MODELING COMPETITION TEAM – SUMMER CHECKLIST

REMIND INSTRUCTIONS

☐ The first document contains instructions for joining the Remind class discussion group. At any time after 7/15/20 and before the first day of school, please follow the instructions and join the group.

SYLLABUS

☐ Document #2 is the course syllabus. Please read it carefully.

☐ Document #3 contains information on the AICE Computer Science syllabus. Please read it carefully.

☐ You must have a three-ring binder and loose-leaf paper the first day of class.

☐ You must have a flash drive the first day of class.

PRACTICE

☐ Document #4 contains a great deal of probability and statistics material that you must understand to be successful in this class. Please study this material carefully. You do not need to complete any of the exercises.

remind

Sign up for important updates from Mr. Dominguez.

Get information for **Everglades High School** right on your phone—not on handouts.

Pick a way to receive messages for **EHS Modeling Team**:

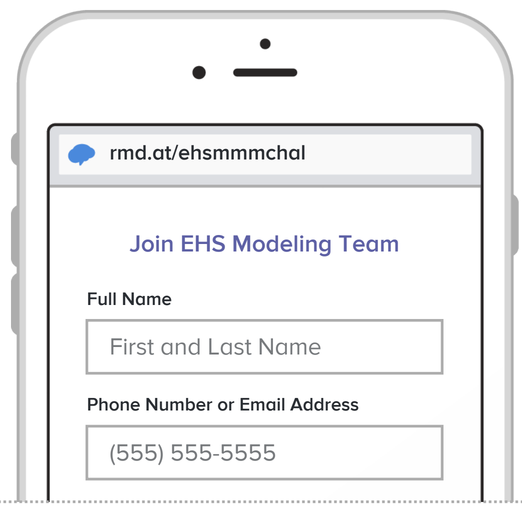
A

If you have a smartphone, get push notifications.

On your iPhone or Android phone, open your web browser and go to the following link:

rmd.at/ehsmmmchal

Follow the instructions to sign up for Remind. You'll be prompted to download the mobile app.



B

If you don't have a smartphone, get text notifications.

Text the message @ehsmmmchal to the number 81010.

If you're having trouble with 81010, try texting @ehsmmmchal to (954) 883-9389.

** Standard text message rates apply.*

Don't have a mobile phone? Go to rmd.at/ehsmmmchal on a desktop computer to sign up for email notifications.

Modeling and Simulation – UNDER CONSTRUCTION

Modeling and Simulation is a high school course designed to meet the requirements of MAP 3127 (Mathematical Modeling) or MAP 3482 (Computer Simulation) in Florida's Statewide Course Numbering System.

The course provides instruction in computer-assisted mathematical analysis and the development of tailored algorithms for solving functional area problems. Students will apply mathematical principles to the development and solving of mathematical models of problems in the social, life, physical and management sciences. Computing projects using MATLAB will be used to deepen the understanding of the concepts and to illustrate the techniques encountered in the course.

Topics include function fitting, regression analysis, optimization, linear and nonlinear programming, queueing models, and discrete and continuous dynamical systems. Specific models that will be explored include population growth, epidemic spread, predator-prey populations, economic systems, and financial markets.

Pre-requisites

1) Students must be available to participate in the MathWorks Mathematics Modeling Challenge (M3C).
2) Open only to juniors and seniors
   - Only juniors and seniors can participate in M3C
3) AP Calculus AB with a grade of B or better and AP Computer Science A with a grade of B or better
   - The ideally prepared student would have also completed statistics, but this course will introduce all the necessary probability and statistics.

Assigned Readings

- A. Dominguez, *Simulation Study Notes*
- S. Ross, *Simulation 4th Edition*
- R. Rubinstein, *Simulation and the Monte Carlo Method*

Additional details

1. M3C is Friday, March 5, 2021, from 8AM to 10PM (tentative). Students must be available during this entire 14-hour window.
2. Only 3-10 students (1-2 teams x 3-5 students) can participate in M3C, but additional students can take this class if interested.

Topic – UNDER CONSTRUCTION	Approx. Timeline	Primary Sources
1) Introduction to Modeling and Simulation	1 day	Ross, Chapter 1 & Section 6.1
2) Single-Server Queueing System	2 days	Ross, Section 6.2 Rubinstein, Section 6.4.1
3) Two-Server Queueing System	2 days	Ross, Sections 6.3 & 6.4
4) Inventory Model	2 days	Ross, Section 6.5
5) Insurance Risk Model	2 days	Ross, Section 6.6
6) Repairman Model	2 days	Ross, Section 6.7 Rubinstein, Section 6.4.2
7) Stock Option Model	2 days	Ross, Sections 6.8 & 8.8
8) Statistical Analysis of Data	3 days	Ross, Chapter 7
9) Curve Fitting of Data and Goodness of Fit Test	2 days	Ross, Sections 9.1 & 9.2

Supplementary Textbook References

- M. Heath, *Scientific Computing 2nd Edition*
  - Chapter 13 discusses simulation models
- R. Pindyck and D. Rubinfeld, *Econometric Models and Economic Forecasts 4th Edition*
  - Chapters 13 and 14 discuss simulation models
- S. Ross, *Introduction to Probability Models 10th Edition*
  - Chapter 11 discusses simulation models
- B. Abraham and J. Ledolter, *Statistical Methods for Forecasting*
- W. Press et al, *Numerical Recipes 3rd Edition*

Web Resources

- M3C website https://m3challenge.siam.org/challenge
- Modeling the Future Challenge website https://www.mtfchallenge.org/
- MathWorks website https://www.mathworks.com/support/learn-with-matlab-tutorials.html
- MIT OCW website https://ocw.mit.edu/courses/mathematics/18-s997-introduction-to-matlab-programming-fall-2011/

2 Syllabus overview

Aims

The aims of this course are to enable students to develop:

- computational thinking skills
- an understanding of the main principles of solving problems using computers
- an understanding of the component parts of computer systems and how they interrelate, including software, data, hardware, communication and people
- an understanding of the different methods of communication and the functionality of networks and the internet
- the skills necessary to apply this understanding to develop computer based solutions to problems

Content overview

AS Level Content

1 Information representation	1.1 Data Representation 1.2 Multimedia – *Graphics*, *Sound* 1.3 Compression
2 Communication	2.1 Networks including the internet
3 Hardware	3.1 Computers and their components 3.2 Logic Gates and Logic Circuits
4 Processor Fundamentals	4.1 Central Processing Unit (CPU) Architecture 4.2 Assembly Language 4.3 Bit manipulation
5 System Software	5.1 Operating System 5.2 Language Translators
6 Security, privacy and data integrity	6.1 Data Security 6.2 Data Integrity
7 Ethics and Ownership	7.1 Ethics and Ownership
8 Databases	8.1 Database Concepts 8.2 Database Management System (DBMS) 8.3 Data Definition Language (DDL) and Data Manipulation Language (DML)
9 Algorithm Design and Problem-Solving	9.1 Computational Thinking Skills 9.2 Algorithms

10 Data Types and structures	10.1 Data Types and Records 10.2 Arrays 10.3 Files 10.4 Introduction to Abstract Data Types (ADT)
11 Programming	11.1 Programming Basics 11.2 Constructs 11.3 Structured Programming
12 Software Development	12.1 Program Development Lifecycle 12.2 Program Design 12.3 Program Testing and maintenance

A Level Content

13 Data Representation	13.1 User-defined data types 13.2 File organisation and access 13.3 Floating-point numbers, representation and manipulation
14 Communication and internet technologies	14.1 Protocols 14.2 Circuit switching, packet switching
15 Hardware and Virtual Machines	15.1 Processers, Parallel Processing and Virtual Machines 15.2 Boolean Algebra and Logic Circuits
16 System Software	16.1 Purposes of an Operating System (OS) 16.2 Translation Software
17 Security	17.1 Encryption, Encryption Protocols and Digital certificates
18 Artificial Intelligence (AI)	18.1 Artificial Intelligence
19 Computational thinking and problem solving	19.1 Algorithms 19.2 Recursion
20 Further Programming	20.1 Programming Paradigms 20.2 File Processing and Exception Handling

Support for Cambridge International AS & A Level Computer Science

Our School Support Hub **www.cambridgeinternational.org/support** provides Cambridge schools with a secure site for downloading specimen and past question papers, mark schemes, grade thresholds and other curriculum resources specific to this syllabus. The School Support Hub community offers teachers the opportunity to connect with each other and to ask questions related to the syllabus.

Assessment overview

At AS Level candidates take papers 1 and 2. At A Level candidates take all four papers. Calculators must not be used in any paper.

Paper 1 Theory Fundamentals

1 hour 30 minutes

75 marks

Paper 1 will assess sections 1 to 8 of the syllabus content.

Written paper.

Externally assessed. Candidates answer all questions.

50% of the AS Level

25% of the A Level

Paper 2 Fundamental Problem-solving and Programming Skills

2 hours

75 marks

Paper 2 will assess sections 9 to 12 of the syllabus content.

Candidates will need to write answers in pseudocode.

Written paper.

Externally assessed. Candidates answer all questions.

50% of the AS Level

25% of the A Level

Paper 3 Advanced Theory

1 hour 30 minutes

75 marks

Paper 3 will assess sections 13 to 20 of the syllabus content.

Written paper.

Externally assessed. Candidates answer all questions.

25% of the A Level

Paper 4 Practical

2 hours 30 minutes

75 marks

Paper 4 will assess sections 19 to 20 of the syllabus content.

Candidates will submit complete program code and evidence of testing.

Candidates will be required to use either Java, VB.NET or Python programming languages.

Externally assessed. Candidates answer all questions on a computer without internet or email facility.

25% of the A Level

There are three routes for Cambridge International AS & A Level Computer Science:

	Route	Paper 1	Paper 2	Paper 3	Paper 4
1	**AS Level only** (Candidates take all AS components in the same exam series)	✓	✓		
2	**A Level** (staged over two years) Year 1 AS Level*	✓	✓		
	Year 2 Complete the A Level			✓	✓
3	**A Level** (Candidates take all components in the same exam series)	✓	✓	✓	✓

* Candidates carry forward their AS Level result subject to the rules and time limits described in the *Cambridge Handbook*.

Candidates following an AS Level route will be eligible for grades a–e. Candidates following an A Level route are eligible for grades A*–E.

Assessment objectives

The assessment objectives (AOs) are:

AO1: Demonstrate knowledge and understanding of the principles and concepts of computer science, including abstraction, logic, algorithms and data representation.

AO2: Apply knowledge and understanding of the principles and concepts of computer science, including to analyse problems in computational terms.

AO3: Design, program and evaluate computer systems to solve problems, making reasoned judgements about these.

Weighting for assessment objectives

The approximate weightings allocated to each of the assessment objectives (AOs) are summarised below.

Assessment objectives as a percentage of each qualification

Assessment objective	Weighting in AS Level %	Weighting in A Level %
AO1	30	30
AO2	40	30
AO3	30	40
Total	100	100

Assessment objectives as a percentage of each component

Assessment objective	Weighting in components %			
	Paper 1	**Paper 2**	**Paper 3**	**Paper 4**
AO1	60	–	60	–
AO2	40	40	40	–
AO3	–	60	–	100
Total	100	100	100	100

LESSON 2 – INTRODUCTION TO PROBABILITY

SAMPLE SPACES AND EVENTS

Consider an experiment whose outcome is not known in advance, due to some random element. The set of all possible outcomes for the experiment is known as the **sample space** of the experiment and is usually designated S. Any subset A of the sample space represents a set of possible outcomes of the experiment and is called an **event**. We say that A happens if one of the outcomes in the event A turns out to be the actual result of the experiment.

For any two events A and B we define a new event A U B, called the **union** of A and B, which consists of all outcomes that are either in A or in B or in both. AUB happens if A happens and/or B happens.

Similarly, we define a new event AB (or $A \cap B$), called the **intersection** of A and B, which consists of all outcomes that are in both A and B. $A \cap B$ happens if and only if A and B both happen

For any event A, we define a new event A^c, called the **complement** of A, which consists of all outcomes in the sample space S that are not in A. A^c happens if and only if A does not happen.

Since the outcome of the experiment must lie in the sample space S, it follows that Sc does not contain any possible outcomes and thus cannot happen. We call S^c the null (or empty) set and designate it ø. If A and B cannot both happen, then $A \cap B = ø$, and we say that A and B are mutually exclusive.

AXIOMS OF PROBABILITY

1. For any event A, $0 \le P(A)$
2. $P(S) = 1$
3. For any finite or countably infinite sequence of mutually exclusive events $A_1, A_2, \ldots$, $P(A_1 U A_2 U \ldots) = P(A_1) + P(A_2) + \ldots$

PROBABILITY THEOREMS

Theorem 1. For any event A, $P(A^c) = 1 - P(A)$

Proof: $A \cup A^c = S$. $P(A \cup A^c) = P(A) + P(A^c)$ by axiom 3 because A and A^c are mutually exclusive. $P(A) + P(A^c) = P(S) = 1$ by axiom 2. Therefore $P(A^c) = 1 - P(A)$

Theorem 2. $P(\emptyset)=0$

Proof: Since $S^c = \emptyset$, from Theorem 1, we can see that $P(\emptyset) = 1 - P(S) = 1 - 1 = 0$.

IMPORTANT FACT. $P(A) = 0$ does **not** imply that $A = \emptyset$. Likewise, $P(A) = 1$ does **not** imply that $A = S$.

Theorem 3. If A and B are events in a sample space S and A is a subset of B, then $P(A) \leq P(B)$.

Proof. $B = A \cup (A^c \cap B)$ by Exercise 1. Also, A and $(A^c \cap B)$ are mutually exclusive.[1] Therefore, $P(B) = P(A) + P(A^c \cap B) \geq P(A)$.

Theorem 4. For any event A, $P(A) \leq 1$.

Proof. Since any event A is a subset of the sample space S, by Theorem 3, $P(A) \leq P(S) = 1$.

Theorem 5. If A and B are any two events in a sample space S, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

This theorem confirms our intuition that the probability that either A or B happens is equal to the probability of A happening plus the probability of B happening minus the probability of both A and B happening since this would have been double-counted as part of both A happening and B happening.

Theorem 6. If A, B and C are any three events in a sample space S, then $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

Theorem 7. $P(A) \geq P(A \cap B)$ and $P(A) \leq P(A \cup B)$

Theorem 8. $P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B)$

Theorem 9. $P(A) + P(B) - 1 \leq P(A \cap B) \leq P(A) + P(B)$

EXAMPLES

Example 1. Consider the fair flipping of a coin. The sample space is {Heads, Tails}. Each of the events {Heads} and {Tails} has equal probability of happening. $P(H) = P(T) = 1/2$.

---

[1] Make sure you understand why this claim is true.

Example 2. Consider the fair tossing of a six-sided die. The sample space is {1, 2, 3, 4, 5, 6}. Each of the six events {1}, {2}, {3}, {4}, {5} and {6} has equal probability of happening. P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6.

Example 3. Consider the act of picking a card from a standard 52-card deck. P(red card) = P(black card) = 1/2. P(a given suit) = 1/4. P(a given rank) = 1/13.

EXERCISES

1) If A is a subset of B, then B = A U ($A^c \cap B$). Verify this by means of a Venn diagram.

2) Prove Theorems 5, 6, 7, 8 and 9.

3) The event "A or B not both" will occur can be written as ($A \cap B^c$) U ($A^c \cap B$). Express the probability of this event in terms of P(A), P(B) and P(A∩B).

REFERENCES

Kreysig, Advanced Engineering Mathematics 8th Edition, pp 1049 – 1155

Mendenhall et al, *Mathematical Statistics with Applications 3rd Edition*, Chapters 1-2

I. Miller & M. Miller, *John Freund's Mathematical Statistics 6th Edition*, Chapters 2-4

S. Ross, *Introduction to Probability Models 10th Edition*, Chapter 1

S. Ross, *Simulation 4th Edition*, Chapter 2

Adapted from source material by Alberto Dominguez

Lesson 2.1: The General Multiplication Rule

Outcomes

- Use the general multiplication rule to calculate the probability of the intersection of two events.
- Interpret probabilities in context.

Classwork

Example 1: Independent Events

Example 1: Independent Events

Do you remember when breakfast cereal companies placed prizes in boxes of cereal? Possibly you recall that when a certain prize or toy was particularly special to children, it increased their interest in trying to get that toy. How many boxes of cereal would a customer have to buy to get that toy? Companies used this strategy to sell their cereal.

One of these companies put one of the following toys in its cereal boxes: a block (B), a toy watch (W), a toy ring (R), and a toy airplane (A). A machine that placed the toy in the box was programmed to select a toy by drawing a random number of 1 to 4. If a 1 was selected, the block (or B) was placed in the box; if a 2 was selected, a watch (or W) was placed in the box; if a 3 was selected, a ring (or R) was placed in the box; and if a 4 was selected, an airplane (or A) was placed in the box. When this promotion was launched, young children were especially interested in getting the toy airplane.

Adapted from source material by Alberto Dominguez

Exercises 1–8

1. **If you bought one box of cereal, what is your estimate of the probability of getting the toy airplane?**

   The probability is $\frac{1}{4}$. I got this answer based on a sample space of $\{B, W, R, A\}$. Each outcome has the same chance of occurring. Therefore, the probability of getting the airplane is $\frac{1}{4}$.

2. **If you bought a second box of cereal, what is your estimate of the probability of getting the toy airplane in the second box? .**

   The probability is again $\frac{1}{4}$. Since the machine that places a toy in the box picks a random number from 1 to 4, the probability that the second toy will be an airplane will again be $\frac{1}{4}$.

3. **If you bought two boxes of cereal, does your chance of getting at least one airplane increase or decrease?**

   The probability of getting at least one airplane increases because the possible outcomes include the following: an airplane in the first box but not the second box, an airplane in the second box but not the first box, and an airplane in both the first and second boxes.

