Gotta Run!

Unit 4 – Probability

Your Improbable Professor Colton

Unit 4 - Outline

Unit 4 - Probability

Intro

- Basic Probability Concepts
- Calculating Probabilities
- Probability Rules
- Probability types: Subjective, Relative Frequency, Theoretical
- Law of Large Numbers

Intro - Probability

- What does ‘probability’ mean to you in everyday conversation?
- Important when working in contexts where we can’t predict things with certainty, random events
- But we can often look things in the long term, the relative frequency over many occurrences

- From another perspective, probability is all about using population information to answer questions about a sample

Basic Probability Concepts

(Probability) Experiment

- An (**probability**) **experiment** is the process by which a random observation / outcome is generated.
- Example: Flip a coin, roll a die, gender of a child

(Simple) Outcome

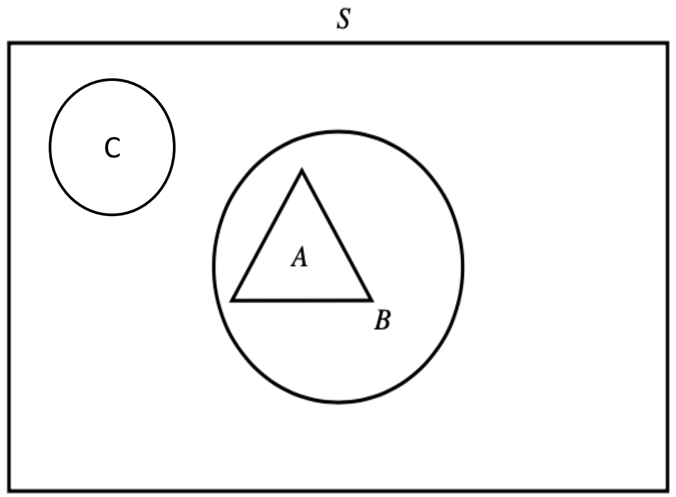
- The **outcomes** are the individual possible things that can happen, like the smallest pieces.
    - Simple means that the outcome cannot be broken up into a collection of other outcomes.
- Example: Flip a coin once, Heads or Tails; Rolling a die, 1 2 3 … 6; Gender, Boy or Girl

Probability

- The **probability** of an outcome is the proportion of times that the outcome occurs in the long run (many, many repetitions)
- Example: Proportion of Heads in 100 tosses, proportion of 3s in 500 rolls

Sample Space

- The set, *S*, of all possible (simple) outcomes of a particular experiment is called the **sample space**.
- Example: Flip a coin twice, S = {HH, HT, TH, TT}
- Can have like a numerical interval **sample space** too.
    - Example: Time until next earthquake: S = [0, infinity)



Event

- An **event** is any collection of possible outcomes of an experiment. In other words, any subset of *S*.
- Example: Flip a coin twice: A = HH, A = HH or HT, A = {TT, HT, TH}, etc.

- Is the sample space S an event?? *YES! S = {HH, HT, TH, TT} is still an event, just happens to be a collection of ALL outcomes!*

LCQ: Sample Spaces

Problem: Find the sample space for the following scenarios.

a) A couple is having twins, what will the genders of the babies be?

b) Flip a coin until the first head is observed.

c) Time to finish an exam with a 2 hr time limit.

LCQ: Sample Spaces

Problem: Find the sample space for the following scenarios.

a) A couple is having twins, what will the genders of the babies be?

Sample Space = {GG, BB, BG}, if order matters S = {GG, BB, BG, GB}

- *These two are different because if order matters, having a Boy first (then a Girl) is different than having a Girl first (then a Boy), even though both have one B and G each → they are different outcomes now*

b) Flip a coin until the first head is observed.

S = {H, TH, TTH, TTTH, TTTTH, ….} → sample spaces can be infinite or uncountable

c) Time to finish an exam with a 2-hour time limit.

