisible by 2 or 5 is $E_2 \cup E_5$ and $E_2 \cap E_5$ is the event that it is divisible by 2 and 5.

It follows that:

$$p(E_2 \cup E_5) = p(E_2) + p(E_5) - p(E_2 \cap E_5)$$
$$= 50/100 + 20/100 - 10/100 = 3/5.$$



Monty Hall Puzzle

Example: You are asked to select one of the three doors to open. There is a large prize behind one of the doors and if you select that door, you win the prize. After you select a door, the game show host opens one of the other doors (which he knows is not the winning door). The prize is not behind the door and he gives you the opportunity to switch your selection. Should you switch?

(*This is a notoriously confusing problem that has been the subject of much discussion . Do a web search to see why!)*

Solution: You should switch. The probability that your initial pick is correct is 1/3. This is the same whether or not you switch doors. But since the game show host always opens a door that does not have the prize, if you switch the probability of winning will be 2/3, because you win if your initial pick was not the correct door and the probability your initial pick was wrong is 2/3.

Can use a computer simulation to confirm this.

Probability Theory

Section 7.2

Section Summary

- Assigning Probabilities
- Probabilities of Complements and Unions of Events
- Conditional Probability
- Independence
- Bernoulli Trials and the Binomial Distribution

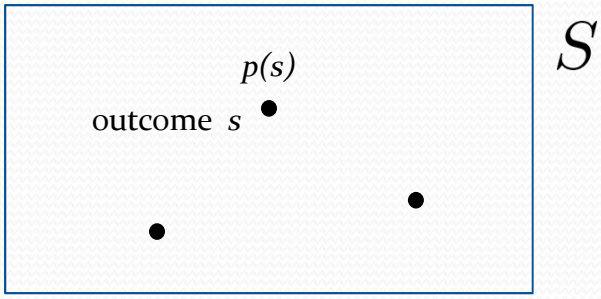
Assigning Probabilities

Laplace's definition from the previous section, assumes that all outcomes are equally likely. Now we introduce a more general definition of probabilities that avoids this restriction.

- Let S be a sample space of an experiment with a finite number of outcomes. We assign a probability $p(s)$ to each outcome s, so that:

i. $0 \le p(s) \le 1$ for each $s \in S$

ii. $\sum_{s \in S} p(s) = 1$



- The function p from the set of all outcomes of the sample space S to interval $[0,1]$ is called a *probability distribution*.

Assigning Probabilities

Example: What probabilities should we assign to the outcomes H (heads) and T (tails) when a fair coin is flipped?

What probabilities should be assigned to these outcomes when the coin is biased so that heads comes up twice as often as tails?

Solution: We have $p(H) = 2p(T)$.

Because $p(H) + p(T) = 1$, it follows that

$$2p(T) + p(T) = 3p(T) = 1.$$

Hence, $p(T) = 1/3$ and $p(H) = 2/3$.

Uniform Distribution

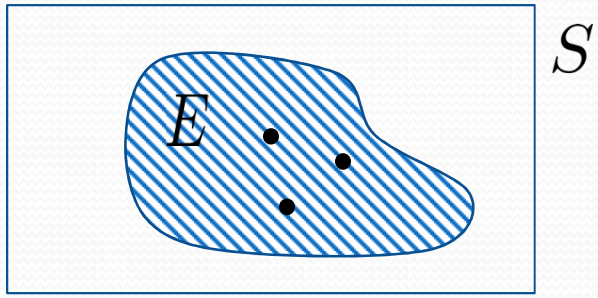
Definition: Suppose that S is a set with n elements. The *uniform distribution* assigns the probability $1/n$ to each element of S. (Note that we could have used Laplace's definition here.)

Example: Consider again the coin flipping example, but with a fair coin. Now $p(H) = p(T) = 1/2$.

Probability of an Event

Definition: The probability of the event E is the sum of the probabilities of the outcomes in E.

$$p(E) = \sum_{s \in S} p(s)$$



- Note that now no assumption is being made about the distribution.

Example

Example: Suppose that a dice is biased so that 3 appears twice as often as each other number, but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this dice?