4. **Do you think the probability of getting at least one airplane from two boxes is greater than 0.5?**

   I think the probability is less than 0.5 as there are many more outcomes that I can describe that do not include the airplane. For example, you could get a watch and a block, a watch and a watch, a watch and a ring, and so on.

A tree is a way to organize the outcomes in a systematic list. A tree diagram provides a strategy to organize and analyze outcomes in other problems presented in this module.

Adapted from source material by Alberto Dominguez

5. List all of the possibilities of getting two toys from two boxes of cereal. (Hint: Think of the possible outcomes as ordered pairs. For example, *BA* would represent a block from the first box and an airplane from the second box.)

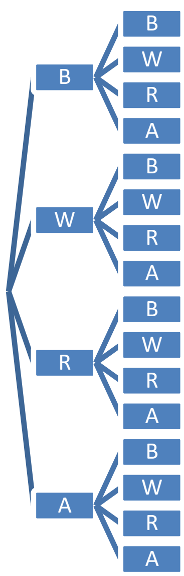
Consider the following ways you might create their lists using the notation:

B for block, W for watch, R for ring, and A for airplane.

The first letter represents the toy found in the first box, and the second letter represents the toy found in the second box. The first column represents getting the block in the first box, followed by each one of the other toys. The second column represents getting the watch in the first box, followed by each one of the other toys. The third column is developed with the ring in the first box, and the fourth column is developed with the airplane in the first box.

BB	***WB***	***RB***	***AB***
BW	***WW***	***RW***	***AW***
BR	***WR***	***RR***	***AR***
BA	***WA***	***RA***	***AA***

The following represents a tree diagram to form the lists:



Adapted from source material by Alberto Dominguez

6. **Based on the list you created, what do you think is the probability of each of the following outcomes if two cereal boxes are purchased?**

   a. **One (and only one) airplane**

      $\frac{6}{16}$ *or* $\frac{3}{8}$

   b. **At least one airplane**

      $\frac{7}{16}$

   c. **No airplanes**

      $\frac{9}{16}$

7. **Consider the purchase of two cereal boxes.**

   a. **What is the probability of getting an airplane in the first cereal box?**

      The probability is $\frac{1}{4}$ *because there are* 4 *possible toys, and each is equally likely to be in the box.*

   b. **What is the probability of getting an airplane in the second cereal box?**

      Again, the probability is $\frac{1}{4}$ *because there are* 4 *possible toys, and each is equally likely to be in the box.*

   c. **What is the probability of getting airplanes in both cereal boxes?**

      The probability is $\frac{1}{16}$ *because there is* 1 *pair in the list (AA) out of the* 16 *that fits this description.*

The event of getting an airplane in the first box purchased and then getting an airplane in the second box purchased is an example of what are called *independent events*. Two events are independent of each other if knowing that one event has occurred does not change the probability that the second event occurs. Point out that because of the way toys are placed in the boxes, knowing the type of toy placed in the first cereal box does not tell us anything about what toy will be found in the second cereal box.

The probability of A and B is denoted as $P(A \text{ and } B)$. This is also called the probability of A intersect B or $P(A \text{ "intersect" } B)$. If A and B are independent events, then the multiplication rule for independent events applies:

Adapted from source material by Alberto Dominguez

$P(A \text{ and } B)$ is the probability that events A and B *both* occur and is the probability of the intersection of A and B. The probability of the intersection of events A and B is sometimes also denoted by $P(A \cap B)$.

Multiplication Rule for Independent Events

If A and B are independent events, $P(A \text{ and } B) = P(A) \cdot P(B)$.

This rule generalizes to more than two independent events. For example:

$P(A \text{ and } B \text{ and } C)$ or $P(A \text{ intersect } B \text{ intersect } C) = P(A) \cdot P(B) \cdot P(C)$.

8. **Based on the multiplication rule for independent events, what is the probability of getting an airplane in both boxes?**

   The probability would be $\frac{1}{16}$ as I would multiply the probability of getting an airplane in the first box $\left(\frac{1}{4}\right)$ by the probability of getting an airplane in the second box, which is also $\frac{1}{4}$.

Example 2: Dependent Events

Example 2 moves to a discussion of dependent events.

The main idea of this example is that as a piece of chocolate is chosen from a box, the piece is not replaced. Since the piece is not replaced, there is one less piece, so the number of choices for the second piece changes. The probability of getting a particular type of chocolate on the second selection is dependent on what type was chosen first.

Two events are dependent if knowing that one event occurring changes the probability that the other event occurs. This second probability is called a *conditional probability*.

Example 2: Dependent Events

Do you remember the famous line, "Life is like a box of chocolates," from the movie *Forrest Gump*? When you take a piece of chocolate from a box, you never quite know what the chocolate will be filled with. Suppose a box of chocolates contains 15 identical-looking pieces. The 15 are filled in this manner: 3 caramel, 2 cherry cream, 2 coconut, 4 chocolate whip, and 4 fudge.

Adapted from source material by Alberto Dominguez

Exercises 9–14

Exercises 9–14

9. If you randomly select one of the pieces of chocolate from the box, what is the probability that the piece will be filled with fudge?

   $\frac{4}{15} \approx 0.2667$

10. If you randomly select a second piece of chocolate (after you have eaten the first one, which was filled with fudge), what is the probability that the piece will be filled with caramel?

   $\frac{3}{14} \approx 0.2143$

The events, ***picking a fudge-filled piece on the first selection*** and ***picking a caramel-filled piece on the second selection***, are called ***dependent events.***

Two events are dependent if knowing that one has occurred changes the probability that the other occurs.

Multiplication Rule for Dependent Events

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

Recall from your previous work with probability in Algebra II that $P(B|A)$ is the conditional probability of event B given that event A occurred. If event A is *picking a fudge-filled piece on the first selection* and event B is *picking a caramel-filled piece on the second selection*, then $P(B|A)$ represents the probability of picking a caramel-filled piece second knowing that a fudge-filled piece was selected first.

Adapted from source material by Alberto Dominguez

11. If $A1$ is the event *picking a fudge-filled piece on the first selection* and $B2$ is the event *picking a caramel-filled piece on the second selection*, what does $P(A1 \text{ and } B2)$ represent? Find $P(A1 \text{ and } B2)$.

    $P(A1 \text{ and } B2)$ represents the probability of picking a fudge-filled piece first and a caramel-filled piece second:

    $\frac{4}{15} \cdot \frac{3}{14} \approx 0.057$

12. What does $P(B1 \text{ and } A2)$ represent? Calculate this probability.

    $P(B1 \text{ and } A2)$ represents the probability of picking a caramel-filled piece first and a fudge-filled piece second:

    $\frac{3}{15} \cdot \frac{4}{14} \approx 0.057$

13. If C represents selecting a coconut-filled piece of chocolate, what does $P(A1 \text{ and } C2)$ represent? Find this probability.

    $P(A1 \text{ and } C2)$ represents the probability of picking a fudge-filled piece first and a coconut-filled piece second:

    $\frac{4}{15} \cdot \frac{2}{14} \approx 0.038$

14. Find the probability that both the first and second pieces selected are filled with chocolate whip.

    $\frac{4}{15} \cdot \frac{3}{14} \approx 0.057$

Exercises 15–17

Exercises 15–17

15. For each of the following, write the probability as the intersection of two events. Then, indicate whether the two events are independent or dependent, and calculate the probability of the intersection of the two events occurring.

    a. The probability of selecting a 6 from the first draw and a 7 on the second draw when two balls are selected without replacement from a container with 10 balls numbered 1 to 10

    Dependent

    $P(6 \text{ first and } 7 \text{ second}) = P(6 \text{ first}) \cdot P(7 \text{ second}|6 \text{ first}) = \frac{1}{10} \cdot \frac{1}{9} = \frac{1}{90} \approx 0.011$

    b. The probability of selecting a 6 on the first draw and a 7 on the second draw when two balls are selected with replacement from a container with 10 balls numbered 1 to 10

    Independent

    $P(6 \text{ first and } 7 \text{ second}) = P(6 \text{ first}) \cdot P(7 \text{ second}) = \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100} = 0.01$

    c. The probability that two people selected at random in a shopping mall on a very busy Saturday both have a birthday in the month of June. Assume that all 365 birthdays are equally likely, and ignore the possibility of a February 29 leap-year birthday.

    Independent

    $P(\text{first person June birthday and second person June birthday}) =$

Adapted from source material by Alberto Dominguez

$P(\text{first person June birthday}) \cdot P(\text{second person June birthday}) =$

$$\frac{30}{365} \cdot \frac{30}{365} = 0.0068$$

d. The probability that two socks selected at random from a drawer containing 10 black socks and 6 white socks will both be black

Dependent

$P(\text{first sock black and second sock black}) =$
$P(\text{first sock black}) \cdot P(\text{second sock black }|\text{first sock black})$

$$= \frac{10}{16} \cdot \frac{9}{15}$$
$$= 0.375$$

16. A gumball machine has gumballs of 4 different flavors: sour apple (A), grape (G), orange (O), and cherry (C). There are six gumballs of each flavor. When $50¢$ is put into the machine, two random gumballs come out. The event $C1$ means a cherry gumball came out first, the event $C2$ means a cherry gumball came out second, the event $A1$ means a sour apple gumball came out first, and the event $G2$ means a grape gumball came out second.

a. What does $P(C2|C1)$ mean in this context?

The probability of the second gumball being cherry knowing the first gumball was cherry

b. Find $P(C1 \text{ and } C2)$.

$$\frac{6}{24} \cdot \frac{5}{23} \approx 0.0543$$

c. Find $P(A1 \text{ and } G2)$.

$$\frac{6}{24} \cdot \frac{6}{23} \approx 0.0652$$

17. Below are the approximate percentages of the different blood types for people in the United States.

Type O 44%

Type A 42%

Type B 10%

Type AB 4%

Consider a group of 100 people with a distribution of blood types consistent with these percentages. If two people are randomly selected with replacement from this group, what is the probability that

a. Both people have type O blood?

$0.44 \cdot 0.44 = 0.1936$

b. The first person has type A blood and the second person has type AB blood?

$0.42 \cdot 0.04 = 0.0168$

Closing

Adapted from source material by Alberto Dominguez

Summary

- **Two events are independent if knowing that one occurs does not change the probability that the other occurs.**
- **Two events are dependent if knowing that one occurs changes the probability that the other occurs.**
- **GENERAL MULTIPLICATION RULE:**

  $\boldsymbol{P(A \text{ and } B) = P(A) \cdot P(B|A)}$

  If A and B are independent events, then $P(B|A) = P(B)$.

Adapted from source material by Alberto Dominguez

Name ____ Date ____________________

Exit Ticket

1. Serena is in a math class of 20 students. Each day for a week (Monday to Friday), a student in Serena's class is randomly selected by the teacher to explain a homework problem. Once a student's name is selected, that student is not eligible to be selected again that week.

   a. What is the probability that Serena is selected on Monday?

   b. What is the probability that Serena is selected on Tuesday given that she was not selected on Monday?

   c. What is the probability that she will be selected on Friday given that she was not selected on any of the other days?

2. Suppose A represents Serena being selected, and B represents Dominic (another student in class) being selected. The event $A1$ means Serena was selected on Monday, and the event $B2$ means Dominic was selected on Tuesday. The event $B1$ means Dominic was selected on Monday, and the event $A2$ means Serena was selected on Tuesday.

   a. Explain in words what $P(A1 \text{ and } B2)$ represents, and then calculate this probability.

   b. Explain in words what $P(B1 \text{ and } A2)$ represents, and then calculate this probability.

Adapted from source material by Alberto Dominguez

Exit Ticket Sample Solutions

1. **Serena is in a math class of 20 students. Each day for a week (Monday to Friday), a student in Serena's class is randomly selected by the teacher to explain a homework problem. Once a student's name is selected, that student is not eligible to be selected again that week.**

   a. **What is the probability that Serena is selected on Monday?**

   $\frac{1}{20} = 0.05$

   b. **What is the probability that Serena is selected on Tuesday given that she was not selected on Monday?**

   $\frac{1}{19} \approx 0.0526$

   c. **What is the probability that she will be selected on Friday given that she was not selected on any of the other days?**

   $\frac{1}{16} = 0.0625$

2. **Suppose A represents Serena being selected, and B represents Dominic (another student in class) being selected. The event $A1$ means Serena was selected on Monday, and the event $B2$ means Dominic was selected on Tuesday. The event $B1$ means Dominic was selected on Monday, and the event $A2$ means Serena was selected on Tuesday.**

   a. **Explain in words what $P(A1 \text{ and } B2)$ represents, and then calculate this probability.**

   It is the probability that Serena is selected on Monday and Dominic is selected on Tuesday.

   $P(A1 \text{ and } B2) = \frac{1}{20} \cdot \frac{1}{19} \approx 0.0026$

   b. **Explain in words what $P(B1 \text{ and } A2)$ represents, and then calculate this probability.**

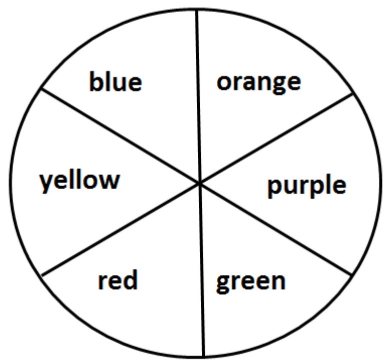
   It is the probability that Dominic is selected on Monday and Serena is selected on Tuesday.

   $P(B1 \text{ and } A2) = \frac{1}{20} \cdot \frac{1}{19} \approx 0.0026$

Adapted from source material by Alberto Dominguez

Problem Set Sample Solutions

1. In a game using the spinner below, a participant spins the spinner twice. If the spinner lands on red both times, the participant is a winner.



   a. The event *participant is a winner* can be thought of as the intersection of two events. List the two events.

   The first spin lands on red, and the second spin lands on red.

   b. Are the two events independent?

   Independent—knowing the first spin landed on red does not change the probability of the second spin landing on red.

   c. Find the probability that a participant wins the game.

   $\frac{1}{6} \cdot \frac{1}{6} \approx 0.0278$

2. The overall probability of winning a prize in a weekly lottery is $\frac{1}{32}$. What is the probability of winning a prize in this lottery three weeks in a row?

   $\frac{1}{32} \cdot \frac{1}{32} \cdot \frac{1}{32} \approx 0.00003$

3. A Gallup poll reported that 28% of adults (age 18 and older) eat at a fast food restaurant about once a week. Find the probability that two randomly selected adults would both say they eat at a fast food restaurant about once a week.

   $0.28 \cdot 0.28 = 0.0784$

Adapted from source material by Alberto Dominguez

4. In the game Scrabble, there are a total of 100 tiles. Of the 100 tiles, 42 tiles have the vowels A, E, I, O, and U printed on them, 56 tiles have the consonants printed on them, and 2 tiles are left blank.

   a. If tiles are selected at random, what is the probability that the first tile drawn from the pile of 100 tiles is a vowel?

   $\frac{42}{100} = 0.42$

   b. If tiles drawn are not replaced, what is the probability that the first two tiles selected are both vowels?

   $\frac{42}{100} \cdot \frac{41}{99} \approx 0.174$

   c. Event A is *drawing a vowel*, event B is *drawing a consonant*, and event C is *drawing a blank tile*. $A1$ means a vowel is drawn on the first selection, $B2$ means a consonant is drawn on the second selection, and $C2$ means a blank tile is drawn on the second selection. Tiles are selected at random and without replacement.

   i. Find $P(A1 \text{ and } B2)$. $= \frac{42}{100} \cdot \frac{56}{99} \approx 0.238$

   ii. Find $P(A1 \text{ and } C2)$. $= \frac{42}{100} \cdot \frac{2}{99} \approx 0.008$

   iii. Find $P(B1 \text{ and } C2)$. $= \frac{56}{100} \cdot \frac{2}{99} \approx 0.011$

5. To prevent a flooded basement, a homeowner has installed two special pumps that work automatically and independently to pump water if the water level gets too high. One pump is rather old and does not work 28% of the time, and the second pump is newer and does not work 9% of the time. Find the probability that both pumps will fail to work at the same time.

   $0.28 \cdot 0.09 \approx 0.025$

6. According to a recent survey, approximately 77% of Americans get to work by driving alone. Other methods for getting to work are listed in the table below.

Method of Getting to Work	Percent of Americans Using This Method
Taxi	0.1%
Motorcycle	0.2%
Bicycle	0.4%
Walk	2.5%
Public Transportation	4.7%
Carpool	10.7%
Drive Alone	77%
Work at Home	3.7%
Other	0.7%

   a. What is the probability that a randomly selected worker drives to work alone?

   0.77

   b. What is the probability that two workers selected at random with replacement both drive to work alone?

   Assume independent $0.77 \cdot 0.77 \approx 0.593$

Adapted from source material by Alberto Dominguez

7. A bag of M&Ms contains the following distribution of colors:

9 blue

6 orange

5 brown

5 green

4 red

3 yellow

Three M&Ms are randomly selected without replacement. Find the probabilities of the following events.

a. All three are blue.

$\frac{9}{32} \cdot \frac{8}{31} \cdot \frac{7}{30} \approx 0.017$

b. The first one selected is blue, the second one selected is orange, and the third one selected is red.

$\frac{9}{32} \cdot \frac{6}{31} \cdot \frac{4}{30} \approx 0.007$

c. The first two selected are red, and the third one selected is yellow.

$\frac{4}{32} \cdot \frac{3}{31} \cdot \frac{3}{30} \approx 0.001$

8. Suppose in a certain breed of dog, the color of fur can either be tan or black. Eighty-five percent of the time, a puppy will be born with tan fur, while 15% of the time, the puppy will have black fur. Suppose in a future litter, six puppies will be born.

a. Are the events *having tan fur* and *having black fur* independent?

