

Discrete Probability

Chapter 7

Because learning changes everything.™

© 2019 McGraw-Hill Education. All rights reserved. Authorized only for instructor use in the classroom. No reproduction or further distribution permitted without the prior written consent of McGraw-Hill Education.

An Introduction to Discrete Probability

Section 7.1



Probability of an Event₁



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 all 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 Definition₁

Example: An urn contains 4 blue balls and 5 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. 6 of these sum to 7. Hence, the probability of obtaining a 7 is $6/36 = 1/6$.

Applying Laplace's Definition₂

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. 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 Definition₃

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

Can you work out the probability of winning the lottery with the biggest prize where you live?

Applying Laplace's Definition₄

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 $50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 = 254,251,200$ 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 Events₁

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 $|\overline{E}| = |S| - |E|$,

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

The Probability of Complements and Unions of Events₂

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 \bar{E} is the event that all of the bits are 1s. The size of the sample space S is 2^{10} . Hence,

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

© 2019 McGraw-Hill Education

The Probability of Complements and Unions of Events₄

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.

It follows that:

$$\begin{aligned} p(E_1 \cup E_2) &= p(E_1) + p(E_2) - p(E_1 \cap E_2) \\ &= 50/100 + 20/100 - 10/100 = 3/5. \end{aligned}$$

© 2019 McGraw-Hill Education

The Probability of Complements and Unions of Events₃

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 $|A \cup B| = |A| + |B| - |A \cap B|$, it follows that

$$\begin{aligned} p(E_1 \cup E_2) &= \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|} \\ &= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|} \\ &= p(E_1) + p(E_2) - p(E_1 \cap E_2). \end{aligned}$$

© 2019 McGraw-Hill Education

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.
- If the initial pick is wrong (with probability $2/3$). Since the game show host always opens a door that does not have the prize, if you switch the probability of winning becomes $2/3$.

© 2019 McGraw-Hill Education

Probability Theory

Section 7.2

© 2019 McGraw-Hill Education

Assigning Probabilities₂

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: For a fair coin, we have $p(H) = p(T) = \frac{1}{2}$.

For a biased coin, 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 $p(H) = 2/3$.

© 2019 McGraw-Hill Education

Assigning Probabilities₁

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.

© 2019 McGraw-Hill Education

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 $p(H) = p(T) = 1/2$.

© 2019 McGraw-Hill Education

Probability of an Event₂

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.

Example

Example: Suppose that a die 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 die?

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

$$p(1) = p(2) = p(4) = p(5) = p(6) = 1/7.$$

$$\text{Hence, } p(E) = p(1) + p(3) + p(5) =$$

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

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, \dots 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 Probability₁

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 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, $p(F) = 8/16 = 1/2$.

Hence,

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

Independence₁

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 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,

$$p(E) = p(F) = 8/16 = 1/2.$$

- 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

$$p(E \cap F) = 1/4 = (1/2)(1/2) = p(E)p(F)$$

Conditional Probability₂

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,

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

Independence₂

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 $p(F) = 3/4$ and $p(E \cap F) = 1/4$. The events E and F are not independent since

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

Pairwise and Mutual Independence

Definition: The events E_1, E_2, \dots, 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

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

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

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

Bernoulli Trials₁

James Bernoulli
(1654 – 1705)



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 Trials₂

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, \dots, 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 order) is $p^k q^{n-k}$. Because there are $C(n, k)$ n -tuples of S s and F s that contain exactly k S s, 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 Variables₁

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.

A random variable is a function. It is not a variable, and it is not random!

In the late 1940s W. Feller and J.L. Doob flipped a coin to see whether both would use “random variable” or the more fitting “chance variable.” Unfortunately, Feller won and the term “random variable” has been used ever since.

The Famous Birthday Problem

The puzzle of finding 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 $\frac{1}{2}$ 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 the n people all have different birthdays.

