Intro to Probability

Al Nosedal.
University of Toronto.

Summer 2019

My momma always said: "Life was like a box of chocolates. You never know what you're gonna get."

Forrest Gump.

Example

Random experiment: Rolling a fair die.
List of all possible outcomes $= \{ 1, 2, 3, 4, 5, 6 \}$.
Probability of 6 $= \frac{1}{6}$.

Definitions

The **sample space S** of a random phenomenon is the set of all possible outcomes.

An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.

Probability is a numerical measure of the likelihood that an event will occur.

Another example

In each of the following situations, describe a sample space S for the random phenomenon.

a. A basketball player shoots three free throws. You record the sequence of hits and misses.

b. A basketball player shoots three free throws. You record the number of baskets she makes.

Solution

H=hit and M=miss

a. S={(H,H,H),(H,H,M),(H,M,H),(H,M,M),
(M,H,H),(M,H,M),(M,M,H),(M,M,M)}

b. S={0,1,2,3}

Problem

How many teams consisting of two individuals can be selected from a group of five individuals?

Solution

{A,B},{A,C},{A,D},{A,E},
{B,C},{B,D},{B,E},
{C,D},{C,E},
{D,E}.
10 teams

Combinations

The number of combinations of N objects taken n at a time is

$$C_n^N = \frac{N!}{n!(N-n)!}$$

where

$$N! = N(N-1)(N-2)...(2)(1)$$

$$n! = n(n-1)(n-2)...(2)(1)$$

and by definition, $0! = 1$.

Problem

How many different committees consisting of a president and a secretary can be selected from a group of five individuals?

Solution

(A,B),(A,C),(A,D),(A,E)
(B,A),(B,C),(B,D),(B,E)
(C,A),(C,B),(C,D),(C,E)
(D,A),(D,B),(D,C),(D,E)
(E,A),(E,B),(E,C),(E,D)
20 committees.

Permutations

The number of permutations on N objects taken n at a time is given by

$$P_n^N = \frac{N!}{(N-n)!}$$

Problem

How many ways can three items be selected from a group of six items?

Problem

How many permutations of three items can be selected from a group of six?

Problem

The Powerball lottery is played twice each week in 28 states, the Virgin Islands, and the District of Columbia. To play Powerball a participant must purchase a ticket and then select five numbers from the digits 1 through 55 and a Powerball number from the digits 1 through 42. To determine the winning numbers for each game, lottery officials draw five white balls out of a drum with 55 white balls, and one red ball out of a drum with 42 red balls. To win the jackpot, a participant's numbers must match the numbers on the five white balls in any order and the number on the red Powerball.

Problem

Eight coworkers at the ConAgra Foods plant in Lincoln, Nebraska, claimed the record \$365 million jackpot on February 18, 2006, by matching the numbers 15-17-43-44-49 and the Powerball number 29. A variety of other cash prizes are awarded each time the game is played. For instance, a prize of \$ 200,000 is paid if the participant's five numbers match the numbers on the five white balls.

a. Compute the number of ways the first five numbers can be selected.

b. What is the probability of winning a prize of \$200,000 by matching the numbers on the five white balls?

c. What is the probability of winning the Powerball jackpot?

Solution

a.

$$\frac{55!}{5!(55-5)!} = 3,478,761$$

b.

$$\frac{1}{3,478,761}$$

c.

$$\text{Number of choices} = (3,478,761)(42) = 146,107,962$$

$$\text{Probability of winning} = \frac{1}{146,107,962}$$

Problem

Consider the experiment of rolling a fair pair of dice. Suppose that we are interested in the sum of the face values showing on the dice.

a. How many sample points are possible?

b. List the sample points.

c. What is the probability of obtaining a value of 7?

d. What is the probability of obtaining a value of 9 or greater?

e. Because each roll has six possible even values (2,4,6,8,10 and 12) and only five possible odd values (3,5,7,9, and 11), the dice should show even values more often than odd values. Do you agree with this statement? Explain.