Yes. Knowing the color of fur for one puppy doesn't affect the probability of fur color for another puppy.

b. What is the probability that one puppy in the litter will have black fur and another puppy will have tan fur?

$0.15 \cdot 0.85 = 0.1275$

c. What is the probability that all six puppies will have tan fur?

$0.85 \cdot 0.85 \cdot 0.85 \cdot 0.85 \cdot 0.85 \cdot 0.85 \approx 0.377$

d. Is it likely for three out of the six puppies to be born with black fur? Justify mathematically.

$0.15 \cdot 0.15 \cdot 0.15 = 0.003\,375$

No. The probability of three puppies being born with black fur is $0.003\,375$. This is not likely to happen.

Adapted from source material by Alberto Dominguez

9. **Suppose that in the litter of six puppies from Exercise 8, five puppies are born with tan fur, and one puppy is born with black fur.**

   a. **You randomly pick up one puppy. What is the probability that puppy will have black fur?**

   $\frac{1}{6} \approx 0.167$

   b. **You randomly pick up one puppy, put it down, and randomly pick up a puppy again. What is the probability that both puppies will have black fur?**

   $\frac{1}{6} \cdot \frac{1}{6} \approx 0.028$

   c. **You randomly pick up two puppies, one in each hand. What is the probability that both puppies will have black fur?**

   0; *this outcome can never happen since there is only one black puppy.*

   d. **You randomly pick up two puppies, one in each hand. What is the probability that both puppies will have tan fur?**

   $\frac{5}{6} \cdot \frac{4}{5} \approx 0.667$

Adapted from source material by Alberto Dominguez

Lesson 2.2: Counting Rules—The Fundamental Counting Principle and Permutations

Outcomes

- Use the fundamental counting principle to determine the number of different possible outcomes for a chance experiment consisting of a sequence of steps.
- Calculate the number of different permutations of a set of n distinct items.
- Calculate the number of different permutations of k items from a set of n distinct items.

Classwork

Example 1: Fundamental Counting Principle

Example 1: Fundamental Counting Principle

A restaurant offers a fixed-price dinner menu for $\$30$. The dinner consists of three courses, and the diner chooses one item for each course.

The menu is shown below:

First Course	Second Course	Third Course
Salad	Burger	Cheesecake
Tomato Soup	Grilled Shrimp	Ice Cream Sundae
French Onion Soup	Mushroom Risotto	
	Ravioli	

Exercises 1–4

In general, how can you find the number of possibilities in situations like this?

- The number of possibilities is the same as the product of the choices for each course (or clothing option). So, in general, you can multiply the number of choices that each option can occur.

This generalization is known as the *fundamental counting principle*. Instead of making a list or a tree diagram, another method for finding the total number of possibilities is to use the fundamental counting principle. Suppose that a process involves a sequence of *steps* or *events*. Let n_1 be the number of ways the first step or event can occur and n_2 be the number of ways the second step or event can occur. Continuing in this way, let n_k be the number of ways the k^{th} stage or event can occur. Then the total number of different ways the process can occur is:
$n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_k$.

In the fixed-price dinner example, there are 3 choices for the first course (step 1), 4 choices for the second course (step

Adapted from source material by Alberto Dominguez

2), and 2 choices for the third course (step 3). Using the fundamental counting principle, there are $3 \cdot 4 \cdot 2$, or 24 total choices. For the fixed- dinner example, you could draw three boxes and label the boxes as shown. Then, identify the number of choices for each box.

First Course (3 choices)	Second Course (4 choices)	Third Course (2 choices)
$\square$	$\square$	$\square$

Exercises 1–4

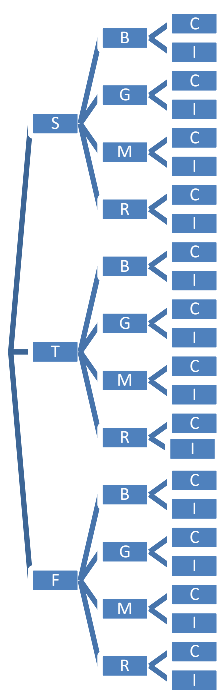
1. Make a list of all the different dinner fixed-price meals that are possible. How many different meals are possible?

Note: The list below shows all the possibilities using abbreviations of the items.

They are listed in order: first course choice, second course choice, and third course choice. For example, SBC would represent salad, burger, and cheesecake.

There are 24 different dinner fixed-price meals possible.

SBC	*TBC*	*FBC*
SBI	*TBI*	*FBI*
SGC	*TGC*	*FGC*
SGI	*TGI*	*FGI*
SMC	*TMC*	*FMC*
SMI	*TMI*	*FMI*
SRC	*TRC*	*FRC*
SRI	*TRI*	*FRI*



Adapted from source material by Alberto Dominguez

2. **For many computer tablets, the owner can set a 4-digit pass code to lock the device.**

   a. **How many digits could you choose from for the first number of the pass code?**

   10

   b. **How many digits could you choose from for the second number of the pass code? Assume that the numbers can be repeated.**

   10

   c. **How many different 4-digit pass codes are possible?**

   $10 \cdot 10 \cdot 10 \cdot 10 = 10000$

   There are 10 choices for the first digit, 10 choices for the second digit, 10 choices for the third digit, and 10 choices for the fourth digit. I used the fundamental counting principle and multiplied the number of choices for each digit together to get the number of possible pass codes.

   d. **How long (in hours) would it take someone to try every possible code if it takes three seconds to enter each possible code?**

   It would take $30,000$ seconds, which is $8\frac{1}{3}$ hr.

3. **The store at your school wants to stock sweatshirts that come in four sizes (small, medium, large, xlarge) and in two colors (red and white). How many different types of sweatshirts will the store have to stock?**

   $4 \cdot 2 = 8$

4. **The call letters for all radio stations in the United States start with either a W (east of the Mississippi River) or a K (west of the Mississippi River) followed by three other letters that can be repeated. How many different call letters are possible?**

   $2 \cdot 26 \cdot 26 \cdot 26 = 35152$

Example 2: Permutations

This example introduces the definition of *permutations*. Order matters for a permutation.

- How many digits can you choose from for the first digit?
  - 10
- How many digits can you choose from for the second digit? (Remember—no repeats)
  - 9
- How many digits can you choose from for the third digit?
  - 8
- How many digits can you choose from for the fourth digit?
  - 7

Adapted from source material by Alberto Dominguez

- How many different 4-digit pass codes are possible if digits cannot be repeated?
  - $10 \cdot 9 \cdot 8 \cdot 7 = 5040$
- Explain how the fundamental counting principle allows you to make this calculation.
  - *The process of choosing the four digits for the pass code involves a sequence of events. There are* 10 *choices for the first digit,* 9 *choices for the second digit, and so on. So, I can multiply the number of choices that each digit in the pass code can occur.*

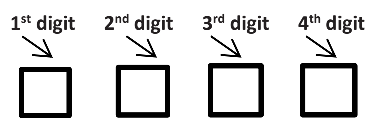
Now introduce the notation for permutations. Explain that finding the number of ordered arrangements of the digits in the pass code is an example of the number of permutations of 10 things taken 4 at a time. This can be written: ${}_{10}P_4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$.

Example 2: Permutations

Suppose that the 4-digit pass code a computer tablet owner uses to lock the device *cannot* have any digits that repeat. For example, 1234 is a valid pass code. However, 1123 is not a valid pass code since the digit 1 is repeated.

An arrangement of four digits with no repeats is an example of a permutation. A permutation is an arrangement in a certain order (a sequence).

How many different 4-digit pass codes are possible if digits cannot be repeated?



Exercises 5–9

Exercises 5–9

5. Suppose a password requires three distinct letters. Find the number of permutations for the three letters in the code if the letters may not be repeated.

$\boldsymbol{{}_{26}P_3 = 26 \cdot 25 \cdot 24 = 15600}$

Adapted from source material by Alberto Dominguez

6. **The high school track has 8 lanes. In the 100-meter dash, there is a runner in each lane. Find the number of ways that 3 out of the 8 runners can finish first, second, and third.**

   ${}_8P_3 = 8 \cdot 7 \cdot 6 = 336$

7. **There are 12 singers auditioning for the school musical. In how many ways can the director choose first a lead singer and then a stand-in for the lead singer?**

   ${}_{12}P_2 = 12 \cdot 11 = 132$

8. **A home security system has a pad with 9 digits (1 to 9). Find the number of possible 5-digit pass codes:**

   a. **If digits can be repeated.**

      $9 \cdot 9 \cdot 9 \cdot 9 \cdot 9 = 59049$

   b. **If digits cannot be repeated.**

      ${}_9P_5 = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 15120$

MP.8

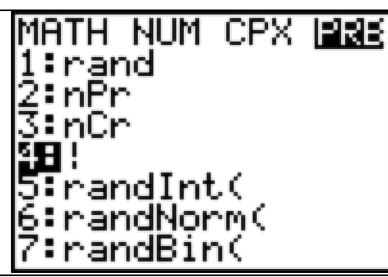
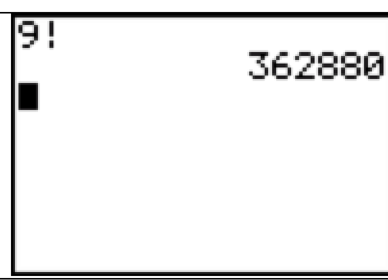
9. **Based on the patterns observed in Exercises 5–8, describe a general formula that can be used to find the number of permutations of n things taken r at a time, or ${}_n\mathrm{P}_r$.**

   Based on the answers to the exercises, a permutation of n things taken r at a time can be found using the formula:

   $$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1).$$

Example 3: Factorials and Permutations

Calculating a Factorial on the TI-84 graphing calculator

1. Enter the integer on the Home screen.	
2. Select the MATH menu.	
3. Scroll to PRB.	
4. Select option 4: !	
5. Press ENTER	

Adapted from source material by Alberto Dominguez

Example 3: Factorials and Permutations

You have purchased a new album with 12 music tracks and loaded it onto your MP3 player. You set the MP3 player to play the 12 tracks in a random order (no repeats). How many different orders could the songs be played in?

This is the permutation of 12 things taken 12 at a time, or
$_{12}P_{12} = 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 479\,001\,600$.
The notation $12!$ is read "12 factorial" and equals $12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$.

Factorials and Permutations

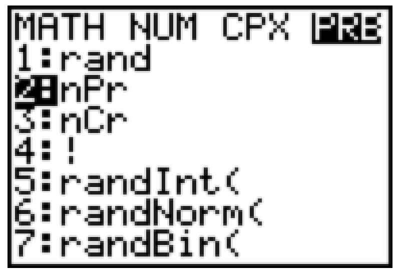
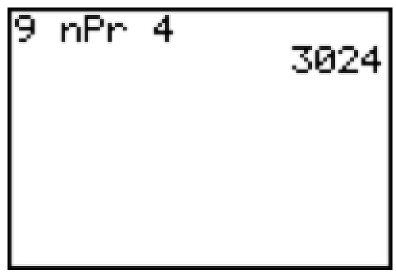
The factorial of a nonnegative integer n is
$$n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot 1.$$
Note: $0!$ is defined to equal 1.

The number of permutations can also be found using factorials. The number of permutations of n things taken r at a time is
$$_nP_r = \frac{n!}{(n-r)!}.$$

- Rewrite the permutation formula by expanding the factorial notation.
  - *Permutation:*
$$\begin{aligned} _nP_r &= \frac{n!}{(n-r)!} \\ &= \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1) \cdot (n-r) \cdot (n-r-1) \cdot \ldots \cdot 1}{(n-r) \cdot (n-r-1) \cdot \ldots \cdot 1} \\ &= n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1) \end{aligned}$$
- How does this formula compare with the general formula you described in Exercise 9?
  - *The formulas are the same. My formula from Exercise 9 was*
$$_nP_r = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1).$$
The permutation formula is just another version of the fundamental counting principle.

Adapted from source material by Alberto Dominguez

Calculating Permutations on the TI-84 graphing calculator

1. Enter the integer on the Home screen.	
2. Select the MATH menu.	
3. Scroll to PRB.	
4. Select option 2: nPr.	
5. Enter 4, and press ENTER.	

Exercises 10–15

Exercises 10–15

10. If $9!$ is $362,880$, find $10!$.

$3,628,800$

11. How many different ways can the 16 numbered pool balls be placed in a line on the pool table?

$16!$ *or* ${}_{16}P_{16} \approx 2 \cdot 10^{13}$

12. Ms. Smith keeps eight different cookbooks on a shelf in one of her kitchen cabinets. How many ways can the eight cookbooks be arranged on the shelf?

$8!$ *or* ${}_{8}P_{8} = 40320$

13. How many distinct 4-letter groupings can be made with the letters from the word *champion* if letters may not be repeated?

$${}_{8}P_{4} = \frac{8!}{(8-4)!} = 1680$$

Adapted from source material by Alberto Dominguez

14. There are 12 different rides at an amusement park. You buy five tickets that allow you to ride on five different rides. In how many different orders can you ride the five rides? How would your answer change if you could repeat a ride?

Different order: ${}_{12}P_5 = \frac{12!}{(12-5)!} = 95040$

Repeat rides: $12^5 = 248\,832$

15. In the summer Olympics, 12 divers advance to the finals of the 3-meter springboard diving event. How many different ways can the divers finish 1st, 2nd, or 3rd?

$${}_{12}P_3 = \frac{12!}{(12-3)!} = 1320$$

Closing

Summary

- Let n_1 be the number of ways the first step or event can occur and n_2 be the number of ways the second step or event can occur. Continuing in this way, let n_k be the number of ways the k^{th} stage or event can occur. Then, based on the fundamental counting principle, the total number of different ways the process can occur is $n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_k$.
- The factorial of a nonnegative integer n is

  $n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot 1$.

  Note: $0!$ is defined to equal 1.
- The number of permutations of n things taken r at a time is

  ${}_n\mathrm{P}_r = \frac{n!}{(n-r)!}$.

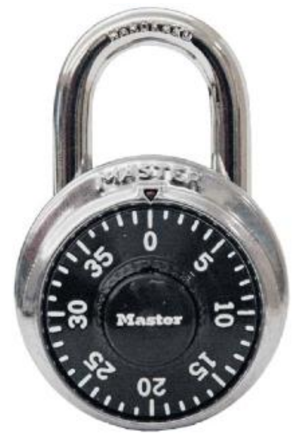
Adapted from source material by Alberto Dominguez

Name ____ Date ____________________

Counting Rules—The Fundamental Counting Principle and Permutations

Exit Ticket

1. The combination for the lock shown below consists of three numbers.
   a. If the numbers can be repeated, how many different combinations are there?



   b. If the numbers cannot be repeated, how many different combinations are there?

2. Jacqui is putting together sets of greeting cards for a school fundraiser. There are four different card options, two different colored envelopes, and four different sticker designs. A greeting card set consists of one type of card, one color for the envelopes, and one sticker design. How many different ways can Jacqui arrange the greeting card sets?

Adapted from source material by Alberto Dominguez

Exit Ticket Sample Solutions

1. The combination for the lock shown below consists of three numbers.

   a. If the numbers can be repeated, how many different combinations are there?

   $40^3 = 64000$

   Because numbers can be repeated, there are 40 choices for each of the three digits. Therefore, I applied the fundamental counting principle.

   b. If the numbers cannot be repeated, how many different combinations are there?

   ${}_{40}P_3 = 59280$

   Since numbers cannot be repeated, this is an example of a permutation.

2. Jacqui is putting together sets of greeting cards for a school fundraiser. There are four different card options, two different colored envelopes, and four different sticker designs. A greeting card set consists of one type of card, one color for the envelopes, and one sticker design. How many different ways can Jacqui arrange the greeting card sets?

   $4 \cdot 2 \cdot 4 = 32$

   By using the fundamental counting principle, there are 32 ways to arrange the sets.

Problem Set Sample Solutions

1. For each of the following, show the substitution in the permutation formula, and find the answer.

   a. ${}_{4}P_4$

   $$\frac{4!}{(4-4)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{0!} = 24$$

   b. ${}_{10}P_2$

   $$\frac{10!}{(10-2)!} = \frac{10!}{8!} = 10 \cdot 9 = 90$$

   c. ${}_{5}P_1$

   $$\frac{5!}{(5-1)!} = \frac{5!}{4!} = 5$$

2. A serial number for a TV begins with three letters, is followed by six numbers, and ends in one letter. How many different serial numbers are possible? Assume the letters and numbers can be repeated.

   $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 26 \approx 4.6 \cdot 10^{11}$

Adapted from source material by Alberto Dominguez

3. In a particular area code, how many phone numbers (###-####) are possible? The first digit cannot be a zero, and assume digits can be repeated.

   $9 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 9\,000\,000$

4. There are four teams in the AFC East division of the National Football League: Bills, Jets, Dolphins, and Patriots. How many different ways can two of the teams finish first and second?

   ${}_4P_2 = 12$

5. How many ways can 3 of 10 students come in first, second, and third place in a spelling contest if there are no ties?

   ${}_{10}P_3 = 720$

6. In how many ways can a president, a treasurer, and a secretary be chosen from among nine candidates if no person can hold more than one position?

   ${}_9P_3 = 504$

7. How many different ways can a class of 22 second graders line up to go to lunch?

   $22! = {}_{22}P_{22} \approx 1.1 \cdot 10^{21}$

8. Describe a situation that could be modeled by using ${}_5P_2$.

   Sample answer: Suppose that there are five members of a family living at home. The first one home has to take out the garbage, and the second one home has to walk the dog. ${}_5P_2$ can be used to model the number of ways the family members can be assigned the different tasks.