S = [0, 2] hours → there is a set min and max time, but you can be any time in between there, so it's a continuous sample space (time is a continues variable)

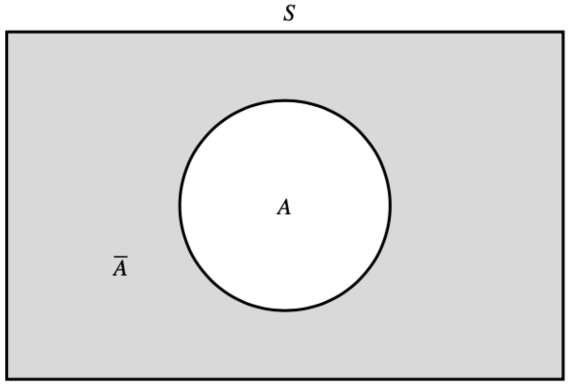
Bonus factoid:

Interval notation: *[0, 2]* *is different than* *(0, 2)*

$0 \leq x \leq 2$ $0 < x < 2$

[] is inclusive *() is exclusive*

More Basic Probability Concepts



Probabilities of Events

- The event itself, A, is just a description in words of the outcome.
- When we talk about the probability of an event A, we are assigning a numerical value to it which represents the likelihood of the event occurring.
- Notation: P(A) = the probability event A is observed

Complements

- Can also talk about the event of Not A occurring, this is called the **Complement = Opposite**.
- Example: Two children, Not having two boys
- Notation: A^C, A' (read as "A prime"), $\bar{A}$, ~~B = Not A~~
- Can find the probability of these as well: P(Not A), $P(A^C)$, P(A'), $P(\bar{A})$

Distribution

- A complete listing of the sample space together with each outcome's probability is called a **distribution** (or a **probability model**).

Outcome:	A	B	C	D	E
Probability:	P(A)	P(B)	P(C)	P(D)	P(E)

Event Notation

- I am just using A here as a general event
- Ex) Let's say we are talking about the cha of rain tomorrow
- I could be more formal and do:
  - A = Rain and P(A)
- **But probably easier to use context and ju say P(Rain)**
  - Or P(Heads) or P(BG) for the previo twins example
- Then the complement would contextually "No Rain" and P(No Rain)

LCQ: Events and Distributions

a) **Find** the distribution for the outcomes of rolling a 6-sided fair die one time:

Outcome:						
Probability:						

b) What is the probability of rolling a 3?

c) What is the probability of rolling an even number?

d) What is the probability of rolling NOT a 5 or 6?

e) If you roll the die twice, what is the probability of getting a sum of 4? *Just think about this one.*

LCQ: Events and Distributions

a) **Find** the distribution for the outcomes of rolling a 6-sided fair die one time:

Outcome:	1	2	3	4	5	6
Probability:	P(1) = 1/6	P(2) = 1/6	1/6	1/6	1/6	1/6

b) What is the probability of rolling a 3?

Event = Rolling a 3, P(3) = 1/6 → just have to look at the probability corresponding to the event of interest

c) What is the probability of rolling an even number?

Event = Even = {2, 4, 6}, P(Even) = 1/6 + 1/6 + 1/6 = 3/6 = ½ → now need to look at ALL the probabilities corresponding to the outcomes of interest

- *Can write final answers as %, fractions or decimals → all are equivalent*

d) What is the probability of rolling NOT a 5 or 6?

4/6 → easy to directly get this answer, or we can think about it using the complement idea as shown below (with more steps for sake of demonstration):

Event = 5 or 6, complement (NOT) 5 or 6 = {5, 6}' = {1, 2, 3, 4} = 4/6

OR can solve this using the Complement Rule discussed later → P(NOT 5 or 6) = 1 – P(5 or 6) = 1 – 2/6 = 4/6 → multiple ways to get this answer, whatever works for you!

LCQ: Events and Distributions

e) If you roll the die twice, what is the probability of getting a sum of 4? *Just think about this one.*

- *I would start by thinking of all the pairs of numbers that equal 4: (1,3) & (2,2). Then consider the probably of rolling those pairs*
- *When thinking about these, we realize that the sample space changed, S = {2, 3, 4, …, 12}. So we would need to find the new distribution, all P(X)s, if we wanted to answer more questions like P(Sum 8), P(Sum < 10), etc..*
- *A good strategy for answering these questions is to first think about the possible ~ options and which sums the go to: (1,1) = 2 (1,2) = 3   (1,3) = 4 …… (6,6) = 12*
- *Then count the options that go to the sum of interest! There is no formula for these questions, and this one is best displayed in the picture shown here:*

Final answer: P(Sum 4) = 3/36 → there is a total of 36 different ~ options (pairs of rolls, order matters) and exactly 3 of them sum to 4

Die	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Calculating Probabilities

Realization

- *This formula / idea is what we were intuitevly doing in the previous LCQ!*
- *Each number was equally likely and we just counted the number of outcomes in our event of interest (like even = 3 outcomes) and divided that by the total number of equally likely outcomes, 6*