Problem (solution)

a. 36

b. (1,1),(1,2),(1,3),(1,4),(1,5),(1,6)
(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)
(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)
(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)
(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)
(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)

c. $P(7)=P\{(6,1)\text{or}(5,2)\text{or}(4,3)\text{or}(3,4)\text{or}(2,5)\text{or}(1,6)\}$
$=\frac{6}{36}$

d. $P(9 \text{ or greater})= \frac{10}{36}$

e. $P(\text{even})=P(2)+P(4)+P(6)+P(8)+P(10)+P(12)=\frac{18}{36}$
$P(\text{odd})=P(1)+P(3)+P(5)+P(7)+P(9)+P(11)=\frac{18}{36}$

Example: Blood Type

In the United States, 44% of the population has type O blood, 42% are type A, 10% are type B, and ?% are type AB. Consider choosing someone at random and determining the person's blood type.

a) What is the probability that the person will have type AB blood?

b) What is the probability that the person will *not* have type O blood?

c) What is the probability that the person will have either type A or type O blood?

Solution

a. Probability that the person will have type AB blood = P(AB) = $1 - 0.96 = 0.04$

b. P(Not O) = P(O^c) = 1 - P(O) = $1 - 0.44 = 0.56$

c. P(A or O) = $0.42 + 0.44 = 0.86$

Probability Rules

1. The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.
2. If S is the sample space in a probability model, then $P(S) = 1$.
3. For any event A, $P(A \text{ does not occur}) = 1 - P(A)$
4. Two events A and B are disjoint if they have no outcomes in common and so can never occur simultaneously. If A and B are disjoint, $P(A \text{ or } B) = P(A) + P(B)$. This is the addition rule for disjoint events.

Problem

A survey of magazine subscribers showed that 45.8% rented a car during the past 12 months for business reasons, 54% rented a car during the past 12 months for personal reasons, and 30% rented a car during the past 12 months for both business and personal reasons.
a. What is the probability that a subscriber rented a car during the past 12 months for business or personal reasons?
b. What is the probability that a subscriber did not rent a car during the past 12 months for either business or personal reasons?

Problem (solution)

B = rented a car for business reasons.
0 = rented a car for personal reasons.
a. Probability of B or O = P($B \cup O$) $= 0.458 + 0.54 - 0.30 = 0.698$
b. P(Neither) = P($(B \cup O)^c$) = 1 - P($B \cup O$)$= 1 - 0.698$

Addition Law

For any two events A and B,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

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

Example

Suppose that we have produced the following table of joint probabilities of smoking and lung disease.

	He is a smoker	He is a nonsmoker
He has lung disease	0.12	0.03
He does not have lung disease	0.19	0.66

Determine the probability that a man smokes or has lung disease.

Solution

P(he is a smoker or has lung disease) $=$

P(he is a smoker and has lung disease)

$+$ P(he is a nonsmoker and has lung disease)

$+$ P(he is a smoker and does not have lung disease)

$= 0.12 + 0.03 + 0.19 = 0.34$

Solution (another way)

Applying the Addition Law we perform the following calculation:
P(he is a smoker or has lung disease) $=$
P(he is a smoker) $+$ P(has lung disease)
- P(he is a smoker and has lung disease)
$= 0.31 + 0.15 - 0.12 = 0.34$

Conditional probability

Problem. Josh and Al are avid tennis players and they enjoy playing matches against each other. They do, however, have one difference of opinion on the court. Al likes to have a nice long warm-up session at the start where they hit the ball back and forth and back and forth. Josh's ideal warm-up is to bend at the waist to tie his sneakers and to adjust his shorts. Al thinks that when they rush through the warm-up, he doesn't play as well.

Conditional probability (cont.)

The following table shows the outcomes of their last 20 matches, along with the type of warm-up before they started keeping score.
Does the type of warm-up have an influence on the outcome of a match?

Warm-up time	Al wins	Josh wins	Total
Less than 10 min.	4	9	13
10 min. or more	5	2	7
Total	9	11	20

Solution

Event A = Al wins the match
Event B = the warm-up time is less than 10 minutes
Event B^c = the warm-up time is 10 minutes or more
$P(A \text{ given } B) = \frac{4}{13} \approx 0.31$
$P(A \text{ given } B^c) = \frac{5}{7} \approx 0.71$
It seems that warm-up time has an effect on the outcome of a match.