9. To order books from an online site, the buyer must open an account. The buyer needs a username and a password.

   a. If the username needs to be eight letters, how many different usernames are possible:

      i. If the letters can be repeated?

         $26^8 \approx 2 \cdot 10^{11}$

      ii. If the letters cannot be repeated?

         ${}_{26}P_8 \approx 6.3 \cdot 10^{10}$

   b. If the password must be eight characters, which can be any of the 26 letters, 10 digits, and 12 special keyboard characters, how many passwords are possible:

      i. If characters can be repeated?

         $48^8 \approx 2.8 \cdot 10^{13}$

      ii. If characters cannot be repeated?

         ${}_{48}P_8 \approx 1.5 \cdot 10^{13}$

Adapted from source material by Alberto Dominguez

c. **How would your answers to part (b) change if the password is case-sensitive? (In other words, *Password* and *password* are considered different because the letter *p* is in uppercase and lowercase.)**

The answers would change because the number of letters that can be used will double to 52, which means the number of characters that can be used is now 74.

So, if characters can be repeated, the answer will be $72^8 \approx 7.2 \cdot 10^{14}$.

If characters cannot be repeated, the answer will be ${}_{72}P_8 \approx 4.8 \cdot 10^{14}$.

10. **Create a scenario to explain why ${}_3P_3 = 3!$.**

Suppose three friends are running in a race. ${}_3P_3$ can be used to model the order in which the three friends finish in 1st, 2nd, and 3rd place. There are three choices for 1st place and two choices for 2nd place, which only leaves one choice for 3rd place. So, there are $3 \cdot 2 \cdot 1$ ways for the friends to finish the race, which is the same as $3!$ or 6.

11. **Explain why ${}_nP_n = n!$ for all positive integers n.**

Using the permutation formula:

$$ {}_nP_r = \frac{n!}{(n-r)!} $$

In this case, $r = n$; therefore:

$$ {}_nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!. $$

Adapted from source material by Alberto Dominguez

Lesson 2.3: Counting Rules—Combinations

Outcome

- Calculate the number of different combinations of k items selected from a set of n distinct items.

Classwork

Example 1: Combinations

This example makes the distinction between situations in which order is important and those in which order is not important. The definition of combinations as a subset of k items selected from a set of n distinct items is formally introduced.

Example 1

Seven speed skaters are competing in an Olympic race. The first-place skater earns the gold medal, the second-place skater earns the silver medal, and the third-place skater earns the bronze medal. In how many different ways could the gold, silver, and bronze medals be awarded? The letters A, B, C, D, E, F, and G will be used to represent these seven skaters.

How can we determine the number of different possible outcomes? How many are there?

Because each outcome is a way of forming an ordered arrangement of 3 things from a set of 7, the total number of possible outcomes is

$${}_7P_3 = \frac{7!}{(7-3)!} = \frac{7!}{4!} = 210.$$

Now consider a slightly different situation. Seven speed skaters are competing in an Olympic race. The top three skaters move on to the next round of races. How many different "top three" groups can be selected?

How is this situation different from the first situation? Would you expect more or fewer possibilities in this situation? Why?

The outcomes in the first situation are medals based on order—first place gets gold, second place gets silver, and third place gets bronze. The outcomes in this situation are not based on order. The top three finishers move on; the others do not. I would expect more possibilities in the first situation because each skater can be first, second, or third and still advance (i.e., there are three different possible finishing positions each skater can attain to earn advancement). In the first situation, there is only one position to earn gold, one for silver, and one for bronze.

Would you consider the outcome where skaters B, C, and A advance to the final to be a different outcome from A, B, and C advancing?

No. This is the same outcome—the same three skaters are advancing to the final competition.

A permutation is an ordered arrangement (a sequence) of k items from a set of n distinct items.

In contrast, a combination is an unordered collection (a set) of k items from a set of n distinct items.

When we wanted to know how many ways there are for seven skaters to finish first, second, and third, order was important. This is an example of a permutation of 3 selected from a set of 7. If we want to know how many possibilities there are for which three skaters will advance to the finals, order is not important. This is an example of a combination of 3 selected from a set of 7.

Adapted from source material by Alberto Dominguez

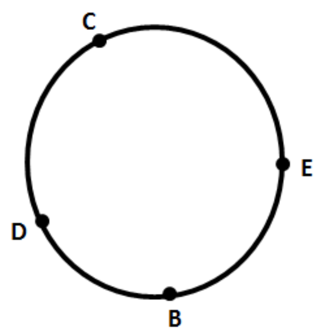
In a permutation, the order of possible outcomes of a situation matters or that outcomes need to happen in a specific order. In a combination, the order of possible outcomes does not matter, or the set of specific outcomes needs to happen in no specific order.

Exercises 1–4

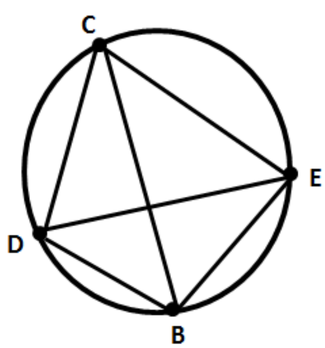
Exercises 1–4

1. **Given four points on a circle, how many different line segments connecting these points do you think could be drawn?**

2. **Draw a circle, and place four points on it. Label the points as shown. Draw segments (chords) to connect all the pairs of points. How many segments did you draw? List each of the segments that you drew. How does the number of segments compare to your answer in Exercise 1?**



Answer:

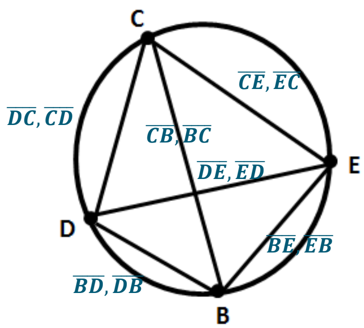


6 segments: $\overline{DC}, \overline{DE}, \overline{DB}, \overline{CB}, \overline{CE}$, ***and*** $\overline{BE}$

You can think of each segment as being identified by a subset of two of the four points on the circle. Chord ED is the same as chord DE. The order of the segment labels is not important. When you count the number of segments (chords), you are counting combinations of two points chosen from a set of four points.

Adapted from source material by Alberto Dominguez

3. **Find the number of permutations of two points from a set of four points. How does this answer compare to the number of segments you were able to draw?**



The diagram is the same, but this answer is double the number of combinations. In a permutation, $\overline{DC}$ is counted as different from $\overline{CD}$. In this case, there are six combinations, and each segment can be represented two ways. 6×2 or 12 permutations

4.

a. **If you add a fifth point to the circle, how many segments (chords) can you draw?**

10 segments

b. **If you add a sixth point, how many segments (chords) can you draw?**

15 segments

Example 2: Combinations Formula

Example 2

Let's look closely at the four examples we have studied so far.

Choosing gold, silver, and bronze medal skaters	Choosing groups of the top three skaters
Finding the number of segments that can be drawn connecting two points out of four points on a circle	Finding the number of *unique* segments that can be drawn connecting two points out of four points on a circle

What do you notice about the way these are grouped?

Sample responses: The examples on the left are permutations, and the examples on the right are combinations. In the left examples, the order of the outcomes matters, while in those on the right, the order does not matter.

Adapted from source material by Alberto Dominguez

> **The number of combinations of k items selected from a set of n distinct items is**
>
> $$_nC_k = \frac{_nP_k}{k!} \text{ or } _nC_k = \frac{n!}{k!\,(n-k)!}.$$

The number of permutations of three skaters from the seven is found in Example 1 to be

$$_7P_3 = \frac{7!}{(7-3)!} = \frac{7!}{4!} = 210.$$

This means that there are 210 different ordered arrangements of three skaters from a set of seven skaters.

There are fewer combinations because, when order is not important, we do not want to count different orders of the same three skaters as different outcomes.

$$3 \cdot 2 \cdot 1 = 3! = 6$$

Consider the set of skaters A, B, and C. There are 6 different ordered arrangements of these three skaters, each of which is counted with the permutations formula. So, if we want combinations of 3 from a set of 7, you would need to divide the number of permutations by 3!. Then,

$$_7C_3 = \frac{_7P_3}{3!} = \frac{7!}{3!\,(7-3)!} = \frac{7!}{3!\,4!} = 35.$$

In general,

$$_nC_k = \frac{_nP_k}{k!} = \frac{n!}{k!\,(n-k)!}.$$

For example, $_9C_4$:

$$\begin{aligned} _9C_4 &= \frac{9!}{4!\,(9-4)!} \\ &= \frac{9!}{4! \cdot 5!} \\ &= \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(4 \cdot 3 \cdot 2 \cdot 1)(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} \\ &= \frac{9 \cdot 8 \cdot 7 \cdot 6}{(4 \cdot 3 \cdot 2 \cdot 1)} \\ &= \frac{3024}{24} \\ &= 126 \end{aligned}$$

- *The number of segments (chords) that can be drawn with four points on a circle is the number of combinations of two points selected from a set of four points.*

$$_4C_2 = \frac{_4P_2}{2!} = 6 \text{ or } _4C_2 = \frac{4!}{2!\,(4-2)!} = 6$$

Adapted from source material by Alberto Dominguez

Exercises 5–11

Exercises 5–11

5. **Find the value of each of the following:**

   a. ${}_{9}C_{2}$ **36**

   b. ${}_{7}C_{7}$ **1**

   c. ${}_{8}C_{0}$ **1**

   d. ${}_{15}C_{1}$ **15**

6. **Find the number of segments (chords) that can be drawn for each of the following:**

   a. 5 points on a circle $\quad {}_{5}C_{2} = 10$

   b. 6 points on a circle $\quad {}_{6}C_{2} = 15$

   c. 20 points on a circle $\quad {}_{20}C_{2} = 190$

   d. n points on a circle $\quad {}_{n}C_{2} = \frac{{}_{n}P_{2}}{2!} = \frac{n!}{2!\,(n-2)!}$

7. **For each of the following questions, indicate whether the question posed involves permutations or combinations. Then, provide an answer to the question with an explanation for your choice.**

   a. **A student club has 20 members. How many ways are there for the club to choose a president and a vice president?**

   Permutations, 380. The role of president is different from that of vice president. The order of outcomes matters.

   b. **A football team of 50 players will choose two co-captains. How many different ways are there to choose the two co-captains?**

   Combinations, 1,225. Regardless of order, two players will attain the same outcome of co-captain.

   c. **There are seven people who meet for the first time at a meeting. They shake hands with each other and introduce themselves. How many handshakes have been exchanged?**

   Combinations, 21. People only shake hands with those they have not greeted yet. Each greeting is unique. Person 1 shaking hands with person 2 is the same event as person 2 shaking hands with person 1.

   d. **At a particular restaurant, you must choose two different side dishes to accompany your meal. If there are eight side dishes to choose from, how many different possibilities are there?**

   Combinations, 28. The order of dishes chosen does not matter. Choosing veggies and mac and cheese is the same as choosing mac and cheese and veggies.

   e. **How many different four-letter sequences can be made using the letters A, B, C, D, E, and F if letters may not be repeated?**

   Permutations, 360. Each four-letter sequence is unique. ABCD is different from BACD even though they contain identical letters.

Adapted from source material by Alberto Dominguez

8. How many ways can a committee of 5 students be chosen from a student council of 30 students? Is the order in which the members of the committee are chosen important?

   No. ${}_{30}C_5 = 142\,506$

9. Brett has ten distinct T-shirts. He is planning on going on a short weekend trip to visit his brother in college. He has enough room in his bag to pack four T-shirts. How many different ways can he choose four T-shirts for his trip?

   ${}_{10}C_4 = 210$

MP.3

10. How many three-topping pizzas can be ordered from the list of toppings below? Did you calculate the number of permutations or the number of combinations to get your answer? Why did you make this choice?

Pizza Toppings

sausage	pepperoni	meatball	onions	olives	spinach
pineapple	ham	green peppers	mushrooms	bacon	hot peppers

${}_{12}C_3 = 220$*; I used combinations because the order of toppings chosen does not matter. My pizza is the same if I order pepperoni, olives, and mushrooms or olives, pepperoni, and mushrooms.*

I calculated the number of combinations—we are choosing 3 *items from the* 12 *distinct toppings, and the order that the* 3 *toppings are chosen is not important.*

11. How can you can distinguish a question about permutations from a question about combinations.

    The key trigger is whether order is important or not.

Closing

Summary

A combination is a subset of k items selected from a set of n distinct items.

The number of combinations of k items selected from a set of n distinct items is

$${}_nC_k = \frac{{}_nP_k}{k!} \quad \text{or} \quad {}_nC_k = \frac{n!}{k!\,(n-k)!}.$$

Adapted from source material by Alberto Dominguez

Name ______________________________ Date ____________________

Exit Ticket

1. Timika is a counselor at a summer camp for young children. She wants to take 20 campers on a hike and wants to choose a pair of students to lead the way. In how many ways can Timika choose this pair of children?

2. Sean has 56 songs on his MP3 player. He wants to randomly select 6 of the songs to use in a school project. How many different groups of 6 songs could Sean select? Did you calculate the number of permutations or the number of combinations to get your answer? Why did you make this choice?

3. A fast food restaurant has the following options for toppings on their hamburgers: mustard, ketchup, mayo, onions, pickles, lettuce, and tomato. In how many ways could a customer choose four different toppings from these options?

4. Seven colored balls (red, blue, yellow, black, brown, white, and orange) are in a bag. A sample of three balls is selected without replacement. How many different samples are possible?

Adapted from source material by Alberto Dominguez

Exit Ticket Sample Solutions

1. **Timika is a counselor at a summer camp for young children. She wants to take 20 campers on a hike and wants to choose a pair of students to lead the way. In how many ways can Timika choose this pair of children?**

   $_{20}C_2 = 190$

2. **Sean has 56 songs on his MP3 player. He wants to randomly select 6 of the songs to use in a school project. How many different groups of 6 songs could Sean select? Did you calculate the number of permutations or the number of combinations to get your answer? Why did you make this choice?**

   $_{56}C_6 = 32\,468\,436$

   I calculated the number of combinations—choosing 6 songs from 56 distinct songs, and order is not important.

3. **A fast food restaurant has the following options for toppings on their hamburgers: mustard, ketchup, mayo, onions, pickles, lettuce, and tomato. In how many ways could a customer choose 4 different toppings from these options?**

   $_7C_4 = 35$

4. **Seven colored balls (red, blue, yellow, black, brown, white, and orange) are in a bag. A sample of three balls is selected without replacement. How many different samples are possible?**

   $_7C_3 = 35$

Problem Set Sample Solutions

1. **Find the value of each of the following:**
   a. $_9C_8$ 9
   b. $_9C_1$ 9
   c. $_9C_9$ 1

2. **Explain why $_6C_4$ is the same value as $_6C_2$.**

   $_6C_4 = \dfrac{6!}{4! \cdot 2!}$ $_6C_2 = \dfrac{6!}{2! \cdot 4!}$

   The denominators are the same. This is because the number of ways to choose four from a set of six is the same as the number of ways to select which two to exclude.

3. **Pat has 12 books he plans to read during the school year. He decides to take 4 of these books with him while on winter break vacation. He decides to take *Harry Potter and the Sorcerer's Stone* as one of the books. In how many ways can he select the remaining 3 books?**

   $_{11}C_3 = 165$

Adapted from source material by Alberto Dominguez

4. **In a basketball conference of 10 schools, how many conference basketball games are played during the season if the teams all play each other exactly once?**

$_{10}C_2 = 45$

5. **Which scenario or scenarios below are represented by $_9C_3$?**

A. **The number of ways 3 of 9 people can sit in a row of 3 chairs**

B. **The number of ways to pick 3 students out of 9 students to attend an art workshop**

C. **The number of ways to pick 3 different entrees from a buffet line of 9 different entrees**

B and C

6. **Explain why $_{10}C_3$ would not be used to solve the following problem:**

There are 10 runners in a race. How many different possibilities are there for the runners to finish first, second, and third?

This is an example of a permutation. The order of how the runners finish is important.

7. **In a lottery, players must match five numbers plus a bonus number. Five white balls are chosen from 59 white balls numbered from 1 to 59, and one red ball (the bonus number) is chosen from 35 red balls numbered 1 to 35. How many different results are possible?**

White ball: $_{59}C_5 = 5\,006\,386$

Red ball: $_{35}C_1 = 35$

Number of possible results: $5\,006\,386 \cdot 35 = 175\,223\,510$

8. **In many courts, 12 jurors are chosen from a pool of 30 perspective jurors.**

a. **In how many ways can 12 jurors be chosen from the pool of 30 perspective jurors?**

$_{30}C_{12} = 86\,493\,225$

b. **Once the 12 jurors are selected, 2 alternates are selected. The order of the alternates is specified. If a selected juror cannot complete the trial, the first alternate is called on to fill that jury spot. In how many ways can the 2 alternates be chosen after the 12 jury members have been chosen?**

$_{18}P_2 = 306$

9. **A band director wants to form a committee of 4 parents from a list of 45 band parents.**

a. **How many different groups of 4 parents can the band director select?**

$_{45}C_4 = 148\,995$

b. **How many different ways can the band director select 4 parents to serve in the band parents' association as president, vice president, treasurer, and secretary?**

$_{45}P_4 = 3\,575\,880$

Adapted from source material by Alberto Dominguez

c. **Explain the difference between parts (a) and (b) in terms of how you decided to solve each part.**

Part (a) is an example of finding the number of combinations—how many ways can 4 parents be chosen from 45 distinct parents. The order is not important.

Part (b) is an example of finding the number of permutations. The order of how the parents are selected is important.

10. **A cube has faces numbered 1 to 6. If you roll this cube 4 times, how many different outcomes are possible?**

Using the fundamental counting principle: $6^4 = 1296$

11. **Write a problem involving students that has an answer of $_6C_3$.**

Sample answer: There are six seniors on the principal's advisory committee on school improvement. The principal would like to select three of the students to attend a workshop on school improvement. How many ways can the principal select three students out of the six students on the advisory committee?