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

- The probability that the birthday of the second person is different from that of the first is $365/366$.
- The probability that the birthday of the third person is different from the other two, when these have two different birthdays, is $364/366$.
- In general, the probability that the j th person has a birthday different from the birthdays of those already in the room, assuming that these people all have different birthdays, is $(366 - (j - 1))/366 = (367 - j)/366$.
- Hence, $p_n = (365/366)(364/366) \cdots (367 - n)/366$.
- Therefore, $1 - p_n = 1 - (365/366)(364/366) \cdots (367 - n)/366$.

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 the probability that at least two of them have the same birthday is greater than $1/2$.

Random Variables₂

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 $p(X = 0) = 1/8$.

Monte Carlo Algorithms

Algorithms that make random choices at one or more steps are called *probabilistic algorithms*.

Monte Carlo algorithms are probabilistic algorithms used to answer decision problems, which are problems that either have “true” or “false” as their answer.

- A Monte Carlo algorithm consists of a sequence of tests. For each test the algorithm responds “true” or “unknown.”
- If the response is “true,” the algorithm terminates with the answer is “true.”
- After running a specified sequence of tests where every step yields “unknown,” the algorithm outputs “false.”
- The idea is that the probability of the algorithm incorrectly outputting “false” should be very small as long as a sufficient number of tests are performed.

Bayes' Theorem

Section 7.3

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:

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

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.

- Let E be the event that Bob has chosen a red ball and F be the event that Bob has chosen the first box.
- By Bayes' theorem the probability that Bob has picked the first box is:

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

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.

Derivation of Bayes' Theorem₁

Recall the definition of the conditional probability $p(E | F)$:

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

From this definition, it follows that:

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

Derivation of Bayes' Theorem₂

On the last slide we showed that

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

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

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

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

Applying Bayes' Theorem₁

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

- the probability that a person who test positive has the disease.
- the probability that a person who test negative does not have the disease.

Should someone who tests positive be worried?

Derivation of Bayes' Theorem

On the last slide we showed that: $p(F|E) = \frac{p(E|F)p(F)}{p(E)}$

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

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

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

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

By the definition of conditional probability,?

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

Hence, $p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\bar{F})p(\bar{F})}$

Applying Bayes' Theorem₂

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 $p(D)$, $p(E|D)$, $p(E|\bar{D})$, $p(\bar{D})$.

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

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

$$\begin{aligned} p(D|E) &= \frac{p(E|D)p(D)}{p(E|D)p(D) + p(E|\bar{D})p(\bar{D})} \\ &= \frac{(0.99)(0.00001)}{(0.99)(0.00001) + (0.005)(0.99999)} \\ &\approx 0.002 \end{aligned}$$

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

Applying Bayes' Theorem₃

What if the result is negative?

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

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

$$\approx 0.9999999$$

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

$$p(D | \bar{E})$$

$$\approx 1 - 0.9999999$$

$$= 0.0000001.$$

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

Bayesian Spam Filters₁

How do we develop a tool for determining whether an email is likely to be spam?

If we have an initial set B of spam messages and 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 $n_G(w)$.

- Estimated probability that a spam message contains w is:
 $p(w) = n_B(w) / |B|$
- Estimated probability that a message that is not spam contains w is:
 $q(w) = n_G(w) / |G|$

Generalized Bayes' Theorem

Generalized Bayes' 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 $\bigcup_i F_i = S$

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

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

Bayesian Spam Filters₂

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

Assuming that it is equally likely that an arbitrary message is spam and is not spam; i.e., $p(S) = \frac{1}{2}$.

Using our empirical estimates of $p(E | S)$ and $p(E | \bar{S})$.

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

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

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

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

Bayesian Spam Filters₃

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: $p(\text{Rolex}) = 250/2000 = .0125$ and $q(\text{Rolex}) = 5/1000 = 0.005$.

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

We class the message as spam and reject the email!