Another solution

$P(A \cap B) = \frac{4}{20}$ $\qquad$ $P(B) = \frac{13}{20}$

$$\frac{P(A \cap B)}{P(B)} = \frac{4/20}{13/20} = \frac{4}{13}$$

$P(A \cap B^c) = \frac{5}{20}$ $\qquad$ $P(B^c) = \frac{7}{20}$

$$\frac{P(A \cap B^c)}{P(B)} = \frac{5/20}{7/20} = \frac{5}{7}$$

Conditional Probability

When $P(A) > 0$, the **conditional probability** of B given A is

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

Conditional Probability

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

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

Independent Events

Two events A and B are independent if

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

or

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

Otherwise, the events are dependent.

Multiplication Law

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

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

Multiplication Law for Independent Events

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

Problem

Suppose that we have two events, A and B, with $P(A) = 0.50, P(B) = 0.60$, and $P(A \cap B) = 0.40$.
a. Find $P(A|B)$.
b. Find $P(B|A)$.
c. Are A and B independent? Why or why not?

Problem (solution)

a. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.40}{0.60} = \frac{4}{6} \approx 0.66$

b. $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{0.40}{0.50} = \frac{4}{5} = 0.80$

c. No, $P(A \cap B) = 0.4 \neq P(A)P(B) = (0.5)(0.6) = 0.3$

Problem

A Morgan Stanley Consumer Research Survey sampled men and women and asked whether they preferred to drink plain bottled water or a sports drink such as Gatorade or Propel Fitness water. Suppose 200 men and 200 women participated in the study, and 280 reported they preferred plain bottled water. Of the group preferring a sports drink, 80 were men and 40 were women. Let

M=the event the consumer is a man

W=the event the consumer is a woman

B= the event the consumer preferred plain bottled water

S= the event the consumer preferred sports drink.

Problem (cont.)

a. What is the probability a person in the study preferred plain bottled water?

b. What is the probability a person in the study preferred a sports drink?

c. What are the conditional probabilities $P(M|S)$ and $P(W|S)$?

d. What are the joint probabilities $P(M \cap S)$ and $P(W \cap S)$?

e. Given a consumer is a man, what is the probability he will prefer a sports drink?

f. Given a consumer is a woman, what is the probability she will prefer a sports drink?

g. Is preference for a sports drink independent of whether the consumer is a man or a woman? Explain using probability information.

Problem (solution)

	M	W	Total
B	120	160	280
S	80	40	120
Total	200	200	400

Problem

a. $P(B) = \frac{280}{400} = 0.70$
b. $P(S) = \frac{120}{400} = 0.30$
c. $P(M|S) = \frac{P(M \cap S)}{P(S)} = \frac{0.2}{0.3} = 0.66$
d. $P(M \cap S) = 0.2$ $\qquad P(W \cap S) = 0.1$
e. $P(S|M) = \frac{P(S \cap M)}{P(M)} = \frac{0.2}{0.5} = 0.40$
f. $P(S|W) = \frac{P(S \cap W)}{P(W)} = \frac{0.1}{0.5} = 0.20$
g. $P(S|M) \neq P(S)$

Problem

A local bank reviewed its credit card policy with the intention of recalling some of its credit cards. In the past approximately 5% of cardholders defaulted, leaving the bank unable to collect the outstanding balance. Hence, management established a prior probability of 0.05 that any particular cardholder will default. The bank also found that the probability of missing a monthly payment is 20% for customers who do not default. Of course, the probability of missing a monthly payment for those who default is 1.

Problem

a. Given that a customer missed one or more monthly payments, compute the posterior probability that the customer will default.
b. The bank would like to recall its card if the probability that a customer will default is greater than 0.20. Should the bank recall its card if the customer misses a monthly payment? Why or why not?