12. **Suppose that a combination lock is opened by entering a three-digit code. Each digit can be any integer between 0 and 9, but digits may not be repeated in the code. How many different codes are possible? Is this question answered by considering permutations or combinations?**

There are 720 possible codes. This question is answered by considering permutations because in a code, the order of the digits is important.

13. **Six musicians will play in a recital. Three will perform before intermission, and three will perform after intermission. How many different ways are there to choose which three musicians will play before intermission? Is this question answered by considering permutations or combinations?**

There are 20 possible ways to choose the musicians. This question is answered by considering combinations because order is not important if we just care about which group of three is before the intermission.

14. **In a game show, contestants must guess the price of a product. A contestant is given nine cards with the numbers 1 to 9 written on them (each card has a different number). The contestant must then choose three cards and arrange them to produce a price in dollars. How many different prices can be formed using these cards? Is this question answered by considering permutations or combinations?**

There are 504 possible prices. The question is answered by considering permutations because the order of the digits is important. $\$123$ is different from $\$312$.

15.

a. **Using the formula for combinations, show that the number of ways of selecting 2 items from a group of 3 items is the same as the number of ways to select 1 item from a group of 3.**

$_3C_2 = \frac{3!}{2!\cdot 1!}$ $\quad\quad$ $_3C_1 = \frac{3!}{1!\cdot 2!}$

b. **Show that $_nC_k$ *and* $_nC_{n-k}$ are equal. Explain why this makes sense.**

$_nC_k = \frac{n!}{k!\cdot(n-k)!}$ $\quad\quad$ $_nC_{n-k} = \frac{n!}{(n-k)!\cdot k!}$

The denominators are the same. This is because the number of ways to choose k from a set of n is the same as the number of ways to select which $(n-k)$ to exclude.

Adapted from source material by Alberto Dominguez

Lesson 2.4: Using Petermutations and Combinations to Compute Probabilities

Outcomes

- Distinguish between situations involving combinations and situations involving permutations.
- Use permutations and combinations to calculate probabilities.
- Interpret probabilities in context.

Classwork

Exercises 1–6

Exercises 1–6

1. **A high school is planning to put on the musical *West Side Story*. There are 20 singers auditioning for the musical. The director is looking for two singers who could sing a good duet. In how many ways can the director choose two singers from the 20 singers?**

   Indicate if this question involves a permutation or a combination.

   Combination; the order that the singers are selected is not important.

2. **The director is also interested in the number of ways to choose a lead singer and a backup singer. In how many ways can the director choose a lead singer and then a backup singer?**

   Indicate if this question involves a permutation or a combination.

   Permutation; the order is important; the lead is "chosen first," and then the backup singer is chosen.

3. **For each of the following, indicate if it is a problem involving permutations, combinations, or neither, and then answer the question posed.**

   a. **How many groups of five songs can be chosen from a list of 35 songs?**

   Combination; the order of the songs is not important; ${}_{35}C_5 = 324\,632$

   b. **How many ways can a person choose three different desserts from a dessert tray of eight desserts?**

   Combination; the order of the desserts is not important; ${}_{8}C_3 = 56$

   c. **How many ways can a manager of a baseball team choose the lead-off batter and second batter from a baseball team of nine players?**

   Permutation; the order of the batters is important; ${}_{9}P_2 = 72$

   d. **How many ways are there to place seven distinct pieces of art in a row?**

   Permutation; the order of each piece of art is important; ${}_{7}P_7$ ***or*** $7! = 5040$

Adapted from source material by Alberto Dominguez

e. **How many ways are there to randomly select four balls without replacement from a container of 15 balls numbered 1 to 15?**

Combination; the order of the balls is not important; $_{15}C_4 = 1365$

4. **The manager of a large store that sells TV sets wants to set up a display of all the different TV sets that they sell. The manager has seven different TVs that have screen sizes between 37 and 43 inches, nine that have screen sizes between 46 and 52 inches, and twelve that have screen sizes of 55 inches or greater.**

a. **In how many ways can the manager arrange the 37- to 43-inch TV sets?**

$7! = 5040$

b. **In how many ways can the manager arrange the 55-inch or greater TV sets?**

$12! = 479\,001\,600$

c. **In how many ways can the manager arrange all the TV sets if he is concerned about the order they were placed in?**

$28!$ *or* $_{28}P_{28} \approx 3.0 \cdot 10^{29}$

5. **Seven slips of paper with the digits 1 to 7 are placed in a large jar. After thoroughly mixing the slips of paper, two slips are picked without replacement.**

a. **Explain the difference between $_7P_2$ and $_7C_2$ in terms of the digits selected.**

$_7P_2$ is the number of permutations of picking two slips from the seven slips of paper. The value of $_7P_2$ is the number of ways that the two slips can be picked from the seven slips of paper in which the order of the digits is important. For example, if the digit 2 is picked first and the digit 1 is picked second, then 21 is considered a different outcome than if 1 is picked first and 2 is picked second, or 12. $_7C_2$ is the combination of picking two slips from the seven slips. It represents the total number of ways two digits can be selected in which order does not matter. Therefore, 21 and 12 are not counted as different outcomes.

b. **Describe a situation in which $_7P_2$ is the total number of outcomes.**

Sample answer: Seven students from your school are eligible to participate in a tennis competition; however, only one student can compete, with one other student designated as a backup. (The backup will compete if the first student is injured or unable to attend.) $_7P_2$ represents how many different pairings of the seven students could be selected for the competition.

c. **Describe a situation in which $_7C_2$ is the total number of outcomes.**

Sample answer: Two students from seven eligible students will receive a prize. Each person is assigned a number from 1 to 7 (with no duplicates). Their numbers are placed in the jar. Two slips are drawn. The students assigned to the selected digits are the winners and will receive the prizes. $_7C_2$ represents the number of ways two people from the group of seven students could win the prizes.

d. **What is the relationship between $_7P_2$ and $_7C_2$?**

$$_7C_2 = \left(\frac{_7P_2}{2!}\right)$$

Adapted from source material by Alberto Dominguez

6. **If you know ${}_nC_k$, and you also know the value of n and k, how could you find the value of ${}_nP_k$?**

The following steps show the relation to be the value of the combination and the permutation:

Recall that ${}_nC_k = \frac{n!}{k!(n-k)!}$.

Multiply each side of the above equation by $k!$, or

$k!\ {}_nC_k = \frac{n!}{(n-k)!}$.

$\frac{n!}{(n-k)!}$ ***is*** ${}_nP_k$***, therefore:***

$k!\ {}_nC_k = {}_nP_k$.

Example 1: Calculating Probabilities

> **Example 1: Calculating Probabilities**
>
> **In a high school, there are 10 math teachers. The principal wants to form a committee by selecting three math teachers at random. If Mr. H, Ms. B, and Ms. J are among the group of 10 math teachers, what is the probability that all three of them will be on the committee?**
>
> **Because every different committee of 3 is equally likely,**
>
> $$P(\text{these three math teachers will be on the committee}) = \frac{\text{number of ways Mr. H, Ms. B, and Ms. J can be selected}}{\text{total number of 3 math teacher committees that can be formed}}.$$
>
> **The total number of possible committees is the number of ways that three math teachers can be chosen from 10 math teachers, which is the number of combinations of 10 math teachers taken 3 at a time or ${}_{10}C_3 = 120$. Mr. H, Ms. B, and Ms. J form one of these selections. The probability that the committee will consist of Mr. H, Ms. B, and Ms. J is $\frac{1}{120}$.**

- How did combinations or permutations help us to answer the question?
  - *The probability of this event is based on equally likely outcomes. We needed to determine how many ways three teachers could be selected from ten teachers. Since order did not matter, we could use a combination to help us determine the number of ways.*

Adapted from source material by Alberto Dominguez

Exercises 7–9

Exercises 7–9

7. **A high school is planning to put on the musical *West Side Story*. There are 20 singers auditioning for the musical. The director is looking for two singers who could sing a good duet.**

   a. **What is the probability that Alicia and Juan are the two singers who are selected by the director?**

$$\frac{1}{{}_{20}C_2} \approx 0.005$$

   This question involves a combination because the order of the two students selected does not matter. The probability of one of the selections (Alicia and Juan) would be 1 divided by the combination.

   b. **The director is also interested in the number of ways to choose a lead singer and a backup singer. What is the probability that Alicia is selected the lead singer and Juan is selected the backup singer?**

$$\frac{1}{{}_{20}P_2} \approx 0.0026$$

   This question involves a permutation because the order of the two singers matters. The probability of one of these selections (Alicia as the lead singer and Juan as the backup) would be 1 divided by the permutation.

8. **For many computer tablets, the owner can set a 4-digit pass code to lock the device.**

   a. **How many different 4-digit pass codes are possible if the digits cannot be repeated?**

$${}_{10}P_4 = 5040$$

   Use a permutation because order matters.

   b. **If the digits of a pass code are chosen at random and without replacement from the digits 0, 1, …, 9, what is the probability that the pass code is 1234?**

$$\frac{1}{5040} \approx 1.98 \cdot 10^{-4} \approx 0.000\,198$$

   The pass code 1234 is 1 out of the total number of possible pass codes. Therefore, the probability would be 1 divided by the permutation representing the total number of pass codes.

   c. **What is the probability that two people, who both chose a pass code by selecting digits at random and without replacement, both have a pass code of 1234?**

$$\frac{1}{5040} \cdot \frac{1}{5040} \approx 3.9 \cdot 10^{-8} \approx 0.000\,000\,039$$

   I multiplied the probability of the first person getting this pass code by the probability of a second person getting this pass code.

9. **A chili recipe calls for ground beef, beans, green pepper, onion, chili powder, crushed tomatoes, salt, and pepper. You have lost the directions about the order in which to add the ingredients, so you decide to add them in a random order.**

   a. **How many different ways are there to add the ingredients?**

$${}_8P_8 \text{ or } 8! = 40320$$

   This problem indicates that the order of adding the ingredients is important. The total number of ways of adding the eight ingredients in which order is important is the permutation indicated.

Adapted from source material by Alberto Dominguez

b. **What is the probability that the first ingredient that you add is crushed tomatoes?**

$$\frac{1}{8} = 0.125$$

There are eight ingredients to pick for my first pick. The probability of selecting crushed tomatoes would be the probability of selecting one of the eight ingredients, or $\frac{1}{8}$.

c. **What is the probability that the ingredients are added in the exact order listed above?**

$$\frac{1}{40320} \approx 2.5 \cdot 10^{-5} \approx 0.000\,025$$

The exact order of adding the ingredients represents 1 of the total number of permutations of the eight ingredients. The probability of selecting this 1 selection would be 1 divided by the total number of permutations.

Adapted from source material by Alberto Dominguez

Example 2: Probability and Combinations

Example 2: Probability and Combinations

A math class consists of 14 girls and 15 boys. The teacher likes to have the students come to the board to demonstrate how to solve some of the math problems. During a lesson, the teacher randomly selects 6 of the students to show their work. What is the probability that all 6 of the students selected are girls?

$$P(\text{all 6 students are girls}) = \frac{\text{number of ways to select 6 girls out of 14}}{\text{number of groups of 6 from the whole class}}$$

The number of ways to select 6 girls from the 14 girls is the number of combinations of 6 from 14, which is ${}_{14}C_6 = 3003$. The total number of groups of 6 is ${}_{29}C_6 = 475\,020$.

The probability that all 6 students are girls is

$$P(\text{all 6 students are girls}) = \frac{{}_{14}C_6}{{}_{29}C_6} = \frac{3003}{475\,020} \approx 0.006$$

Exercises 10–11

Exercises 10–11

10. There are nine golf balls numbered from 1 to 9 in a bag. Three balls are randomly selected without replacement to form a 3-digit number.

    a. How many 3-digit numbers can be formed?

    $${}_9P_3 = 504$$

    Order is important. As a result, I formed the permutation of selecting three golf balls from the nine golf balls.

    b. How many 3-digit numbers start with the digit 1?

    $$1 \cdot {}_8P_2 = 1 \cdot 8 \cdot 7 = 56$$

    If you select the digit 1 first, then there are 8 digits left for the other two digits. The number of ways of picking two of the remaining eight digits would be how many 3-digit numbers are formed with the digit 1 in the first position.

    c. What is the probability that the 3-digit number formed is less than 200?

    $$\frac{{}_8P_2}{{}_9P_3} \approx 0.111$$

    The probability of a 3-digit number formed that is less than 200 would be the probability formed by the number of 3-digit numbers that start with 1 (part (b)) divided by the total number of 3-digit numbers (part (a)).

Adapted from source material by Alberto Dominguez

11. **There are eleven seniors and five juniors who are sprinters on the high school track team. The coach must select four sprinters to run the 800-meter relay race.**

    a. **How many 4-sprinter relay teams can be formed from the group of 16 sprinters?**

    $${}_{16}C_4 = 1820$$

    b. **In how many ways can two seniors be chosen to be part of the relay team?**

    $${}_{11}C_2 = 55$$

    c. **In how many ways can two juniors be chosen to be part of the relay team?**

    $${}_{5}C_2 = 10$$

    d. **In how many ways can two seniors and two juniors be chosen to be part of the relay team?**

    $${}_{11}C_2 \cdot {}_{5}C_2 = 550$$

    e. **What is the probability that two seniors and two juniors will be chosen for the relay team?**

    $$\frac{{}_{11}C_2 \cdot {}_{5}C_2}{{}_{16}C_4} = \frac{550}{1820} \approx 0.302$$

Closing

Permutations and combinations can help us determine how many outcomes there are for an event, which can be used to calculate the probability.

Summary

- **The number of permutations of n things taken k at a time is**

$${}_nP_k = \frac{n!}{(n-k)!}.$$

- **The number of combinations of k items selected from a set of n distinct items is**

$${}_nC_k = \frac{{}_nP_k}{k!} \text{ or } {}_nC_k = \frac{n!}{k!(n-k)!}.$$

- **Permutations and combinations can be used to calculate probabilities.**

Adapted from source material by Alberto Dominguez

Name ____ Date ____________________

Using Permutations and Combinations to Compute Probabilities

Exit Ticket

1. An ice cream shop has 25 different flavors of ice cream. For each of the following, indicate whether it is a problem that involves permutations, combinations, or neither.

   a. What is the number of different 3-scoop ice cream cones that are possible if all three scoops are different flavors, and a cone with vanilla, strawberry, and chocolate is different from a cone with vanilla, chocolate, and strawberry?

   b. What is the number of different 3-scoop ice cream cones that are possible if all three scoops are different flavors, and a cone with vanilla, strawberry, and chocolate is considered the same as a cone with vanilla, chocolate, and strawberry?

   c. What is the number of different ice cream cones if all three scoops could be the same, and the order of the flavors is important?

2. A train consists of an engine at the front, a caboose at the rear, and 27 boxcars that are numbered from 1 to 27.

   a. How many different orders are there for cars that make up the train?

   b. If the cars are attached to the train in a random order, what is the probability that the boxcars are in numerical order from 1 to 27?

Adapted from source material by Alberto Dominguez

3. The dance club at school has 22 members. The dance coach wants to send four members to a special training on new dance routines.

   a. The dance coach will select four dancers to go to the special training. Is the number of ways to select four dancers a permutation, a combination, or neither?

   b. If the dance coach chooses at random, how would you determine the probability of selecting dancers Laura, Matthew, Lakiesha, and Santos?

Adapted from source material by Alberto Dominguez

Exit Ticket Sample Solutions

Note: The Exit Ticket does not contain problems similar to Example 2 or Exercises 10–11.

1. **An ice cream shop has 25 different flavors of ice cream. For each of the following, indicate whether it is a problem that involves permutations, combinations, or neither.**

   a. **What is the number of different 3-scoop ice cream cones that are possible if all three scoops are different flavors, and a cone with vanilla, strawberry, and chocolate is different from a cone with vanilla, chocolate, and strawberry?**

      Permutation

   b. **What is the number of different 3-scoop ice cream cones that are possible if all three scoops are different flavors, and a cone with vanilla, strawberry, and chocolate is considered the same as a cone with vanilla, chocolate, and strawberry?**

      Combination

   c. **What is the number of different ice cream cones if all three scoops could be the same, and the order of the flavors is important?**

      Neither; $25^3 = 15625$

2. **A train consists of an engine at the front, a caboose at the rear, and 27 boxcars that are numbered from 1 to 27.**

   a. **How many different orders are there for cars that make up the train?**

      $1 \cdot 27! \cdot 1 \approx 1.1 \cdot 10^{28}$

   b. **If the cars are attached to the train in a random order, what is the probability that the boxcars are in numerical order from 1 to 27?**

      $\frac{1}{27!} \approx 9.2 \cdot 10^{-29}$

3. **The dance club at school has 22 members. The dance coach wants to send four members to a special training on new dance routines.**

   a. **The dance coach will select four dancers to go to the special training. Is the number of ways to select four dancers a permutation, a combination, or neither?**

      The order of the dancers is not important; therefore, the number of ways of selecting four dancers would be a combination of four dancers from 22 possible dancers.

   b. **If the dance coach chooses at random, how would you determine the probability of selecting dancers Laura, Matthew, Lakiesha, and Santos?**

      The probability of selecting one of the combinations would be 1 divided by the total number of combinations.