Solution: We want the probability of the event $E = \{1,3,5\}$. We have $p(3) = 2/7$ and

$$p(1) = p(2) = p(4) = p(5) = p(6) = 1/7 \quad \text{(why?)}$$

Hence, $p(E) = p(1) + p(3) + p(5) =$

$$1/7 + 2/7 + 1/7 = 4/7.$$

Probabilities of Complements and Unions of Events

- Complements: $p(\overline{E}) = 1 - p(E)$ still holds. Since each outcome is in either E or $\overline{E}$, but not both,

$$\sum_{s \in S} p(s) = 1 = p(E) + p(\overline{E}).$$

- Unions: $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$ also still holds under the new definition.

Combinations of Events

Theorem: If E_1, E_2, … is a sequence of pairwise disjoint events in a sample space S, then

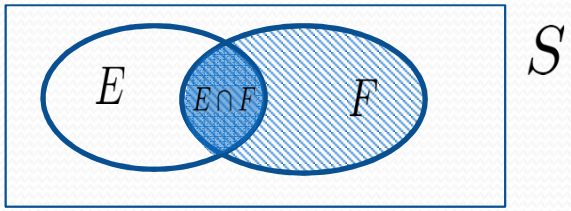
$$p\left(\bigcup_i E_i\right) = \sum_i p(E_i)$$

see Exercises 36 and 37 for the proof

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:

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



Example: A bit string of length four is generated at random so that each of the 16 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: Let E be the event that the bit string contains at least two consecutive 0s, and F be the event that the first bit is a 0.

- Since $E \cap F$ = {0000, 0001, 0010, 0011, 0100}, $p(E \cap F)$=5/16.
- Because 8 bit strings of length 4 start with a 0, p(F) = 8/16= ½.

Hence,

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{5/16}{1/2} = \frac{5}{8}.$$

Conditional Probability

Example: What is the conditional probability that a family with two children has two boys, given that they have at least one boy. Assume that each of the possibilities *BB*, *BG*, *GB*, and *GG* is equally likely where *B* represents a boy and *G* represents a girl.

Solution: Let E be the event that the family has two boys and let F be the event that the family has at least one boy. Then $E = \{BB\}$, $F = \{BB, BG, GB\}$, and $E \cap F = \{BB\}$.

- It follows that p(F) = 3/4 and $p(E \cap F) = 1/4$.

Hence,

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Independence

Intuition: two events are independent if the occurrence of one of the events gives us no information about whether or not the other event will occur; that is, the events have no influence on each other.

Formal Definition: The events E and F are independent if and only if

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

Note that independence of events E and F implies

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

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

Independence

Example: Suppose E is the event that a randomly generated bit string of length four begins with a 1 and F is the event that this bit string contains an even number of 1s. Are E and F independent if the 16 bit strings of length four are equally likely?

Solution: There are eight bit strings of length four that begin with a 1, and eight bit strings of length four that contain an even number of 1s.

- Since the number of bit strings of length 4 is 16,

$$p(E) = p(F) = 8/16 = ½.$$

- Since $E \cap F = \{1111, 1100, 1010, 1001\}$, $p(E \cap F) = 4/16 = 1/4$.

We conclude that E and F are independent, because

$$p(E \cap F) = 1/4 = (½)\,(½) = p(E)\,p(F)$$

Independence

Example: Assume (as in the previous example) that each of the four ways a family can have two children (BB, GG, BG,GB) is equally likely. Are the events E, that a family with two children has two boys, and F, that a family with two children has at least one boy, independent?

Solution: Because $E = \{BB\}$, $p(E) = 1/4$. We saw previously that that $p(F) = 3/4$ and $p(E \cap F) = 1/4$. The events E and F are not independent since

$$p(E)\, p(F) = 3/16 \neq \tfrac{1}{4} = p(E \cap F) .$$

Pairwise and Mutual Independence

Definition: The events $E_1, E_2, \ldots, E_n$ are *pairwise independent* if and only if $p(E_i \cap E_j) = p(E_i)\, p(E_j)$ for all pairs i and j with $i \leq j \leq n$.