Problem

D = Default, D^c = customer doesn't default, M = missed payment.

a. $P(D|M) = \frac{P(D \cap M}{P(M)} = \frac{P(D \cap M)}{P(D \cap M) + P(D^c \cap M)}$

$P(D) = 0.05$ $\quad$ $P(D^c) = 0.95$

$P(M|D^c) = 0.20$ $\quad$ $P(M|D) = 1$

$P(D|M) = \frac{(0.05)(1)}{(0.05)(1) + (0.95)(0.20)}$

Problem

Suppose that 80 percent of used car buyers are good credit risks. Suppose, further, that the probability is 0.7 that an individual who is a good credit risk has a credit card, but that this probability is only 0.4 for a bad credit risk. Calculate the probability

a. a randomly selected car buyer has a credit card.

b. a randomly selected car buyer who has a credit card is a good risk.

c. a randomly selected car buyer who does not have a credit card is a good risk.

Solution

a. $A =$ selecting a good credit risk.
$B =$ selecting an individual with a credit card.
$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$
$P(B) = (0.8)(0.7) + (0.2)(0.4) = 0.64$
b. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} = \frac{(0.8)(0.7)}{0.64} = \frac{7}{8}$
c.$P(A|B^c) = \frac{P(A \cap B^c)}{P(B)} = \frac{P(A)P(B^c|A)}{P(B^c)} = \frac{(0.8)(0.3)}{0.36} = \frac{2}{3}$

Bayes' Theorem

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

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

where $A_1 \cup A_2 = S$ and $A_1 \cap A_2 = \varnothing$.

Problem

An actuary studied the likelihood that different types of drivers would be involved in at least one collision during any one-year period. The results of the study are presented below.

Type of driver	Percentage of all drivers	Probability of at least one collision
Teen	8%	0.15
Young adult	16%	0.08
Midlife	45%	0.04
Senior	31%	0.05
Total	100%	

Given that a driver has been involved in at least one collision in the past year, what is the probability that the driver is a young adult driver?

Solution

Y = Young adult and C = collision. We have that $P(C|Y)P(Y) = (0.16)(0.08)$, $P(C|T)P(T) = (0.08)(0.15)$, $P(C|M)P(M) = (0.45)(0.04)$ and $P(C|S)P(S) = (0.31)(0.05)$.

$$P(Y|C) = \frac{(0.16)(0.08)}{(0.16)(0.08) + (0.08)(0.15) + (0.45)(0.04) + (0.31)(0.05)}$$

$$P(Y|C) = 0.21955$$

Example

An insurance company classifies drivers as High Risk, Standard and Preferred. 10% of the drivers in a population are High Risk, 60% are Standard and the rest are preferred. The probability of accident during a period is 0.3 for a High Risk driver, 0.2 for a Standard driver and 0.1 for a preferred driver.

1. Given that a person chosen at random had an accident during this period, find the probability that the person is Standard.
2. Given that a person chosen at random has not had an accident during this period, find the probability that the person is High Risk.

Solution

Let H, S, P and A stand for High-Risk, Standard, Preferred and Accident and W stand for H, S or P.

	$Pr(W)$	$Pr(A\|W)$	$Pr(A \cap W) = Pr(A\|W)Pr(W)$
H	0.10	0.30	0.03
S	0.60	0.20	0.12
P	0.30	0.10	0.03
Total	1		

Solution (1)

1. From Bayes' Formula, taking the figures from the table,

$$Pr(S|A) = \frac{Pr(A|S)Pr(S)}{Pr(A|S)Pr(S) + Pr(A|H)Pr(H) + Pr(A|P)Pr(P)}$$

$$Pr(S|A) = \frac{0.12}{0.18} = \frac{2}{3}$$

Solution (2)

2. For the second part, note that $Pr(A^c|W) = 1 - Pr(A|W)$. This is because if a fraction of p amongst W get into an accident, then the fraction $1 - p$ of W does not get into an accident. You can draw another table or observe that

$$Pr(H|A^c) = \frac{Pr(A^c|H)Pr(H)}{Pr(A^c)} = \frac{(1-0.3)(0.1)}{1-0.18} = \frac{0.07}{0.82} = \frac{7}{82}.$$

Homework?

Suppose it is known that 1% of the population suffers from a particular disease. A blood test has a 97% chance of identifying the disease for diseased individuals, but also has a 6% chance of falsely indicating that a healthy person has the disease.

a) What is the probability that a person will have a positive blood test?

b) If your blood test is positive, what is the chance that you have the disease?

c) If your blood test is negative, what is the chance that you do not have the disease?