Adapted from source material by Alberto Dominguez

Problem Set Sample Solutions

1. **For each of the following, indicate whether it is a question that involves permutations, combinations, or neither, and then answer the question posed.**
   a. **How many ways can a coach choose two co-captains from 16 players in the basketball team?**

      Combination; order does not matter; ${}_{16}C_2 = 120$

   b. **In how many ways can seven questions out of ten be chosen on an examination?**

      Combination; the order of the questions does not matter; ${}_{10}C_7 = 120$

   c. **Find the number of ways that 10 women in the finals of the skateboard street competition can finish first, second, and third in the X Games final.**

      Permutation; the order of the women is important in determining first, second, and third place; ${}_{10}P_3 = 720$

   d. **A postal zip code contains five digits. How many different zip codes can be made with the digits 0–9? Assume a digit can be repeated.**

      Neither; the digits can repeat; $10^5 = 100\,000$

2. **Four pieces of candy are drawn at random from a bag containing five orange pieces and seven brown pieces.**
   a. **How many different ways can four pieces be selected from the 12 colored pieces?**

      ${}_{12}C_4 = 495$

   b. **How many different ways can two orange pieces be selected from five orange pieces?**

      ${}_{5}C_2 = 10$

   c. **How many different ways can two brown pieces be selected from seven brown pieces?**

      ${}_{7}C_2 = 21$

3. **Consider the following:**
   a. **A game was advertised as having a probability of 0.4 of winning. You know that the game involved five cards with a different digit on each card. Describe a possible game involving the cards that would have a probability of 0.4 of winning.**

      Sample answer: Have the numbers $1, 2, 3, 4, 5$ on the cards. A card is selected at random. If the number is even, you win the game.

Adapted from source material by Alberto Dominguez

b. A second game involving the same five cards was advertised as having a winning probability of 0.05. Describe a possible game that would have a probability of 0.05 or close to 0.05 of winning.

Sample answer: Given that this probability is considerably less than the probability in Exercise 3, part (a), students would be expected to consider games in which more than one card is randomly selected, and the numbers formed from the selections would be involved in winning the game. If they randomly select two cards from the five possible cards, the probability of picking one of the possible 2-digit numbers without replacement is $\frac{1}{{}_5P_2}$, or 0.05. Encourage students to experiment with their suggestions by approximating the winning probabilities.

4. You have five people who are your friends on a certain social network. You are related to two of the people, but you do not recall who of the five people are your relatives. You are going to invite two of the five people to a special meeting. If you randomly select two of the five people to invite, explain how you would derive the probability of inviting your relatives to this meeting.

The number of ways of picking two people does not involve order. There is only one way I could pick the two relatives, and that would be if my pick involved both relatives. I would divide 1 by the total number of ways I could pick two people from five people, or the combination of picking two out of five.

5. Charlotte is picking out her class ring. She can select from a ruby, an emerald, or an opal stone, and she can also select silver or gold for the metal.

a. How many different combinations of one stone and one type of metal can she choose?

6; I multiplied the number of different stones by the number of different metals.

b. If Charlotte selects a stone and a metal at random, what is the probability that she would select a ring with a ruby stone and gold metal?

$$\frac{1}{6}$$

6. In a lottery, three numbers are chosen from 0 to 9. You win if the three numbers you pick match the three numbers selected by the lottery machine.

a. What is the probability of winning this lottery if the numbers cannot be repeated?

$$\frac{1}{{}_{10}C_3} \approx 0.008$$

b. What is the probability of winning this lottery if the numbers can be repeated?

$$\frac{1}{10^3} = 0.001$$

c. What is the probability of winning this lottery if you must match the exact order that the lottery machine picked the numbers?

$$\frac{1}{{}_{10}P_3} \approx 0.0014$$

Adapted from source material by Alberto Dominguez

7. The store at your school wants to stock T-shirts that come in five sizes (small, medium, large, XL, XXL) and in two colors (orange and black).

   a. How many different type T-shirts will the store have to stock?

      10

   b. At the next basketball game, the cheerleaders plan to have a T-shirt toss. If they have one T-shirt of each type in a box and select a shirt at random, what is the probability that the first randomly selected T-shirt is a large orange T-shirt?

      $\frac{1}{10}$

8. There are 10 balls in a bag numbered from 1 to 10. Three balls are selected at random without replacement.

   a. How many different ways are there of selecting the three balls?

      ${}_{10}C_3 = 120$

   b. What is the probability that one of the balls selected is the number 5?

      $\frac{{}_{9}C_2}{{}_{10}C_3} = 0.3$

9. There are nine slips of paper numbered from 1 to 9 in a bag. Four slips are randomly selected without replacement to form a 4-digit number.

   a. How many 4-digit numbers can be formed?

      ${}_{9}P_4 = 3024$

   b. How many 4-digit numbers start with the digit 1?

      ${}_{8}P_3 = 336$

10. There are fourteen juniors and twenty-three seniors in the Service Club. The club is to send four representatives to the state conference.

   a. How many different ways are there to select a group of four students to attend the conference from the 37 Service Club members?

      ${}_{37}C_4 = 66045$

   b. How many ways are there to select exactly two juniors?

      ${}_{14}C_2 = 91$

   c. How many ways are there to select exactly two seniors?

      ${}_{23}C_2 = 253$

   d. If the members of the club decide to send two juniors and two seniors, how many different groupings are possible?

      ${}_{14}C_2 \cdot {}_{23}C_2 = 23023$

Adapted from source material by Alberto Dominguez

e. **What is the probability that two juniors and two seniors are selected to attend the conference?**

$$\frac{{}_{14}C_2 \cdot {}_{23}C_2}{{}_{37}C_4} = \frac{23023}{66045} \approx 0.349$$

11. **A basketball team of 16 players consists of 6 guards, 7 forwards, and 3 centers. The coach decides to randomly select 5 players to start the game. What is the probability of 2 guards, 2 forwards, and 1 center starting the game?**

$$\frac{{}_{6}C_2 \cdot {}_{7}C_2 \cdot {}_{3}C_1}{{}_{16}C_5} \approx 0.216$$

12. **A research study was conducted to estimate the number of white perch (a type of fish) in a Midwestern lake. 300 perch were captured and tagged. After they were tagged, the perch were released back into the lake. A scientist involved in the research estimates there are $1,000$ perch in this lake. Several days after tagging and releasing the fish, the scientist caught 50 perch of which 20 were tagged. If this scientist's estimate about the number of fish in the lake is correct, do you think it was likely to get 20 perch out of 50 with a tag?**

Assume the total number of fish was $1,000$ perch. The probability of getting 20 of 50 perch tagged would be based on the number of ways to get 20 perch from the 300 tagged fish, multiplied by the number of ways of getting 30 perch from the 700 fish that are not tagged, divided by the number of ways of picking 50 fish from $1,000$ fish. This result is the requested probability, or

$$\frac{({}_{300}C_{20})({}_{700}C_{30})}{{}_{1000}C_{50}}$$

The above probability is approximately 0.04. This is a small probability and not likely to occur.

LESSON 3 – INTRODUCTION TO STATISTICS

DEFINITIONS

Statistics is a theory of information, with inference making as its objective. The large body of data that is the target of our interest is the **population**. The subset selected from the population to be subjected to detailed study is the **sample**.

The *mean* of a sample of n measured responses $y_1, y_2, \ldots, y_n$ is given by

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

The corresponding population mean is denoted μ.

We usually cannot measure the value of the population mean μ; rather, μ is an unknown constant that we may want to estimate using sample information. The mean of a set of measurements only locates the center of the distribution of data; by itself, it does not provide an adequate description of a set of measurements. Two sets of measurements could have widely different frequency distributions but equal means. To describe data adequately, we must also define measures of data variability. The most common measure of variability used in statistics is the variance, which is a function of the deviations of the sample measurements from their mean.

There is a standard MATLAB function **mean** to calculate the mean of a set of measurements.

https://www.mathworks.com/help/matlab/ref/mean.html

The *variance* of a sample of measurements $y_1, y_2, \ldots, y_n$ is the sum of the square of the differences between the measurements and their mean, divided by $n - 1$. Symbolically, the sample variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 .$$

The corresponding population variance is denoted by the symbol σ^2.

Notice that we divided by n−1 instead of by n in our definition. There is a theoretical reason for this choice of divisor (which we may or may not explore later in the course).

As with mean, we usually cannot measure the value of the population variance; rather, it is an unknown estimated using sample information.

There are standard MATLAB functions **var** and **std** to calculate variance and standard deviation, respectively.

https://www.mathworks.com/help/matlab/ref/var.html

https://www.mathworks.com/help/matlab/ref/std.html

ALTERNATIVE (SHORTCUT) FORMULA FOR STANDARD DEVIATION

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]$$

I only included the formula above because it comes from the same book (and therefore uses the same notation) as all the previous formulas. However, this is a very ugly and unclear way to present this idea. Here is a better way to present the same formula.

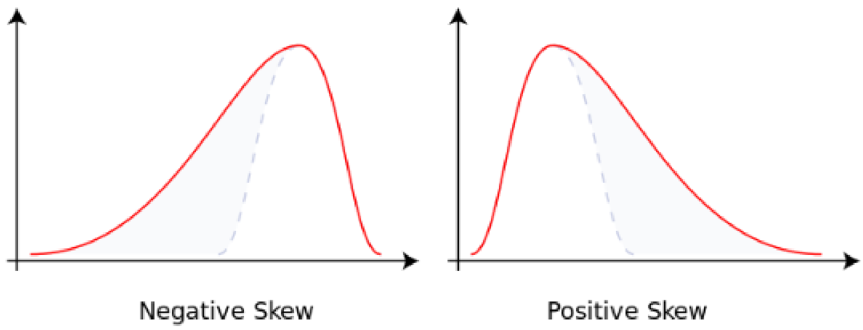
$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{n(\Sigma X^2) - (\Sigma X)^2}{n(n-1)}$$

The derivation is included in Modules on Canvas, but you don't really need to understand the derivation if you can apply the formula.

The idea behind this shortcut is that instead of having to calculate the mean first and then calculate the deviation of each value y_i from this mean (which requires going through the data set twice), we can simply calculate the average of the values squared y_i^2 and subtract the square of the average value, which can be accomplished in a single loop through the data set. This is usually much simpler to program and work with computationally.

SKEWNESS

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or undefined. For a unimodal distribution (i.e., a distribution that exhibits a single peak), negative skew commonly indicates that the tail is on the left side of the distribution, and positive skew indicates that the tail is on the right.



Many textbooks teach a rule of thumb stating that the mean is right of the median under right skew and left of the median under left skew. This rule fails with surprising frequency. It can fail in multimodal distributions or in distributions where one tail is long but the other is heavy. Most commonly, though, the rule fails in discrete distributions where the areas to the left and right of the median are not equal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

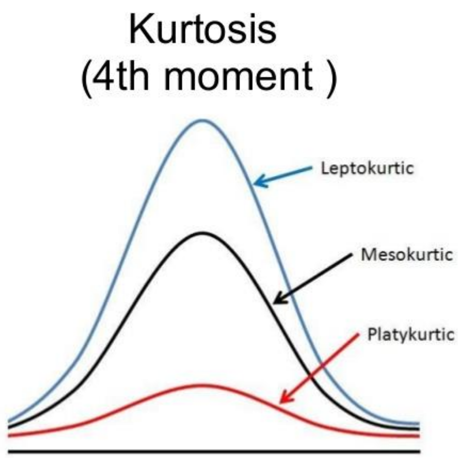
https://www.mathworks.com/help/stats/skewness.html

KURTOSIS

Kurtosis is a measure of the combined sizes of the two tails. It measures the amount of probability in the tails. The value is often compared to the kurtosis of the normal distribution, which is equal to 3. Many books re-define kurtosis so that the kurtosis of the normal distribution comes out to 0.

Excess kurtosis means the distribution of event outcomes have lots of instances of outlier results, causing "fat tails" on the bell-shaped distribution curve. This means the event in question is prone to extreme outcomes.

Mesokurtic distributions have a kurtosis equal to that of the normal distribution. In contrast, a **leptokurtic** distribution has kurtosis greater than that of the normal distribution, meaning it has fatter tails and the probability of extreme events is greater than that implied by the normal curve. A **platykurtic** distribution has kurtosis less than that of the normal distribution, meaning it has thinner tails and the probability of extreme events is less than that implied by the normal curve.



https://classconnection.s3.amazonaws.com/65/flashcards/2185065/jpg/kurtosis-142C1127AF2178FB244.jpg

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/kurtosis.html

HYPOTHESIS TESTING

Many times, the objective of a statistical test is to test a hypothesis concerning the values of one or more population parameters. We generally have a theory—a research hypothesis—about the parameter(s) that we wish to support. For example, suppose that a political candidate, Jones, claims that he will gain more than 50% of the votes in a city election and thereby emerge as the winner. If we do not believe Jones's claim, we might seek to support the research hypothesis that Jones is not favored by more than 50% of the electorate. Support for this research hypothesis, also called the alternative hypothesis, is obtained by showing (using the sample data as evidence) that the converse of the alternative hypothesis, called the null hypothesis, is false.

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps.

1. Formulate the null hypothesis H_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H_a (commonly, that the observations show a real effect combined with a component of chance variation).

2. Identify a test statistic that can be used to assess the truth of the null hypothesis.

3. Compute the p-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the p-value, the stronger the evidence against the null hypothesis.

4. Compare the p-value to an acceptable significance value α. If $p \leq \alpha$, the observed effect is statistically significant, the null hypothesis is rejected, and the alternative hypothesis is accepted. If $p > \alpha$, the observed effect is not statistically significant, and the null hypothesis is not rejected.

The Elements of a Statistical Test

1. Null hypothesis, H_0
2. Alternative hypothesis, H_a
3. Test statistic
4. Rejection region

The functioning parts of a statistical test are the test statistic and an associated rejection region (RR). The test statistic is a function of the sample measurements on which the statistical decision will be based. RR specifies the values of the test statistic for which the null hypothesis is to be rejected in favor of the alternative hypothesis. If for a sample, the computed value of the test statistic falls in the RR, we reject the null hypothesis H_0 and accept the alternative hypothesis H_a. If the value of the test statistic does not fall into the RR, we ~~accept~~ **fail to reject** H_0. Finding a good RR for a statistical test is a challenging problem that we do not have time to explore.

Definition. A type I error is made if H_0 is rejected when H_0 is the correct hypothesis. The probability of a type I error is denoted by α. A type II error is made if H_0 is not rejected but H_a is the correct hypothesis. The probability of a type II error is denoted by β.

"REJECT" VS "FAILURE TO REJECT"

The p-value is probabilistic.

This means that when we interpret the result of a statistical test, we do not know what is true or false, only what is likely.

Rejecting the null hypothesis means that there is enough statistical evidence that the null hypothesis does not look likely. Otherwise, it means that there is not enough statistical evidence to reject the null hypothesis.

We may think about the statistical test in terms of the dichotomy of rejecting and accepting the null hypothesis. The danger is that if we say that we "*accept*" the null hypothesis, the language suggests that the null hypothesis is true. Instead, it is safer to say that we "*fail to reject*" the null hypothesis, as in, there is insufficient statistical evidence to reject it.

When reading "*reject*" vs "*fail to reject*" for the first time, it is confusing to beginners. You can think of it as "*reject*" vs "*accept*" in your mind, as long as you remind yourself that the result is probabilistic and that even an "*accepted*" null hypothesis still has a small probability of being wrong.

COMMON P-VALUE MISINTERPRETATIONS

True or False Null Hypothesis

The interpretation of the p-value does not mean that the null hypothesis is true or false.

It means that we have chosen to reject or not reject the null hypothesis at a specific statistical significance level based on empirical evidence and the chosen statistical test.

You are limited to making probabilistic claims, not crisp binary or true/false claims about the result.

p-value as Probability

A common misunderstanding is that the p-value is a probability of the null hypothesis being true or false given the data. This is incorrect.

Instead, the p-value can be thought of as the probability of the data given the pre-specified assumption embedded in the statistical test.

It allows us to reason about whether the data fits the hypothesis. Not the other way around.

The p-value is a measure of how likely it is that the data sample would be observed if the null hypothesis were true.

Post-Hoc Tuning

It does not mean that you can re-sample your domain or tune your data sample and re-run the statistical test until you achieve a desired result.

It also does not mean that you can choose your p-value after you run the test. This is called p-hacking or hill climbing and will mean that the result you present will be fragile and not representative. In science, this is at best inaccurate, and at worst unethical.

EXERCISES

Wackerly, Chapter 1 Exercises 23, 27-29, 31 (these exercises are all based on the z values reading)

REFERENCES

https://en.wikipedia.org/wiki/Misuse_of_p-values

https://machinelearningmastery.com/statistical-hypothesis-tests/

Wackerly et al, *Mathematical Statistics with Applications 7th Edition*, Chapters 1 and 10

Weisstein, Eric W. "Hypothesis Testing." From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/HypothesisTesting.html

LESSON 4 – ASSORTED HYPOTHESIS TESTS

TEACHER'S NOTE: You do not need to master all the details of every test in this lesson. Just read for high-level understanding of the ideas. The point of including all the tests was to create a single document that you could use as a reference when you need to perform hypothesis testing.

NORMALITY

A large fraction of the field of statistics is concerned with data that assumes that it was drawn from a normal (aka Gaussian) distribution.

If methods are used that assume a normal distribution, and your data was drawn from a different distribution, the findings may be misleading or plain wrong.

There are several techniques that you can check if your data sample is normal or sufficiently normal to use the standard techniques, or sufficiently non-normal to instead use non-parametric statistical methods.

This is a key decision point when it comes to choosing statistical methods for your data sample. We can summarize this decision as follows:

```
If Data Is Normal
    Use Parametric Statistical Methods
Else
    Use Nonparametric Statistical Methods
```

There is also some middle ground where we can assume that the data is normal enough to use parametric methods or that we can use data preparation techniques to transform the data to be sufficiently normal to use the parametric methods.

VISUAL NORMALITY CHECKS

We can create plots of the data to check whether it is normal.

These checks are qualitative, so less accurate than the statistical methods we will calculate in the next section. Nevertheless, they are fast and like the statistical tests, must still be interpreted before you can make a call about your data sample.