The events are *mutually independent* if

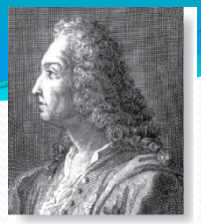
$$p(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = p(E_{i_1}) p(E_{i_2}) \cdots p(E_{i_m})$$

whenever i_j, $j = 1,2,\ldots,m$, are integers with

$$1 \leq i_1 < i_2 < \cdots < i_m \leq n \quad \text{and } m \geq 2.$$

NOTE: mutually independent events are pairwise independent, but some pairwise independent events are not mutually independent.

James Bernoulli
(1854 – 1705)



Bernoulli Trials

Definition: Suppose an experiment can have only two possible outcomes, *e.g.*, the flipping of a coin or the random generation of a bit.

- Each performance of the experiment is called a *Bernoulli trial*.
- Often, one outcome is called a *success* and the other a *failure*.
- If p is the probability of success and q the probability of failure, then $p + q = 1$.

- Many problems involve determining the probability of k successes when an experiment consists of n mutually independent Bernoulli trials.

Bernoulli Trials

Example: A coin is biased so that the probability of heads is 2/3. What is the probability that exactly four heads occur when the coin is flipped seven times?

Solution: There are $2^7 = 128$ possible outcomes. The number of ways four of the seven flips can be heads is $C(7,4)$. The probability of each of the outcomes is $(2/3)^4(1/3)^3$ since the seven flips are independent. Hence, the probability that exactly four heads occur is

$$C(7,4)\,(2/3)^4(1/3)^3 = (35\cdot 16)/\,2^7 = 560/\,2187.$$

Probability of k Successes in n Independent Bernoulli Trials.