Bayesian Spam Filters using Multiple Words₂

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

$$\begin{aligned} r(\text{stock, undervalued}) &= \frac{p(\text{stock}) p(\text{undervalued})}{p(\text{stock}) p(\text{undervalued}) + q(\text{stock}) q(\text{undervalued})} \\ &= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)} \approx 0.930 \end{aligned}$$

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

Bayesian Spam Filters using Multiple Words₁

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 $p(S) = \frac{1}{2}$** .

$$\begin{aligned} p(S | E_1 \cap E_2) &= \frac{p(E_1 | S) p(E_2 | S)}{p(E_1 | S) p(E_2 | S) + p(E_1 | \bar{S}) p(E_2 | \bar{S})} \\ r(w_1, w_2) &= \frac{p(w_1) p(w_2)}{p(w_1) p(w_2) + q(w_1) q(w_2)} \end{aligned}$$

Bayesian Spam Filters using Multiple Words₃

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

$$\begin{aligned} p\left(S | \bigcap_{i=1}^k E_i\right) &= \frac{\prod_{i=1}^k p(E_i | S)}{\prod_{i=1}^k p(E_i | S) + \prod_{i=1}^k p(E_i | \bar{S})} \\ r(w_1, w_2, \dots, w_n) &= \frac{\prod_{i=1}^k p(w_i)}{\prod_{i=1}^k p(w_i) + \prod_{i=1}^k q(w_i)} \end{aligned}$$

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

Expected Value and Variance

Section 7.4

© 2019 McGraw-Hill Education

Expected Value₁

Definition: The *expected value* (or *expectation* or *mean*) of the random variable $X(s)$ on the sample space S is equal to

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

Example-Expected Value of a Die: Let X be the number that comes up when a fair die is rolled. What is the expected value of X ?

Solution: The random variable X takes the values 1, 2, 3, 4, 5, or 6. Each has probability $1/6$. It follows that

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6 = \frac{21}{6} = \frac{7}{2}.$$

© 2019 McGraw-Hill Education

Expected Value₂

Theorem 1: If X is a random variable and $p(X = r)$ is the probability that $X = r$, so that

$$p(X = r) = \sum_{s \in S, X(s)=r} p(s), \text{ then}$$
$$E(X) = \sum_{r \in X(S)} p(X = r)r.$$

Proof: Suppose that X is a random variable with range $X(S)$ and let $p(X = r)$ be the probability that X takes the value r . Consequently, $p(X = r)$ is the sum of the probabilities of the outcomes s such that $X(s) = r$. Hence,

$$E(X) = \sum_{r \in X(S)} p(X = r)r.$$

© 2019 McGraw-Hill Education

Expected Value₃

Theorem 2: The expected number of successes when n mutually independent Bernoulli trials are performed, where, the probability of success on each trial, $p = np$.

Proof: Let X be the random variable equal to the number of success in n trials. By Theorem 2 of section 7.2, $p(X = k) = C(n, k)p^kq^{n-k}$. Hence,

$$E(X) = \sum_{k=1}^n kp(X = k) \text{ by Theorem 1}$$

© 2019 McGraw-Hill Education

Expected Value₄

$$\begin{aligned}
 E(X) &= \sum_{k=1}^n kp(X=k) && \text{from previous page} \\
 &= \sum_{k=1}^n kC(n,k)p^kq^{n-k} && \text{by Theorem 2 in Section 7.2} \\
 &= \sum_{k=1}^n nC(n-1,k-1)p^kq^{n-k} && \text{by Exercise 21 in Section 6.4} \\
 &= np \sum_{k=1}^n C(n-1,k-1)p^{k-1}q^{n-k} && \text{factoring np from each term} \\
 &= np \sum_{j=0}^{n-1} C(n-1,j)p^jq^{n-1-j} \\
 &= np(p+q)^{n-1} && \text{by the binomial theorem} \\
 &= np. && \text{because } p+q=1
 \end{aligned}$$

We see that the expected number of successes in n mutually independent Bernoulli trials is np .