Histogram Plot

A simple and commonly used plot to check the distribution of a sample of data quickly is the histogram.

A data sample from a population with a normal distribution will display the familiar bell shape.

Quantile-Quantile Plot

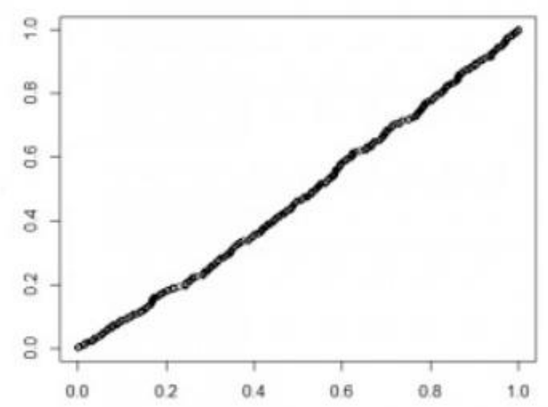
Another popular plot for checking the distribution of a data sample is the quantile-quantile plot, Q-Q plot, or QQ plot for short.

This plot generates its own sample of the idealized distribution that we are comparing with, in this case the normal distribution. The idealized samples are divided into groups called quantiles. Each data point in the sample is paired with a similar member from the idealized distribution at the same cumulative distribution.

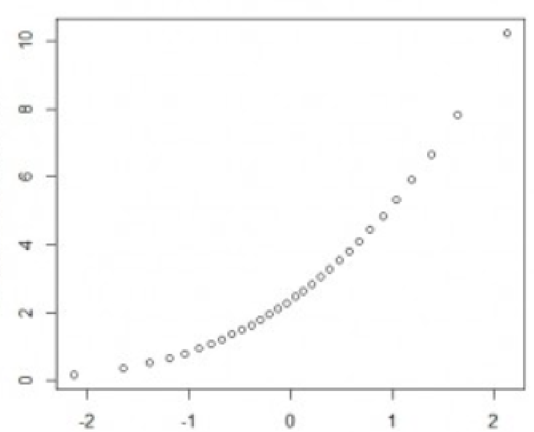
The resulting points are plotted as a scatter plot with the idealized value on the x-axis and the data sample on the y-axis.

A perfect match for the distribution will be shown by a line of dots on a 45-degree angle from the bottom left of the plot to the top right. Often a line is drawn on the plot to help make this expectation clear. Deviations by the dots from the line shows a deviation from the expected distribution.

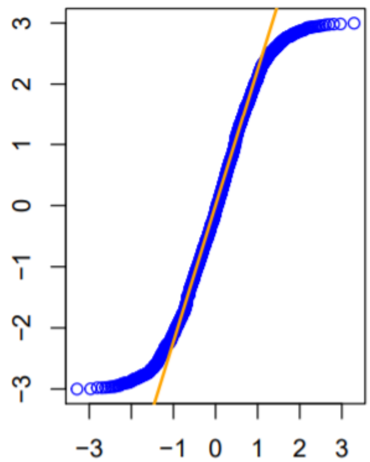
https://en.wikipedia.org/wiki/Q–Q_plot



The figure above is for data that is approximately normal.



The figure above is for data that is definitely not normal.



In the graph above, notice the points fall along a line in the middle of the graph, but curve off in the extremities. QQ plots that exhibit this distinctive S-shaped behavior usually mean your data have more extreme values than would be expected if they truly came from a normal distribution. Many data sets in the real world exhibit this kind of behavior.

EXTREME EVENTS IN FINANCE

One data set that exhibits this behavior is stock market returns. In finance, extremal events happen more often than one might expect, particularly if one incorrectly assumes that stock market returns are normally distributed.

- The 1987 stock market crash. On Monday, October 19, 1987, the Dow Jones Industrial Average dropped 22.6%, an event approximately 25 standard deviations away from the mean. If stock market returns were normally distributed, this would be essentially impossible.
- The 1990 Japanese real estate and stock market collapse. On 1/1/1990, the Nikkei 225 stock market index stood at 38916, near its all-time high. As of 8/23/2019, nearly three full decades later, it stood at only 20711, almost 50% lower.
- The 1997 "Asian Tiger" slowdown and 1998 Russian default resulting in the infamous collapse of the Long-Term Capital Management (LTCM) hedge fund collapse.
- The 2000 collapse of the technology and communications bubble. People focus on the dot-com collapse, but the stock market value lost in telecoms was many multiples larger than the value lost in dot-coms. Lucent stock went from an all-time high of $69 to 55 cents. WorldCom had a peak market value of $70 billion, but in just two years it collapsed in the then-largest bankruptcy in history.
- The collapse of the 2007 real estate bubble and the 2008-9 stock market crash.

COMPARISON OF ACTUAL S&P500 QUARTERLY RETURNS VS EXPECTED RETURNS BASED ON ASSUMPTION OF NORMALITY

No. of Quarterly Occurrences

Severe Quarterly Losses	**Actual** S&P 500 Returns	**Expected** S&P 500 Returns	**Unexpected Tail Risk Events** *
(20%)	42	17	**+25**
(22%)	27	8	**+19**
(24%)	28	4	**+24**
(26%)	29	2	**+27**
(28%)	20	1	**+19**
(30%)	10		**+10**
(32%)	7		**+7**
(34%)	2		**+2**
(36%)	3		**+3**
(38%)			
(40%)	1		**+1**
Total	169	32	**+137**

Actual frequency of severe losses occurred **5.3x** more frequently than expected.

EXTREME EVENTS IN INSURANCE

Property-casualty insurance data also indicates that extreme events happen more often than would be expected if losses were normally distributed.

- In 2017, the US exhibited three "once in a decade" hurricanes (Harvey, Irma, and Maria), each causing in excess of $30 billion in losses.
- In 2004, Florida was directly impacted by four major hurricanes – Charley, Frances, Ivan, and Jeanne. This was followed by Dennis, Katrina and Wilma in 2005. Based on insurance company loss frequency models, this should have been essentially impossible. The Florida home insurance market has still not recovered from these losses.

STATISTICAL NORMALITY TESTS

Shapiro-Wilk Test

Tests whether a data sample has a normal distribution.

Assumptions

- Observations in each sample are independent and identically distributed (iid).

Interpretation

- H_0: the sample has a normal distribution.
- H_a: the sample does not have a normal distribution.

Swtest.m

Shapiro-Wilk test on Wikipedia

D'Agostino's K^2 Test

Tests whether a data sample has a normal distribution.

Assumptions

- Observations in each sample are independent and identically distributed (iid).

Interpretation

- H_0: the sample has a normal distribution.
- H_a: the sample does not have a normal distribution.

DP.m

D'Agostino's K^2 test on Wikipedia

Anderson-Darling Test

Tests whether a data sample has a normal distribution.

Assumptions

- Observations in each sample are independent and identically distributed (iid).

Interpretation

- H_0: the sample has a normal distribution.
- H_a: the sample does not have a normal distribution.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/adtest.html

Anderson-Darling test on Wikipedia

WHAT TEST SHOULD YOU USE?

We have covered a few normality tests, although not all of the tests that exist of course. But which test do you use? I recommend using them all on your data, where appropriate. The question then becomes, how do you interpret the results? What if the tests disagree, which they often will? I have two suggestions for you to help think about this question.

Hard Fail

Your data may not be normal for lots of different reasons. Each test looks at the question of whether a sample was drawn from a normal distribution from a slightly different perspective. A failure of one normality test means that your data is not normal. As simple as that.

You can either investigate why your data is not normal and perhaps use data preparation techniques to make the data more normal or you can start looking into the use of nonparametric statistical methods instead of the parametric methods.

Soft Fail

If some of the methods suggest that the sample is normal and some not, then perhaps take this as an indication that your data is approximately but not exactly normal. In many situations, you can treat your data as though it is normal and proceed with your chosen parametric statistical methods.

CORRELATION TESTS

Pearson's Correlation Coefficient

Tests whether two samples have a linear relationship. Pearson's linear correlation coefficient is the most commonly used linear correlation coefficient. Values of the correlation coefficient can range from –1 to +1. A value of –1 indicates perfect negative correlation, while a value of +1 indicates perfect positive correlation. A value of 0 indicates no correlation.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H_0: the two samples are independent.
- H_a: there is a dependency between the samples.

This is a built-in function in MATLAB.

https://www.mathworks.com/help/matlab/ref/corrcoef.html

Pearson's correlation coefficient on Wikipedia

Spearman's Rank Correlation and Kendall's Rank Correlation

Test whether two samples have a monotonic relationship.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0: the two samples are independent.
- H_a: there is a dependency between the samples.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/corr.html

Spearman's rank correlation coefficient on Wikipedia

Kendall rank correlation coefficient on Wikipedia

Chi-Squared Test

Tests whether two categorical variables are related or independent.

Assumptions

- Observations used in the calculation of the contingency table are independent.
- 25 or more examples in each cell of the contingency table.

Interpretation

- H_0: the two samples are independent.
- H_a: there is a dependency between the samples.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/chi2gof.html

Chi-Squared test on Wikipedia

PARAMETRIC STATISTICAL HYPOTHESIS TESTS

Student's t-test

Tests whether the means of two independent samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H_0: the means of the samples are equal.
- H_a: the means of the samples are unequal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/ttest2.html

Student's t-test on Wikipedia

Paired Student's t-test

Tests whether the means of two paired samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.
- Observations across each sample are paired.

Interpretation

- H_0: the means of the samples are equal.
- H_a: the means of the samples are unequal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/ttest.html

Analysis of Variance Test (ANOVA)

Tests whether the means of two or more independent samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H_0: the means of the samples are equal.
- H_a: one or more of the means of the samples are unequal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/anova1.html

Analysis of variance on Wikipedia

Repeated Measures ANOVA Test

Tests whether the means of two or more paired samples are significantly different.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.
- Observations across each sample are paired.

Interpretation

- H_0: the means of the samples are equal.
- H_a: one or more of the means of the samples are unequal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/repeatedmeasuresmodel.ranova.html

NONPARAMETRIC STATISTICAL HYPOTHESIS TESTS

Mann-Whitney U Test

Tests whether the distributions of two independent samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0: the distributions of both samples are equal.
- H_a: the distributions of both samples are not equal.

mannWhitney.m

[Mann-Whitney U test on Wikipedia]

Wilcoxon Signed-Rank Test

Tests whether the distributions of two paired samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.
- Observations across each sample are paired.

Interpretation

- H_0: the distributions of both samples are equal.
- H_a: the distributions of both samples are not equal.

https://www.mathworks.com/help/stats/signrank.html

Wilcoxon signed-rank test on Wikipedia

Kruskal-Wallis H Test

Tests whether the distributions of two or more independent samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- H_0: the distributions of all samples are equal.
- H_a: the distributions of one or more samples are not equal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/kruskalwallis.html

Kruskal-Wallis one-way analysis of variance on Wikipedia

Friedman Test

Tests whether the distributions of two or more paired samples are equal or not.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.
- Observations across each sample are paired.

Interpretation

- H_0: the distributions of all samples are equal.
- H_a: the distributions of one or more samples are not equal.

This is a built-in function in MATLAB's Statistics and Machine Learning Toolbox.

https://www.mathworks.com/help/stats/friedman.html

Friedman test on Wikipedia

REFERENCES

https://machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/

https://machinelearningmastery.com/statistical-hypothesis-tests/

https://machinelearningmastery.com/a-gentle-introduction-to-normality-tests-in-python/

Embrechts et al, *Modeling Extremal Events for Insurance and Finance*

Hardy, *Investment Guarantees*

LESSON 5 – RANDOM NUMBER GENERATORS AND INTRODUCTION TO MONTE CARLO METHODS

RANDOM NUMBER GENERATORS

We will not attempt to define the term "random." Physical processes that are usually considered random, such as flipping a coin or rolling a die, are actually deterministic if we know enough about the initial conditions under which they were thrown. The seemingly random behavior of so-called chaotic dynamical systems is actually due to extreme sensitivity to initial conditions. Weather provides a good example of such a system, as predictions become impossible more than a few days out due to our imperfect knowledge of the values of the variables underlying the system.

One useful way to think about it [Heath, Section 13.2] is to characterize a sequence as random if it has no shorter description than itself. That is, the terms in the sequence cannot be defined by a rule.

PSEUDO-RANDOM NUMBER GENERATORS

Computer algorithms for generating random numbers supply sequences of terms that seem to be random. However, the algorithm itself provides the description of the sequence, so it cannot truly be random. Such algorithms are generally referred to as pseudo-random number generators (PRNGs). PRNG algorithms generally require one or more initial values, called seeds.

PROPERTIES OF RANDOM NUMBER SEQUENCES AND PSEUDO-RANDOM NUMBER GENERATORS

1) Random sequences should have a long period, meaning that they supply a large number of terms before repeating. A common manifestation of this problem is PRNGs that have suitable long periods for some seed values but shorter than expected periods for other seed values (known as weak seeds). A minimum acceptable period is 2^{64} or about 2×10^{19}. Never use an algorithm with unknown period.
2) Random sequences should be reproducible, meaning we should be able to start the generator with specified initial conditions to obtain the same sequence of terms. This property is of the utmost importance in testing software or peer-reviewing scientific papers based on simulations.

3) Random sequences should be unpredictable. Even with full knowledge of every term in the sequence up to the current term, there is no way to predict the next term in the sequence. In particular, the pattern should pass statistical tests of randomness.
4) Random sequences should exhibit the appropriate distribution, covering all portions of the expected range appropriately. For example, a random sequence drawn from the interval [0, 1) should exhibit uniformity of distribution if a large number of terms is generated. As another example, if a random sequence intended to generate normally distributed variables with mean 0 and variance 1, then no interval on the real number should display an "unusually high" or "unusually low" number of terms if a large number of terms is generated. There are statistical pattern detection tests to check for appropriate behavior.
5) Random sequences should be uncorrelated, having no statistical relationship between terms.
6) Generators should be efficient, executing rapidly and requiring little memory, since most simulations require a large number of terms.
7) Generators should have a sound mathematical basis. In the early development of algorithms, it was thought that complexity in an algorithm would by itself ensure randomness in the results. This resulted in many algorithms that turned out to have highly non-random properties, e.g., John von Neumann's "middle square" algorithm.
8) Generators should support parallel computing. Parallel computing (i.e., running separate threads of your program through different processors) requires an instance of the generator for each thread. Each of these instances needs to be independent of the others. Some older algorithms that otherwise have good statistical properties are not usable for parallel computing.
9) Generators should be portable, able to run on different types of computer and produce the same sequences on each.

It is very difficult to satisfy all of these requirements simultaneously. The study of PRNG algorithms is an active area of research in the field of numerical analysis. We will have to develop an understanding of the relative strengths and weaknesses of the algorithms we use in our simulations, in order to be able to identify and possibly mitigate deficiencies in conclusions drawn from simulations. However, we do not have to become experts in the statistics of random numbers. We will not be developing any PRNG algorithms (although we might decide to implement algorithms developed by others).

A few general observations will suffice for our purposes:

- Java's PRNG (contained in Math.Random or java.util.random) has some flaws, specifically it has a relatively short period and it fails some randomness tests. It is suitable for some quick modeling or simple applications, but we should not rely it on for serious modeling such as official competitions.

- Despite some initial improvements made in Excel 2003 and further improvements in Excel 2010, the PRNG in Microsoft Excel still has serious quality issues. It should not be used in this class, with the possible exception of quick prototypes when trying to decide on an appropriate model.
- MATLAB and Python built-in PRNGs both use the high-quality Mersenne Twister (MT) algorithm. Therefore, it will not be necessary to code our own PRNG in either of those environments. The function in MATLAB for accessing MT is `rand`. MT is slower than many generators that do as well in tests and require only a few lines of code, but since it is built-in (i.e., requires NO lines of code for our purposes) we will use it.

TYPES OF PSEUDO-RANDOM NUMBER GENERATORS

Since we not be developing our own PRNG algorithms, this section will contain only the barest outline of some different types of algorithms.

- Congruential generators have the form $x_k = (ax_{k-1} + b) \pmod M$ where a, b, and M are selected to provide the generator with good statistical properties. Congruential PRNGs are seeded with an initial value x_0. These algorithms were once considered state of the art, but the present consensus among numerical analysts is to avoid congruential generators.
- Fibonacci generators calculate a new term as a difference, sum or product of previous terms.
- Bit-shift methods generate terms by operating on the individual bits in the numerical representation of previous terms.
- Most recently, highly efficient Multiply-with-Carry generators were developed starting in 1991 by George Marsaglia.

QUASI-RANDOM NUMBER GENERATORS

For some applications achieving reasonably uniform coverage of the sampled volume (avoiding the clumping that often occurs with random sequences) is more important than whether the sample points are truly random. For such applications, sequences of numbers called quasi-random or low-discrepancy sequences are developed. There are a number of algorithms for generating such sequences, but we will defer a detailed discussion of them until we identify a need for them in our work.

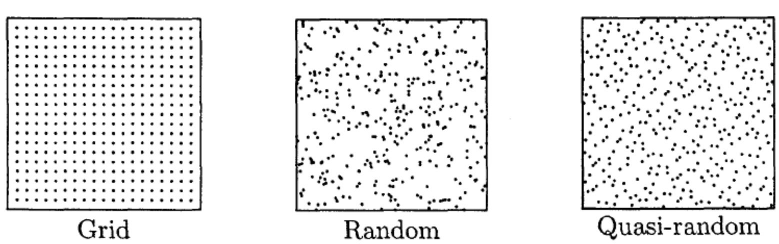


Figure 13.1: Three methods for sampling the unit square using the same number of points.