Calculating Probabilities with Equally Likely Outcomes

- If a sample space S has n equally likely outcomes, and an event A has k outcomes, then

$$P(A) = \frac{number\ of\ outcomes\ in\ A}{number\ of\ outcomes\ in\ S} = \frac{k}{n}$$

Sampling from a Population

- Sampling an individual from a population is a probability experiment.
- The population is the sample space and members of the population are equally likely outcomes.
- Examples: Academic Office has information about student's majors.
  - What is the probability a randomly selected student is a Statistics major?
    - *435 equally likely students to be selected, 150 are Stats majors → P(Stats) = 150/435*
  - What is the probability a randomly selected student is an Art or Chemistry major?
    - *Event A = Art or Chemistry, which has 105 + 180 = 285 students → P(Art or Chem) = 285/435*

Total Students: 435	

Major	**Students**
Statistics	150
Art	105
Chemistry	180

Unusual Events

- An unusual event is one whose probability is very small.
- A rule of thumb is that any event whose probability is less than 0.05 is considered to be unusual.
- Example: For example, in a college of 5000 students, 150 are math majors.
  - A student is selected at random and turns out to be a math major. Is this an unusual event?
  - *P(Math) = 150 / 5000 = 0.03 = 3% → unusual because probability is < 5%, which means it would be weird/odd/unexpected to select a math major*

LCQ: Distributions 2

Find the distribution for the number of heads when flipping two coins:

X:			
P(X):			

LCQ: Distributions 2

Find *the <u>distribution</u> for the <u>number of heads</u> when flipping two coins:*

First think about the Sample space:

- *Like we've seen before with flipping coins, so S = {TT, TH, HH} → WRONG!! this wouldn't be the sample space because of the context!*
  - *Have to look at the question we are trying to answer! We want the 'number of heads when flipping two coins!!*

X (outcome):	*0*	*1*	*2*
P(X):	~~*1/3*~~	~~*1/3*~~	~~*1/3*~~

Next is the probabilities:

- *Three outcomes in S, so all must be 1/3 → WRONG!! Think about the possible ~ options!*
- *Possible options ~ {TT = 0, TH = 1, HT = 1, HH = 2}, each has ¼ probability*
- *So then "map" each option to an outcome and add up the probabilities!*
- *Now we get:*

X (outcome):	*0*	*1*	*2*
P(X):	*1/4*	*2/4*	*1/4*

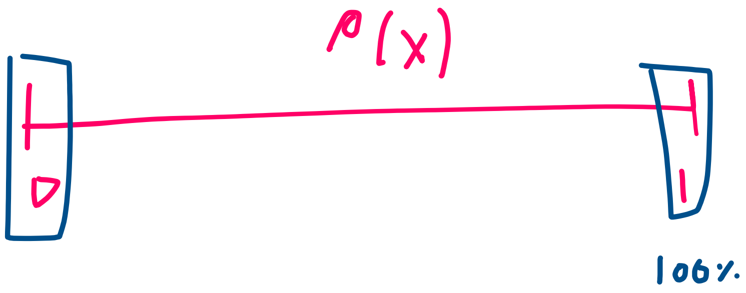
CORRECT now!

<u>Realization</u>

This is an example when all outcomes (the ones in the final table) are NOT equally likely!

When we break it down into the possible ~ options, those are equally likely though, which helped us find the correct probabilities

Probability Rules



All **probabilities** have the following **rules** or **characteristics**:

1. All **probabilities** are between zero and one: $0 \leq P(x) \leq 1$

    - Using interval notation P(x) = [0,1] (inclusive)
    - Probability of 0 means it cannot occur; Probability of 1 means it always occurs.
    - So, this means the probability of any event A must be $0 \leq P(\text{A}) \leq 1$.

2. The sum of the **probabilities** of all outcomes in a sample space is one: $\sum P(x) = 1$

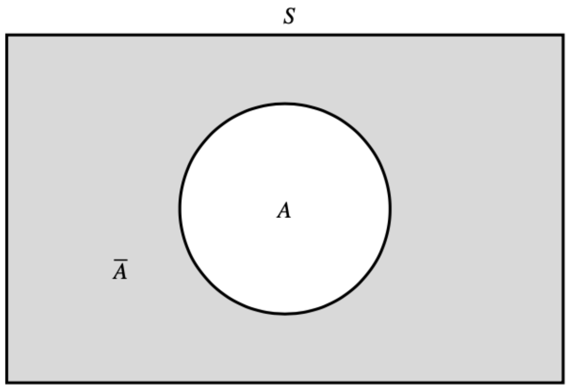
Outcome, X:	1	2	3	4	5	6	*Total*
Probability, P(X):	1/6	1/6	1/6	1/6	1/6	1/6	1

- If we have a valid probability distribution, the sum of all simple outcomes probabilities' equals 1.
    - AND all of the individual probabilities meet the first rule!