Theorem 2: The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is

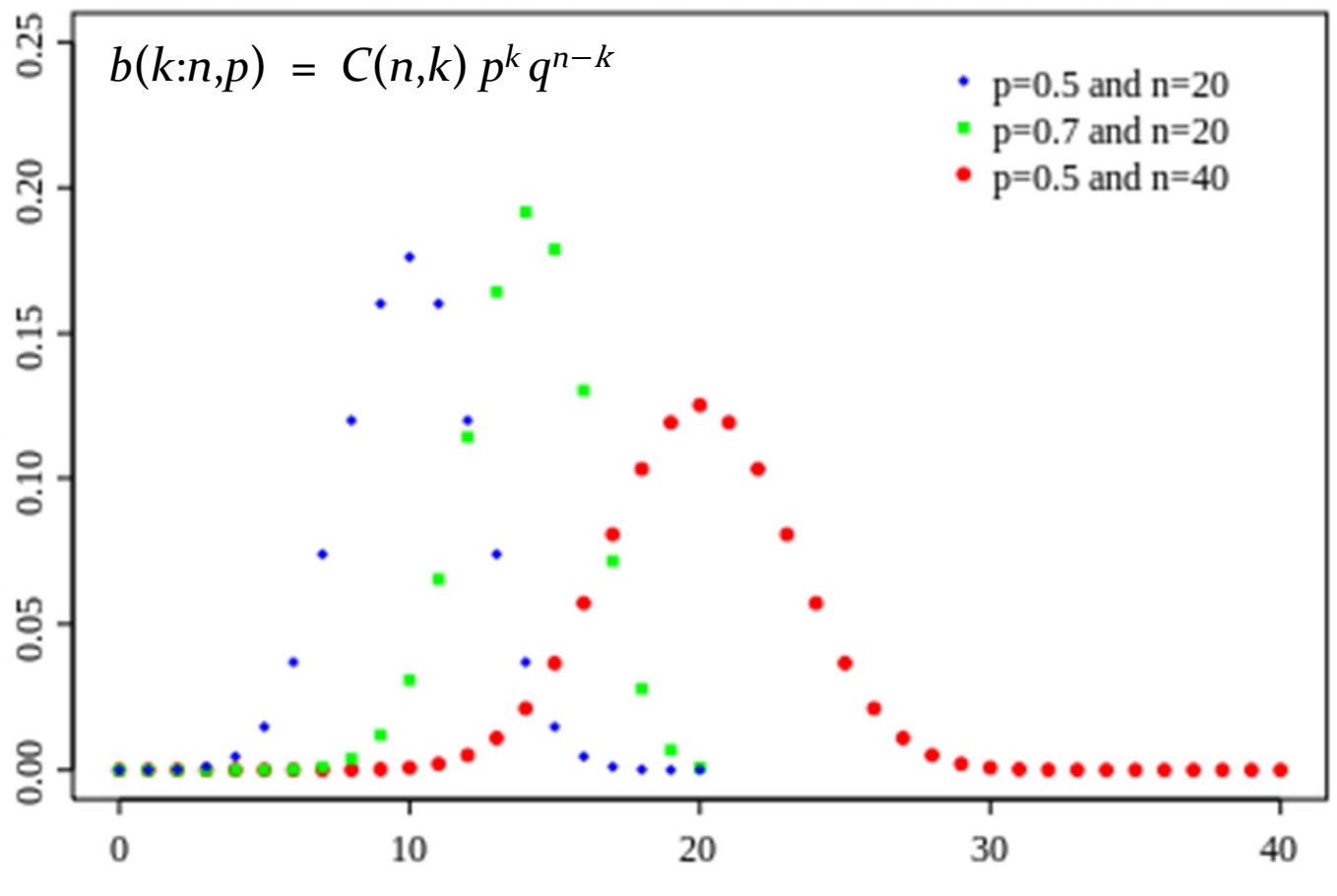
$$C(n,k)\, p^k\, q^{n-k}$$

Proof: The outcome of n Bernoulli trials is an n-tuple $(t_1,t_2,\ldots,t_n)$, where each is t_i either S (success) or F (failure). The probability of each outcome of n trials consisting of k successes and $n - $ k failures (in any given order) is $p^k q^{n-k}$. Because there are $C(n,k)$ n-tuples of Ss and Fs that contain exactly k Ss, the probability of k successes is $C(n,k)p^k q^{n-k}$. ◀

- We denote by $b(k{:}n,p)$ the probability of k successes in n independent Bernoulli trials with p the probability of success. Viewed as a function of k, $b(k{:}n,p)$ is the *binomial distribution*. By Theorem 2,

$$b(k{:}n,p) = C(n,k)p^k q^{n-k}.$$

Probability of *k* Successes in *n* Independent Bernoulli Trials.



source: WikipediA

Bayes' Theorem

Section 7.3

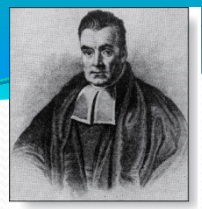
Section Summary

- Bayes' Theorem
- Generalized Bayes' Theorem
- Bayesian Spam Filters

Motivation for Bayes' Theorem

- Bayes' theorem allows us to use probability to answer questions such as the following:
  - Given that someone tests positive for having a particular disease, what is the probability that they actually do have the disease?
  - Given that someone tests negative for the disease, what is the probability, that in fact they do have the disease?
- Bayes' theorem has applications to medicine, law, artificial intelligence, engineering, and many diverse other areas.

Thomas Bayes
(1702-1761)



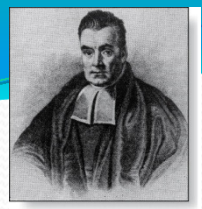
Bayes' Theorem

Bayes' 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:

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

- This helps when conditional probability $p(E|F)$ is easier to estimate compared to $p(F|E)$.

Thomas Bayes
(1702-1761)



Bayes' Theorem

- **Example**: We have two boxes. The first box contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects 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 selected a ball from the first box $p(F|E)$?
  - Let E be the event that Bob has chosen a red ball and F be the event that Bob has chosen the first box.

$$p(E|F) = \frac{7}{2+7} \qquad p(E|\overline{F}) = \frac{3}{4+3} \qquad p(F|E) = ?$$

  - By Bayes' theorem the probability that Bob has picked the first box is:

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

$$p(F|E) = \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{7/18}{38/63} = \frac{49}{76} \approx 0.645.$$

Derivation of Bayes' Theorem

- Recall the definition of the conditional probability $p(E|\mathrm{F})$:

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

- From this definition, it follows that:

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

continued →

Derivation of Bayes' Theorem

On the last slide we showed that

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

Equating the two formulas for p(E F) shows that

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

Solving for $p(E|F)$ and for $p(F|E)$ tells us that

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

continued →

Derivation of Bayes' Theorem

On the last slide we showed that:

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

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

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

because $E = E \cap S = E \cap (F \cup \overline{F}) = (E \cap F) \cup (E \cap \overline{F})$

and $(E \cap F) \cap (E \cap \overline{F}) = \emptyset$

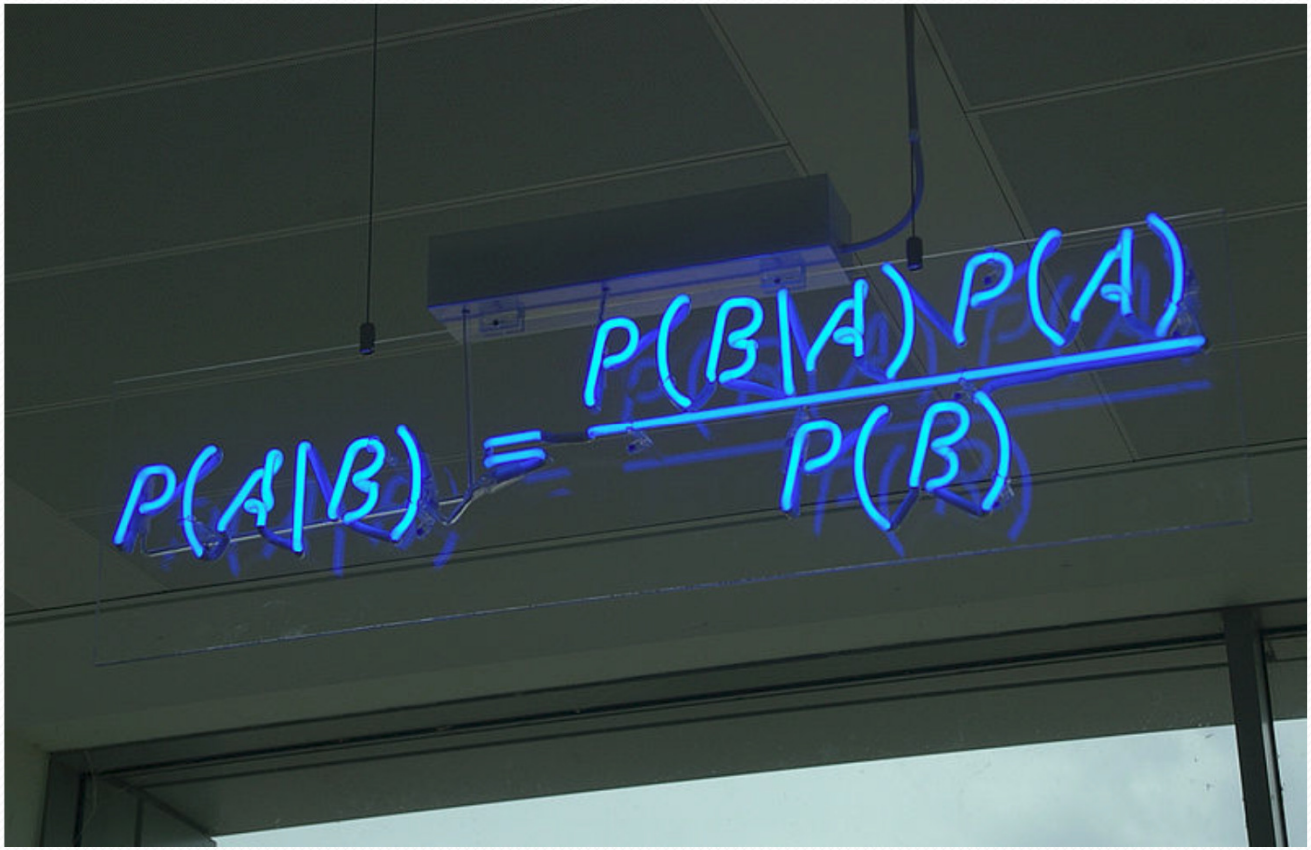
By the definition of conditional probability,

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

Hence,

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

Simple form of Bayes' Theorem



A blue neon sign at the Autonomy Corporation, Cambridge, showing the simple statement of Bayes' Theorem.