Linearity of Expectations₂

Expected Value in the Hatcheck Problem: A new employee started a job checking hats, but forgot to put the claim check numbers on the hats. So, the n customers just receive a random hat from those remaining. What is the expected number of hat returned correctly?

Solution: Let X be the random variable that equals the number of people who receive the correct hat. Note that $X = X_1 + X_2 + \dots + X_n$,

where $X_i = 1$ if the i th person receives the hat and $X_i = 0$ otherwise.

- Because it is equally likely that the checker returns any of the hats to the i th person, it follows that the probability that the i th person receives the correct hat is $1/n$. Consequently (by Theorem 1), for all i

$$E(X_i) = 1 \cdot p(X_i = 1) + 0 \cdot p(X_i = 0) = 1 \cdot 1/n + 0 = 1/n.$$

- By the linearity of expectations (Theorem 3), it follows that:

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_n) = n \cdot \frac{1}{n} = 1.$$

Consequently, the average number of people who receive the correct hat is exactly 1. (Surprisingly, this answer remains the same no matter how many people have checked their hats!)

Linearity of Expectations₁

The following theorem tells us that expected values are linear. For example, the expected value of the sum of random variables is the sum of their expected values.

Theorem 3: If $X_i, i = 1, 2, \dots, n$ with n a positive integer, are random variables on S , and if a and b are real numbers, then

- $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$
- $E(aX + b) = aE(X) + b.$

Linearity of Expectations₃

Expected Number of Inversions in a Permutation: The ordered pair (i, j) is an *inversion* in a permutation of the first n positive integers if $i < j$, but j precedes i in the permutation.

Example: There are six inversions in the permutation of 3, 5, 1, 4, 2

(1, 3), (1, 5), (2, 3), (2, 4), (2, 5), (4, 5).

Find the average number of inversions in a random permutation of the first n integers.

Solution: Let $I_{i,j}$ be the random variable on the set of all permutations of the first n positive integers with $I_{i,j} = 1$ if (i, j) is an inversion of the permutation and $I_{i,j} = 0$ otherwise. If X is the random variable equal to the number of inversions in the permutation, then $X = \sum_{1 \leq i < j \leq n} I_{i,j}.$

- Since it is equally likely for i to precede j in a randomly chosen permutation as it is for j to precede i , we have: $E(I_{i,j}) = 1 \cdot p(I_{i,j} = 1) + 0 \cdot p(I_{i,j} = 0) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}$
- Because there are C_n^2 pairs i and j with $1 \leq i < j \leq n$, by the linearity of expectations (Theorem 3), we have:
$$E(X) = \sum_{1 \leq i < j \leq n} E(I_{i,j}) = \binom{n}{2} \cdot \frac{1}{2} = \frac{n(n-1)}{2} \cdot \frac{1}{2}.$$

Consequently, it follows that there is an average of $\frac{n(n-1)}{4}$ inversions in a random permutation of the first n positive integers.

The Geometric Distribution

Definition 2: A random variable X has *geometric distribution* with parameter p if $p(X = k) = (1 - p)^{k-1}p$ for $k = 1, 2, 3, \dots$, where p is a real number with $0 \leq p \leq 1$.

Theorem 4: If the random variable X has the geometric distribution with parameter p , then $E(X) = 1/p$.

Example: Suppose the probability that a coin comes up tails is p . What is the expected number of flips until this coin comes up tails?

- The sample space is $\{T, HT, HHT, HHHT, \dots\}$.
- Let X be the random variable equal to the number of flips in an element of the sample space; $X(T) = 1$, $X(HT) = 2$, $X(HHT) = 3$, etc.
- By Theorem 4, $E(X) = 1/p$. $\sum_{i=1}^{\infty} i \times (1 - p)^{i-1}p = \frac{1}{p}$

Variance₁

Deviation: The *deviation* of X at $s \in S$ is $X(s) - E(X)$, the difference between the value of X and the mean of X .