- This does NOT mean the sum of the probabilities of all EVENTS equals one…
    - If A = even number, B = odd number and C = 3 → $\text{P(A)} + \text{P(B)} + \text{P(C)} > 1$

      P(Even) = ½ = P(Odd), P(3) = 1/6

Complement Probabilities

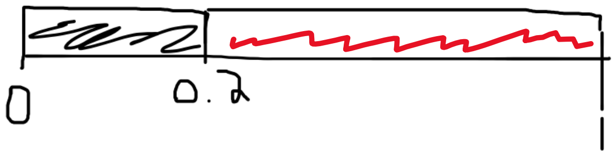


- Here is another probability rule → **The Complement Rule**!
- If we have Event A, then the probability of the complement of A, A’ is:

  $P(A') = 1 - P(A)$

  - Remember we subtract from one because the entire probability distribution is equal to 1.
- Simple example: What is the probability of A’?

  P(A) = 0.2 *P(A’) = 1 – P(A) = 0.8*

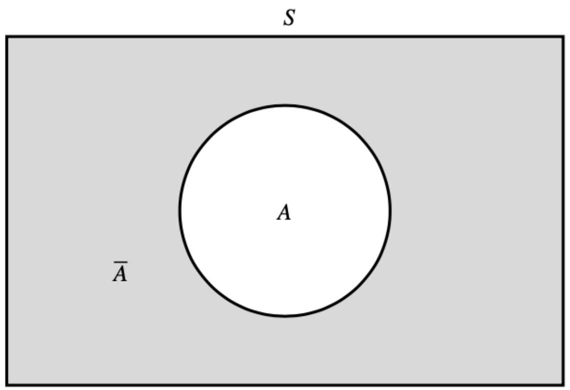


- Can do this with more complicated events.
  - Very useful for problems with “at least” or “no more than”, etc.
  - Example:
    - Find $P(X > 3)$
    - Find P(X no more than 10)

$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$

$P(X > 3) = P(4) + P(5) + \ldots + P(13)$
$= 1 - P(X \leq 2)$

Complement Probabilities

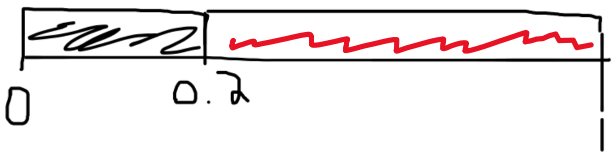


- Here is another probability rule! → **The Complement Rule**!
- If we have Event A, then the probability of the complement of A, A' is:

  $$P(A') = 1 - P(A)$$

  - Remember we subtract from one because the entire probability distribution is equal to 1.
- Simple example: What is the probability of A'?

  P(A) = 0.2 *P(A') = 1 – P(A) = 0.8*



- Can do this with more complicated events.
  - Very useful for problems with "at least" or "no more than", etc.
  - Example:
    - Find P(X > 3)
    - Or Find P(X no more than 10)

$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$

$P(X > 3) = P(4) + P(5) + \dots + P(13)$

$= 1 - P(X \leq 2)$

Probability Types

There is two different (valid) ways we can think about probabilities.

Personal (Subjective) Probabilities

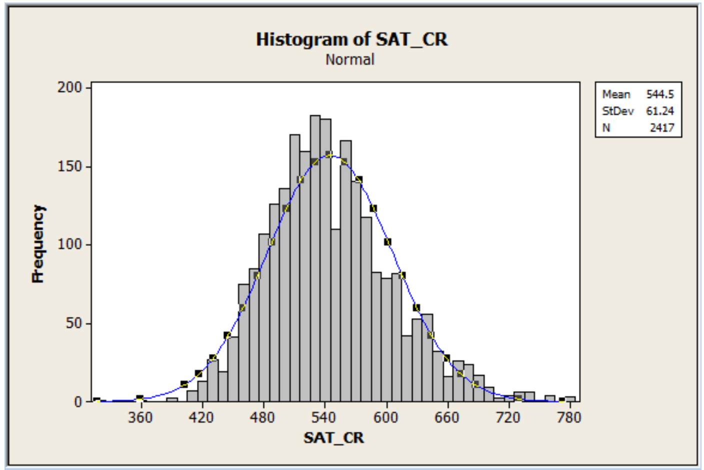
- Based on personal experience; just educated guesses based on what "you" have experienced.
- Examples:
  - Guessing the forecast as you get ready in the morning.
  - Gas tank is on Empty and thinking about the probability you can make it back home.
- These aren't valid ways we are going to discuss probability in a stats context.