Interpretation of the Simple Form of Bayes' Theorem

- Bayes' Theorem links the degree of belied in a proposition before and after accounting for evidence.
- Proposition A, Evidence B
- $p(A)$ = prior probability (initial degree of belief in A)
- $p(A|B)$ = posterior probability (degree of belief in A after having accounted for B)
- $p(B|A) \,/\, p(B)$ = the support provided for A by B

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)}{p(B)} \cdot p(A)$$

- Suppose someone told you they had a conversation with a person on a train
- If you knew nothing else about this conversation, you would compute the probability that this person was a woman as 50%
- Now, suppose you were also told that the person had long hair
- Bayes' theorem can be used to calculate the probability that the person is a woman, given the additional knowledge we have.

How to solve the train problem

- W = event that the conversation partner is a woman
- L = the conversation partner has long hair
- Suppose we know that 75% of women have long hair and 15% of men have long hair. These are statistics that can be directly estimated.

$$p(L|W) = 0.75 \qquad p(L|\overline{W}) = 0.15$$

- What about $p(W|L)$?

$$p(W|L) = \frac{p(L|W)p(W)}{p(L|W)p(W) + p(L|\overline{W})p(\overline{W})}$$

$$p(W|L) = \frac{0.75 \cdot 0.5}{0.75 \cdot 0.5 + 0.15 \cdot 0.5} = 0.83$$

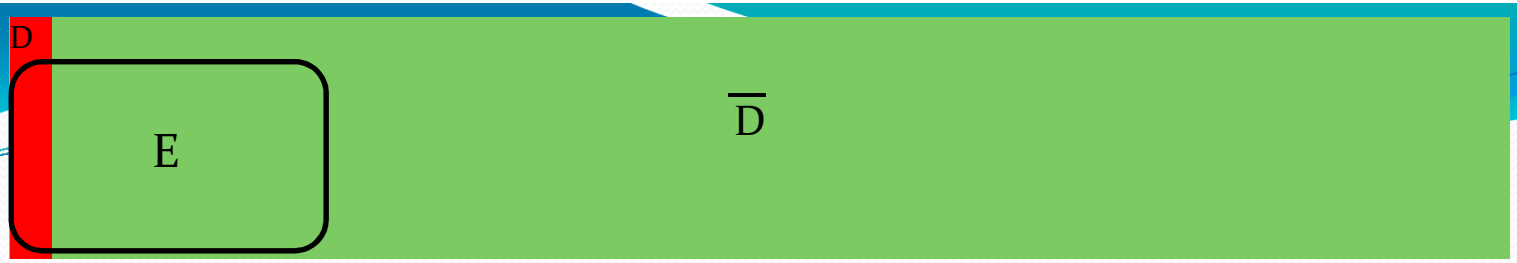
Applying Bayes' Theorem

Example: Suppose that one person in 100,000 has a particular disease. There is a test for the disease that gives a positive result 99% of the time when given to someone with the disease. When given to someone without the disease, 99.5% of the time it gives a negative result. Find

a) the probability that a person who test positive has the disease.

b) the probability that a person who test negative does not have the disease.

- Should someone who tests positive be worried?



Applying Bayes' Theorem

Solution: Let D be the event that the person has the disease, and E be the event that this person tests positive. We need to compute $p(D|E)$ from $p(\mathrm{D})$, $p(E|D)$, $p(\ E\ |\overline{D})$, $p(\overline{D})$.

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

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

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

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

$$\approx 0.002$$

Can you use this formula to explain why the resulting probability is surprisingly small?

So, don't worry too much, if your test for this disease comes back positive.

Applying Bayes' Theorem

- What if the result is negative?

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

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

$$\approx 0.9999999$$

So, the probability you have the disease if you test negative is

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

- So, it is extremely unlikely you have the disease if you test negative.

Generalized Bayes' Theorem

Generalized Bayes' Theorem: Suppose that E is an event from a sample space S and that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i}^{n} F_i = S$.