Definition 4: Let X be a random variable on the sample space S . The *variance* of X , denoted by $V(X)$ is

$$V(X) = \sum_{s \in S} (X(s) - E(X))^2 p(s).$$

That is $V(X)$ is the weighted average of the square of the deviation of X . The standard deviation of X , denoted by $\sigma(X)$ is defined to be $\sqrt{V(X)}$.

Theorem 6: If X is a random variable on a sample space S , then $V(X) = E(X^2) - E(X)^2$.

$$\begin{aligned} V(X) &= \sum (X(s) - E(X))^2 p(s) = \sum (X(s)^2 - 2E(X)X(s) + E(X)^2) p(s) \\ &= \sum X(s)^2 p(s) - 2E(X) \sum X(s) p(s) + E(X)^2 \sum p(s) = E(X^2) - E(X)^2 \end{aligned}$$

Corollary 1: If X is a random variable on a sample space S and $E(X) = \mu$, then $V(X) = E((X - \mu)^2)$.

$$E((X - \mu)^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - E(X)^2 = V(X)$$

Independent Random Variables

Definition 3: The random variables X and Y on a sample space S are independent if

$$p(X = r_1 \text{ and } Y = r_2) = p(X = r_1) \cdot p(Y = r_2).$$

Theorem 5: If X and Y are independent variables on a sample space S , then $E(XY) = E(X)E(Y)$.

see text for full proof

Variance₂

Example: What is the variance of the random variable X , where $X(t) = 1$ if a Bernoulli trial is a success and $X(t) = 0$ if it is a failure, where p is the probability of success and q is the probability of failure?

Solution: Because X takes only the values 0 and 1, it follows that $X^2(t) = X(t)$. Hence,

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p) = pq.$$

Variance of the Value of a Die: What is the variance of a random variable X , where X is the number that comes up when a fair die is rolled?

Solution: We have $V(X) = E(X^2) - E(X)^2$. In an earlier example, we saw that $E(X) = 7/2$. Note that

$$E(X^2) = 1/6(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = 91/6.$$

We conclude that $V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$.

Variance

Irenée-Jules Bienaymé
(1796-1878)



Bienaymé's Formula: If X and Y are two independent random variables on a sample space S , then $V(X + Y) = V(X) + V(Y)$. Furthermore, if $X_i, i = 1, 2, \dots, n$, with n a positive integer, are pairwise independent random variables on S , then

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n).$$

Example: Find the variance of the number of successes when n independent Bernoulli trials are performed, where on each trial, p is the probability of success and q is the probability of failure.

Solution: Let X_i be the random variable with $X_i((t_1, t_2, \dots, t_n)) = 1$ if trial t_i is a success and $X_i((t_1, t_2, \dots, t_n)) = 0$ if it is a failure. Let $X = X_1 + X_2 + \dots + X_n$. Then X counts the number of successes in the n trials.

- By Bienaymé's Formula, it follows that $V(X) = V(X_1) + V(X_2) + \dots + V(X_n)$.
- By the previous example, $V(X_i) = pq$ for $i = 1, 2, \dots, n$.

Hence, $V(X) = npq$.

Chebyshev's Inequality

Chebyshev's Inequality: Let X be a random variable on a sample space S with probability function p . If r is a positive real number, then

$$p(|X(s) - E(X)| \geq r) \leq V(X) / r^2.$$

Example: Suppose that X is a random variable that counts the number of tails when a fair coin is tossed n times. Note that X is the number of successes when n independent Bernoulli trials, each with probability of success $1/2$ are done. Hence, $E(X) = n/2$ and $V(X) = n/4$.

By Chebyshev's inequality with $r = \sqrt{n}$,

$$p(|X(s) - n/2| \geq \sqrt{n}) \leq (n/4) / (\sqrt{n})^2 = 1/4.$$

This means that the probability that the number of tails that come up on n tosses deviates from the mean, $n/2$, by more than \sqrt{n} is no larger than $1/4$.