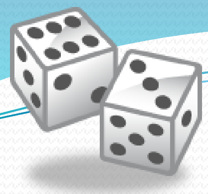
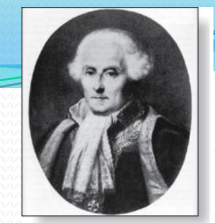
Discrete Probability

Chapter 7



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.

- Here is how Laplace defined the probability of an event:

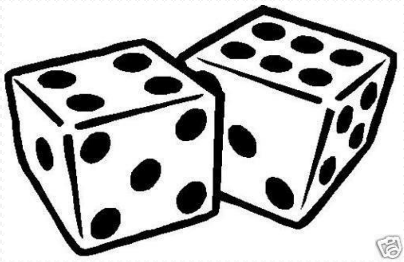
  Definition: If S is a finite sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the *probability* of E is $p(E) = |E|/|S|$.

- 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 10000 ways to pick four digits.

- Since there is only 1 way to pick the correct digits, the probability of winning the large prize is $1/10000 = 0.0001$.

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. 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 $50 \cdot 49 \cdot 47 \cdot 46 = 254{,}251{,}200$ 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

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

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

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

◀

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_1 be the event that the integer is divisible by 2 and E_2 be the event that it is divisible 5? Then the event that the integer is divisible by 2 or 5 is $E_1 \cup E_2$ and $E_1 \cap E_2$ is the event that it is divisible by 2 and 5.

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$
$$= 50/100 + 20/100 - 10/100 = 3/5.$$

Probability Theory

Section 7.2

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. $0 \leq p(s) \leq 1$ 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 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.

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

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

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 bit strings of length 4 is 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

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

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

Example: Suppose E is the event that a randomly generated bit string of length four begins with 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 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

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

Pairwise and Mutual Independence

Definition: The events E_1, E_2, …, E_n are *pairwise independent* if and only if $p(E_i \cap E_j) = p(E_i)\,\mathrm{p}(E_j)$ 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

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

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.*
- 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 $k - 1$ failures (in any 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,

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

Random Variables

Definition: A *random variable* is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

Example: Suppose that a coin is flipped three times. Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome. Then $X(t)$ takes on the following values:

$$X(HHH) = 3, X(TTT) = 0,$$

$$X(HHT) = X(HTH) = X(THH) = 2,$$

$$X(TTH) = X(THT) = X(HTT) = 1.$$

The Famous Birthday Problem

- Find the number of people needed in a room to ensure that the probability of at least two of them having the same birthday is more than ½ has a surprising answer, which we now find.

Solution: We assume that all birthdays are equally likely and that there are 366 days in the year. First, we find the probability p_n that n people have different birthdays:

$$p_n = (365/366)(364/366)\cdots(367 - n)/366.$$

Now, imagine the people entering the room one by one. The probability that at least two have the same birthday is $1 - p_n$.

Checking various values for n with computation help tells us that for $n = 22$, $1 - p_n \approx 0.457$, and for $n = 23$, $1 - p_n \approx 0.506$.
Consequently, a minimum number of 23 people are needed so that that the probability that at least two of them have the same birthday is greater than 1/2.

Random Variables

Definition: The *distribution* of a random variable X on a sample space S is the set of pairs $(r, p(X = r))$ for all $r \in X(S)$, where $p(X = r)$ is the probability that X takes the value r.

Example: Suppose that a coin is flipped three times. Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome. Then $X(t)$ takes on the following values:

$X(HHH) = 3, X(TTT) = 0,$

$X(HHT) = X(HTH) = X(THH) = 2,$

$X(TTH) = X(THT) = X(HTT) = 1.$

Each of the eight possible outcomes has probability 1/8. So, the distribution of $X(t)$ is $p(X = 3) = 1/8$, $p(X = 2) = 3/8$, $p(X = 1) = 3/8$, and $p(X = 0) = 1/8$.

Bayes' Theorem

Section 7.3

Motivation

- Suppose that one person in 100000 has a particular disease. There is a test for the disease that gives a positive result 99% when given to someone with the disease. When given to someone without the disease, 99.5% it gives a negative result.

  Find

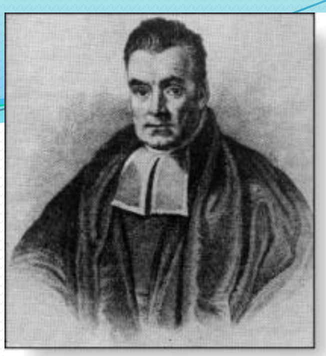
  a) the probability that a person who tests positive has the disease actually

  b) the probability that a person who tests negative does not have the disease actually

- Should someone who tests positive be worried?

Bayes' Theorem

Thomas Bayes
(1702-1761)



- 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:

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

- Bayes' theorem has applications to medicine, law, artificial intelligence, engineering, and many diverse other areas.

Bayes' Theorem

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

Ex. We have two boxes. The first box contains 2 green balls and 7 red balls. The second contains 4 green balls and 3 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?

- Let *E* be the event that Bob has chosen a red ball and
- Let *F* be the event that Bob has chosen the first box.

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

Derivation of Bayes' Theorem

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

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

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

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

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

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

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

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

Applying Bayes' Theorem (1/3)

Ex. 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.

Then, what is

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 (2/3)

Let D be the event that the person has the disease

Let E be the event that this person tests positive.

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

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

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

$$= \frac{(0.99)(0.00001)}{(0.99)(0.00001) + (0.005)(0.99999)} \approx 0.002$$

Applying Bayes' Theorem (3/3)

What if the result is negative?

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

$$= \frac{(0.995)(0.99999)}{(0.995)(0.99999) + (0.01)(0.00001)} \approx 0.9999999$$

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

Generalized Bayes' Theorem

- **Generalized Bayes' Theorem**: Suppose that E is an event from a sample space S and that F_1, F_2, …, 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

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

Bayesian Spam Filters

- If we have a set B of spam messages and a set G 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 $n_G(w)$, respectively
  - Estimated probability that a spam email contains w: $p(w) = n_B(w)/|B|$
  - Estimated probability that a 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,

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

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

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

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

$r(w)$ 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 $p(S) = ½$.

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

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

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

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.