Assume that $p(E) \neq 0$ for $i = 1, 2, \ldots, n$. Then

$$p(F_j|E) = \frac{p(E|F_j)p(F_j)}{\sum_{i=1}^{n} p(E|F_i)p(F_i)}.$$

Exercise 17 *asks for the proof.*

Bayesian Spam Filters

- How do we develop a tool for determining whether an email is likely to be spam?
- If we have an initial set ***B****(ad)* of spam messages and set ***G****(ood)* of non-spam messages. We can use this information along with Bayes' law to predict the probability that a new email message is spam.
- We look at a particular word w, and count the number of times that it occurs in B and in G; $n_B(w)$ and $n_G(w)$.
  - Estimated probability that spam email contains w:
    $$p(w) = n_B(w)/|B|$$
  - Estimated probability that non-spam email contains w:
    $$q(w) = n_G(w)/|G|$$

continued →

Bayesian Spam Filters

- Let S be the event that the message is spam, and E be the event that the message contains the word w.
- Using Bayes' Rule,

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

Assuming that it is equally likely that an arbitrary message is spam and is not spam; i.e., $p(S) = ½$.

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

Note: If we have data on the frequency of spam messages, we can obtain a better estimate for p(s). (*See Exercise* 22.)

Using our empirical estimates of
$p(w) = p(E \mid S)$
$q(w) = p(E \mid \overline{S})$.

$$r(w) = \frac{p(w)}{p(w) + q(w)}$$

$r(w)$ estimates the probability that the message is spam. We can class the message as spam if $r(w)$ is above a threshold.

Bayesian Spam Filters

Example: We find that the word "Rolex" occurs in 250 out of 2000 spam messages and occurs in 5 out of 1000 non-spam messages. Estimate the probability that an incoming message is spam. Suppose our threshold for rejecting the email is 0.9.

Solution: $p(Rolex) = 250/2000 = .0125$ and $q(Rolex) = 5/1000 = 0.005$.

$$r(Rolex) = \frac{p(Rolex)}{p(Rolex) + q(Rolex)} = \frac{0.125}{0.125 + .005} = \frac{0.125}{0.125 + .005} \approx 0.962$$

We class the message as spam and reject the email!

Bayesian Spam Filters using Multiple Words

- Accuracy can be improved by considering more than one word as evidence.
- Consider the case where E_1 and E_2 denote the events that the message contains the words w_1 and w_2 respectively.
- We make the simplifying assumption that the events are independent. And again we assume that $p(S) = ½$.

$$p(S|E_1 \cap E_2) = \frac{p(E_1|S)p(E_2|S)}{p(E_1|S)p(E_2|S) + p(E_1|\overline{S})p(E_2|\overline{S})}$$

$$r(w_1, w_2) = \frac{p(w_1)p(w_2)}{p(w_1)p(w_2) + q(w_1)q(w_2)}$$

Bayesian Spam Filters using Multiple Words

Example: We have 2000 spam messages and 1000 non-spam messages. The word "stock" occurs 400 times in the spam messages and 60 times in the non-spam. The word "undervalued" occurs in 200 spam messages and 25 non-spam.

Solution: $p(stock) = 400/2000 = .2$, $q(stock) = 60/1000 = .06$, $p(undervalued) = 200/2000 = .1$, $q(undervalued) = 25/1000 = .025$

$$r(stock, undervalued) = \frac{p(stock)p(undervalued)}{p(stock)p(undervalued) + q(stock)q(undervalued)}$$

$$= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)} \approx 0.930$$

If our threshold is .9, we class the message as spam and reject it.

Bayesian Spam Filters using Multiple Words

- In general, the more words we consider, the more accurate the spam filter. With the independence assumption if we consider *k* words:

$$p(S|\bigcap_{i=1}^{k} E_i) = \frac{\Pi_{i=1}^{k} p(E_i|S)}{\Pi_{i=1}^{k} p(E_1|S) + \Pi_{i=1}^{k} p(E_i|\overline{S})}$$

$$r(w_1, w_2, \ldots w_n) = \frac{\Pi_i^k p(w_i)}{\Pi_{i=1}^{k} p(w_i) + \Pi_{i=1}^{k} q(w_i)}$$

We can further improve the filter by considering pairs of words as a single block or certain types of strings.