Theoretical Probability

- (Probability distribution) This is based on the underlying theoretical "model" of the experiment.
- Kinda like the scientific probabilities.
- Example: Theoretical probability of flipping a fair coin is obviously P(H) = 0.5 and P(T) = 0.5
  - Probability of selecting a Diamond from a standard 52 card deck = 1/4.

Empirical (Relative Frequency) Probability

- Requires collecting data to count outcomes that actually occur.
- Can think of this as the experimental (observed) probabilities.
- Example: If I flip a coin 10 times and get 7 Heads and 3 Tails
  - Empirical Probabilities: P(H) = 0.7 and P(T) = 0.3
- Small samples may not reflect accurate probabilities.



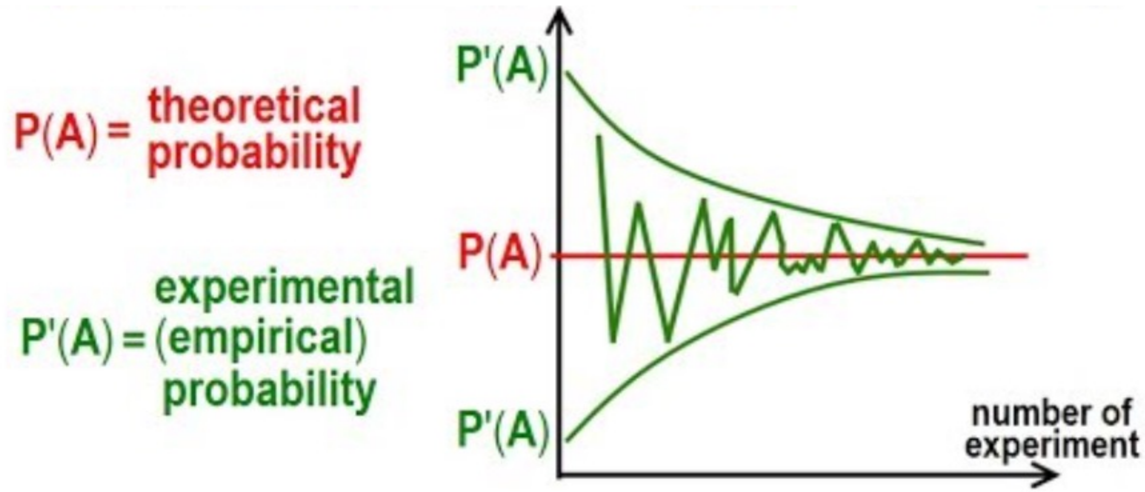
https://www.softschools.com/math/probability_and_statistics/the_normal_distribution_empirical_rule/

Law of Large Numbers

Law of Large Numbers

- This is an important idea that states: the more times an experiment is repeated (increased sample size), the closer the **empirical** (relative frequency) probability gets to the **theoretical** probability.
- In short, big samples are more reliable than small samples when determining relative frequency probabilities.

- Cool applet demonstration:
  - https://www.stapplet.com/largenum.html



Law of Averages, Misconception!

<u>"Law of Averages"</u>

- This is the idea that you can be "due" for something, e.g. due to make the next shot, due to roll a 6, etc.
- This notion is a <u>misconception</u>, a fallacy! It makes no sense in Statistics!
- With <u>independent</u> scenarios (like rolling die, or spinner landing on red), the <u>results of previous runs</u> have <u>NO impact</u> on what happens <u>next</u>. It does NOT change the future probabilities!
    - Flipping 10 heads in a row doesn't make the next toss more likely to be tails, it's still 50/50!!

- This fake Law of Averages is often confused with the Law of Large Numbers!
    - The difference is maybe subtle, but important!
    - The "<u>Law of Averages</u>" is attempting to say things about the <u>short term probabilities</u>!
    - Whereas the the of <u>Law of Large Numbers</u> refers to the <u>long-term probabilities</u>, over many many trials the overall relative probabilities will get close to the true theoretical probabilities!

<u>Poker Example</u>

- Let's say you win 35% of the poker hands that you play
- If you lose 10 hands in a row, the law of averages is trying to tell you that you are more likely to win the next one because of the previous losses…. But this is NOT true!
- Now let's say you played 100 hands, the law of large numbers tells us that you would have one close to 35% of the hands overall! This should be roughly TRUE! Closer to 35% as you play more and more hands

PROBLEM SESSION!!!!!!!!!!!

Problem #23

A toy company is preparing to market an electronic game for young children that "randomly" generates a color. They suspect, however, that the way the random color is determined may not be reliable, so they ask the programmers to perform tests and report the frequencies of each outcome. Are each of the following probability assignments possible? Why or why not?

	Probabilities of ...			
	Red	Yellow	Green	Blue
a)	0.25	0.25	0.25	0.25
b)	0.10	0.20	0.30	0.40
c)	0.20	0.30	0.40	0.50
d)	0	0	1.00	0
e)	0.10	0.20	1.20	−1.50

Problem #23 Solution

a) Yes
b) Yes
c) No, probabilities sum to more than 1
d) Yes
e) No, probabilities must be between 0 and 1, inclusive.

Problem #17

Respond to the following questions:

a) A casino claims that its roulette wheel is truly random. What should that claim mean?

b) A reporter on *Market Place* says that there is a 50% chance that the NASDAQ will hit a new high in the next month? What is the meaning of such a phrase?

Problem #17 Solution

a) Every slot is equally likely
b) This is likely a personal probability

Problem #19

Even though commercial airlines have excellent safety records, in the weeks following a crash, airlines often report a drop in the number of passengers, probably because people are afraid to risk flying.

a) A travel agent suggests that since the law of averages makes it highly unlikely to have two plane crashes within a few weeks of each other, flying soon after a crash is the safest time. What do you think?

b) If the airline industry proudly announces that it has set a new record for the longest period of safe flights, would you be reluctant to fly? Are the airlines due to have a crash?

Problem #19 Solution

a) This is silly; whether or not an airplane crashes does not affect the probability of other planes crashing. We can assume the events are independent.

b) There is no such thing as the law of averages. The long period of safe flights does not change the probability of a crash.

Problem #21

Insurance companies collect annual payments from homeowners in exchange for paying to rebuild houses that burn down.

a) Why should you be reluctant to accept a $300 payment from your neighbor to replace his house should it burn down during the coming year?
b) Why can the insurance company make that offer?

Problem #21 Solution

a) Because, while it is unlikely that it would happen, if it did happen you would not have enough money to rebuild the house – it would cost you a lot more than $300.

b) Because the insurance company collects $300 from hundreds of thousands of customers, while only a few of their customers will need to file a claim.

Problem #27

In developing their warranty policy, an automobile company estimates that over a 1-year period, 17% of their new cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs. If you buy a new car from them, what is the probability that your car will need:

a) No repairs?
b) No more than one repair?
c) Some repairs?

Problem #27 Solution

a) P(No repairs) = 1 – 0.17 – 0.07 – 0.04 = 0.72
b) P(No more than 1 repair) = P(0) + P(1) = 0.72 + 0.17 = 0.89
c) P(Some repairs) = P(1 or more) = 1 – P(0) = 1 – 0.72 = 0.28

Problem #35

The Mars company says that before the introduction of purple, yellow made up 20% of their plain M&M candies, red made up another 20%, and orange, blue, and green each made up 10%. The rest were brown.

a) If you picked an M&M at random from a pre-purple bag of candies, what is the probability that it was:
   i. Brown?
   ii. Yellow or orange?
   iii. Not green?
   iv. Striped?

Problem #35 Solution

a) P(Brown) = 0.3

P(Yellow or Orange) = 0.3

P(not green) = 0.90

P(striped) = 0

Problem #3

In many state lotteries, you can choose which numbers to play. Consider a common form in which you choose 5 numbers. Which of the following strategies can improve your chance of winning? If the method works, explain why. If not, explain why using appropriate statistics terms.

a) Always play 1, 2, 3, 4, 5
b) Choose the numbers that did come up in the most recent lottery drawing because they are "hot."

Problem #3 Solution

a) Since each number is equally likely to be selected for the lottery this won’t help, but it won’t hurt, either.
b) Won’t work, each number is equally likely, just because the numbers are “hot” doesn’t change the probability of their being selected.