(Heath, *Scientific Computing 2nd Edition*, Figure 13.1, page 516)

EXAMPLE 1

(a) Write a code snippet that simulates the flipping of a fair coin.

```
if rand < 0.5
    disp("H")
else
    disp("T")
end
```

(b) Write a code snippet that simulates the tossing of a fair six-sided die.

```
disp(floor(rand*6 + 1))
```

WHAT IS A MONTE CARLO SIMULATION?

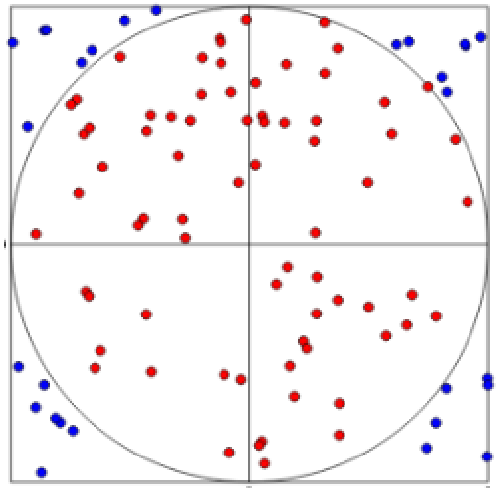
Monte Carlo simulations are used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. It is a technique used to understand the impact of risk and uncertainty in prediction and forecasting models.

WHAT IS A MONTE CARLO INTEGRATION?

Monte Carlo integration is a technique for numerical integration using random numbers. It is a particular Monte Carlo method that numerically computes a definite integral. While other algorithms usually evaluate the integrand at a regular grid, Monte Carlo methods randomly choose points at which the integrand is evaluated. This method is particularly useful for higher-dimensional integrals and is therefore highly useful in many financial engineering problems.

EXAMPLE 2

Write a simulation to compute the area of a circle of diameter 1 inscribed in the unit square.



```
N = 10000000;
r2 = (2*rand(1,N)-1).^2 + (2*rand(1,N)-1).^2;
disp(4*sum(r2<=1)/N)

>> Example11
    3.1418
```

Study this solution code carefully. There are two important features that we will use throughout this course.

1. Note that “native” MATLAB code uses arrays rather than loops for repetitive calculations. For a Java programmer, this is not the most intuitive way to code this function, but we should make a conscious effort to avoid loops because array calculations are hundreds of times more efficient.

2. The function `rand` produces a random number x between 0 and 1. By calculating 2x-1 we transform this into a random number between -1 and 1. See discussion question 1 for a generalization of this transformation.

DISCUSSION QUESTIONS

1) If you have a random number generator for a uniform distribution on (0, 1), how can you use it to generate random numbers uniformly distributed on (a, b)? [Heath, Question 13.9]

2) If you have a random number generator for a normal distribution with mean 0 and variance 1, how can you use it to generate random numbers normally distributed with mean μ and variance σ^2? [Heath, Question 13.10]

CODING EXERCISES

3) Write a simulation to compute the volume of a sphere of diameter 1 inscribed in the unit cube. [Heath, Exercise 13.10]

4) **Birthday Problem:** Write a simulation that determines the smallest number of persons required for the probability to be greater than 50% that two persons in a group have the same birthday. [Heath, Exercise 13.6]
Optional follow-up: Evaluate the validity of your simulation by deriving the result analytically.

5) **Random Walk Problem:** A random walk is a sequence of equal steps, each taken in an independent random direction. Consider the one-dimensional case in which the walk begins at the origin and proceeds by successive steps of unit length with the direction, positive or negative, chosen randomly with equal probability. Write a simulation to implement a random walk of n steps and determine the distance from the origin at the end of the walk. [Heath, Exercise 13.13]

6) **Monty Hall Problem: A** game show host leads you to a wall with three closed doors. Behind one of the doors is the automobile of your dreams, and behind each of the other two is a can of dog food. The three doors all have even chances of hiding the automobile. The host, a person who knows precisely what is behind each of the three doors, explains how the game will work. First, you will choose a door without opening it, knowing that after you have done so, the host will open one of the two remaining doors to reveal a can of dog food. (HINT: Think very carefully about what this means!) When this has been done, you will be given the opportunity to switch doors; you will win whatever is behind the door you choose at this stage of the game. Do you raise your chances of winning the automobile by switching doors? [Marko Boon, Problem 1.2]

7) **Saint Petersburg Paradox:** A fair coin is flipped until it comes up heads the first time. At that point the player wins $\$2^n$ where n is the number of times the coin was flipped. Show that the theoretical expected value of the payoff of this game is

infinite. However, in practice, the actual average payoff of this game turns out to be a small number. Write a simulation that estimates the average payoff of this game.

OPTIONAL ADDITIONAL EXERCISES

1) A sequence of random numbers distributed uniformly on [0, 1) should have mean $\mu = \int_0^1 x dx = \frac{1}{2}$ and variance $\sigma^2 = \int_0^1 (x-\mu)^2 dx = \frac{1}{12}$. For each of the PRNGs discussed in this lesson (Excel, Python, Java, MATLAB), generate series of 1000 and 10000 terms, and determine the mean and variance of each. [Heath, Exercise 13.4]

2) Write a simulation of a poker hand, based on a standard 52 card deck (4 suits – hearts, clubs, diamonds, spaces – each with 13 ranked cards – 2-10, J, Q, K, A) with no jokers and no wild cards. Use your simulation to determine the probability of each type of five-card poker hand – four of a kind, full house (i.e., 3 cards of one kind and 2 cards of another kind), flush (all five cards of the same suit), straight (five cards in a row, A can be either the low card or the high card), straight flush (a hand that is both a straight and a flush), three of a kind, two pairs, one pair, and busted hand (all hands that don't meet one of the previous definitions).
Optional follow-up: Evaluate the validity of your simulation by calculating the probabilities analytically.

3) You have taken ten different pairs of socks to the laundromat, and during the washing, six socks are lost. In the best-case scenario, you will still have seven matching pairs left. In the worst-case scenario, you will have four matching pairs left. (Make sure that you understand why these two statements are true. That will help you solve this exercise.) Write a simulation to estimate the probabilities that you have respectively 4, 5, 6, or 7 matching pairs left? [Marko Boon, Problem 1.1]

REFERENCES

M. Boon, Simulation Lecture Notes, Lecture 1, https://www.win.tue.nl/~marko/2WB05/

M. Heath, *Scientific Computing 2nd Edition*, Chapter 13

T. Kennedy, *Monte Carlo Methods*, Chapter 3 (https://www.math.arizona.edu/~tgk/mc/book.pdf)

W. Press et al, *Numerical Recipes 3rd Edition*, Section 7.1

S. Ross, *Introduction to Probability Models 10th Edition*, Chapter 11

S. Ross, *Simulation 4th Edition*, Chapter 3

http://www.andygranowitz.com/2015/01/12/r-code-simulation-st-petersburg-paradox.html

https://plato.stanford.edu/entries/paradox-stpetersburg/

https://en.wikipedia.org/wiki/Randomness_tests

https://en.wikipedia.org/wiki/St._Petersburg_paradox

LESSON 6 – GENERATING DISCRETE RANDOM VARIABLES

DISCRETE RANDOM VARIABLE WITH PROBABILITY MASS FUNCTION

Suppose we are given a discrete random variable X having probability mass function $P\{X = x_j\} = p_j$ with j = 0, 1, …, and $\Sigma p_j = 1$.

To generate a value for this variable, we first create a random number U that is uniformly distributed on (0,1), indicated as U(0,1), then we set X as follows:

- x_1 if $0 < U < p_0$
- x_2 if $p_0 < U < p_1$, etc.
- In general $X = x_j$ if $\sum_0^{j-1} p_i < U < \sum_0^{j} p_i$
- …

This can be written as an algorithm as follows:

1. Generate a random number U(0,1)
2. If $U < p_0$, then set $X = x_0$ and stop
3. If $U < p_0 + p_1$, then set $X = x_1$ and stop
4. If $U < p_0 + p_1 + p_2$, then set $X = x_2$ and stop
5. …

The amount of time it takes to generate a random variable by this method is proportional to the number of intervals searched. Therefore, it is worthwhile to consider the possible values of x_j in decreasing order of p_j.

Example: Simulate a random variable X such that P(1) = 0.20, P(2) = 0.15, P(3) = 0.25, and P(4) = 0.40. Generate U(0,1), and set X = 4 if U < 0.40, X = 3 if U < 0.65, X = 1 if U < 0.85, and X = 2 otherwise.

DISCRETE UNIFORM RANDOM VARIABLE

A random variable for which the outcomes m, m+1, …, n are all equally likely can be generated with the simple algorithm:

1. Generate a random number U(0,1); k = n – m + 1
2. Set X = m + kU

MATLAB has a built-in function **rand** to generate a random number U(0,1).
https://www.mathworks.com/help/matlab/ref/rand.html

RANDOM PERMUTATIONS

Suppose we are interested in generating a permutation of the integers 1, 2, …, n, such that all n! possible orderings are equally likely.

1. Let $P_1, P_2, \ldots, P_n$ be any permutation of 1, 2, …, n, e.g., take $P_j = j$.
2. Set k = n
3. Generate a random number U(0,1) and let I = int(kU) + 1
4. Interchange the values of P_I and P_k.
5. Decrement k and if k > 1 go to Step 3.
6. Let $P_1, P_2, \ldots, P_n$ is the desired random permutation.

One property of this algorithm is that it can be modified to generate a random subset of size r of the integers 1, 2, …, n. Just follow the algorithm until the positions n through n-r+1 are filled, and the elements in these positions constitute a random subset of size r. Note that if r > n/2, then for efficiency we would use this algorithm to choose a random subset of size n-r and use the elements not in the set as our random subset.

GEOMETRIC RANDOM VARIABLE

A geometric random variable $P\{X = i\} = p(1-p)^{i-1} = (1-q)q^{i-1}$ represents the time of the first success when independent trials each with success of probability p are executed. Note that the probability $1 - P\{X > j-1\} = 1 - q^{j-1}$. Therefore, we can generate the desired value of X by the following algorithm.

1. Generate a random number U(0,1).
2. Set X equal to the value of j for which $1 - q^{j-1} \le U < 1 - q^j$, or equivalently $q^j < 1 - U \le q^{j-1}$. This value is calculated using the formula X = Int(ln U / ln q) + 1.

MATLAB's Statistics and Machine Learning Toolbox has a function **geornd** to generate geometric random variables.

https://www.mathworks.com/help/stats/geornd.html

BERNOULLI RANDOM VARIABLES

A single Bernoulli random variable P(1) = 1 – P(0) = p can be generated with the simple algorithm:

1. Generate a random number U(0,1).
2. If U < p, then set X = 1, else set X = 0.

Therefore, a sequence could be generated simply by looping through this algorithm the desired number of times. This can be done very efficiently using MATLAB's vector functionality.

```
bernp1 = rand(n,1) <= p
```

Another way to generate such a sequence is to realize that if the value of geometric random variable is n, this can be interpreted as a sequence of n-1 failures followed by a single success, which represents a Bernoulli sequence. This is used in many development environments because it can be more efficient than the direct method. However, in MATLAB the direct method is very efficient and is therefore preferred because it is more understandable than the more obscure method involving a geometric random variable.

POISSON RANDOM VARIABLE

A Poisson random variable can be generated with the following algorithm:

1. Generate a random number U(0,1).
2. i = 0; p = $e^{-\lambda}$; F = p
3. If U < F, then set X = I and stop
4. p = λp/(i+1); F += p; i++
5. Go to step 3

MATLAB's Statistics and Machine Learning Toolbox has a function **poissrnd** to generate Poisson random variables.

https://www.mathworks.com/help/stats/poissrnd.html

BINOMIAL (N,P) RANDOM VARIABLE

A binomial (n, p) random variable can be generated with the following algorithm:

1. Generate a random number U(0,1).
2. c = p/(1-p); i = 0; r = $(1\text{-}p)^n$; F = r
3. If U < F, then set X = I and stop
4. r *= [c(n-i)/(i+1)]; F += r; i++
5. Go to step 3

Another way of generating a binomial (n, p) variable is by using its interpretation as the number of success in b independent Bernoulli trials where each trial is a success with probability p. Therefore, we can simulate n Bernoulli trials and count the successes.

MATLAB's Statistics and Machine Learning Toolbox has a function **binornd** to generate binomial random variables.

https://www.mathworks.com/help/stats/binornd.html

NEGATIVE BINOMIAL RANDOM VARIABLE

A negative binomial (n, p) random variable can be generated with the following straightforward algorithm:

1. Generate n geometric(p) random variables.

2. Set $X = Y_1 + \ldots + Y_n$

There are other more efficient algorithms. If we encounter a situation where we need to generate a large number of negative binomial random variables, we can research some of these more efficient algorithms.

MATLAB's Statistics and Machine Learning Toolbox has a function **nbinrnd** to generate negative binomial random variables.

https://www.mathworks.com/help/stats/nbinrnd.html

REJECTION METHOD FOR GENERATING DISCRETE RANDOM VARIABLES

Rejection Method

STEP 1: Simulate the value of Y, having probability mass function q_j.

STEP 2: Generate a random number U.

STEP 3: If $U < p_Y/cq_Y$, set $X = Y$ and stop. Otherwise, return to Step 1.

The rejection method is pictorially represented in Figure 4.1.

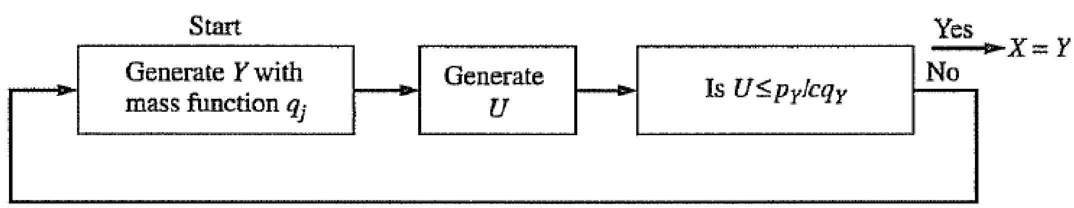


Figure 4.1. Acceptance–rejection.

EXERCISES

1) Write code that implements generators for each of the above-mentioned discrete random variable algorithms.

2) A pair of fair six-sided dice are to be rolled continually until all the possible outcomes 2, …, 12 have occurred at least once. Develop a simulation that estimates the expected number of dice rolls that are needed. [Ross, Simulation, Exercise 4.7]

3) Simulate the random variable X that has the probability function that p(n) = 0.11 when n = 5, 7, 9, 11, 13 and p(n) = 0.09 when n = 6, 8, 10, 12, 14. [Ross, Simulation, Exercise 4.14]

REFERENCES

M. Boon, Simulation Lecture Notes, Lecture 8, https://www.win.tue.nl/~marko/2WB05/

S. Ross, *Introduction to Probability Models 10th Edition*, Chapter 11

S. Ross, *Simulation 4th Edition*, Chapter 4

LESSON 7 – GENERATING CONTINUOUS RANDOM VARIABLES

UNIFORM DISTRIBUTION

We've used the uniform distribution before.

The MATLAB function `rand` returns a single uniformly distributed random number in the open interval (0,1).

$F(x) = (x – a) / (b – a)$

$f(x) = 1 / (b – a)$

$E[X] = (b – a) / 2$

$Var [X] = (b – a)^2 /12$

https://www.mathworks.com/help/matlab/ref/rand.html

NORMAL (GAUSSIAN) DISTRIBUTION

In MATLAB, the way to work with a normal distribution is to create a `NormalDistribution` probability distribution object in one of the following three ways:

- Create a distribution with specified parameter values using `makedist`.
- Fit a distribution to data using `fitdist`.
- Interactively fit a distribution to data using the Distribution Fitter app.

https://www.mathworks.com/help/stats/prob.normaldistribution.html

https://en.wikipedia.org/wiki/Normal_distribution

EXPONENTIAL DISTRIBUTION

$F(x) = 1 – e^{-\lambda x}$

$f(x) = \lambda e^{-\lambda x}$

$E[X] = 1/\lambda$

Standard Deviation of $X = 1/\lambda$

This distribution can be used to model any phenomenon with a positive value and a known mean $1/\lambda$. In particular, it is used to model time until failure/breakdown for a component or a machine, time between two consecutive customer orders, etc. Its fundamental property and claim to fame is the memoryless property, i.e., $P(X > x+t \mid X > t) = P(X > x)$. In other words, how long you've waited already is completely irrelevant when trying to predict how much longer you will have to wait. The exponential distribution is the only continuous distribution with this property. A process of consecutive arrivals when the inter-arrival times are independent and exponentially distributed is known as a Poisson process.

In MATLAB, the way to work with a normal distribution is to create an `ExponentialDistribution` probability distribution object.

The exponential distribution is not the same as the class of exponential families of distributions, which is a large class of probability distributions that includes the exponential distribution as one of its members, but also includes many others.

The exponential distribution is a special case of the gamma distribution.

GAMMA DISTRIBUTION

https://en.wikipedia.org/wiki/Gamma_distribution

https://www.mathworks.com/help/stats/gamma-distribution.html

BETA DISTRIBUTION

https://en.wikipedia.org/wiki/Beta_distribution

https://www.mathworks.com/help/stats/beta-distribution.html

INVERSE TRANSFORM METHOD FOR GENERATING CONTINUOUS RANDOM VARIABLES

Let U be a uniform (0,1) random variable. For any distribution function F, the random variable X defined by $X = F^{-1}(U)$ has distribution F. (Proof omitted)

REJECTION METHOD FOR GENERATING CONTINUOUS RANDOM VARIABLES

The Rejection Method

STEP 1: Generate Y having density g.
STEP 2: Generate a random number U.
STEP 3: If $U \le \frac{f(Y)}{cg(Y)}$, set $X = Y$. Otherwise, return to Step 1.

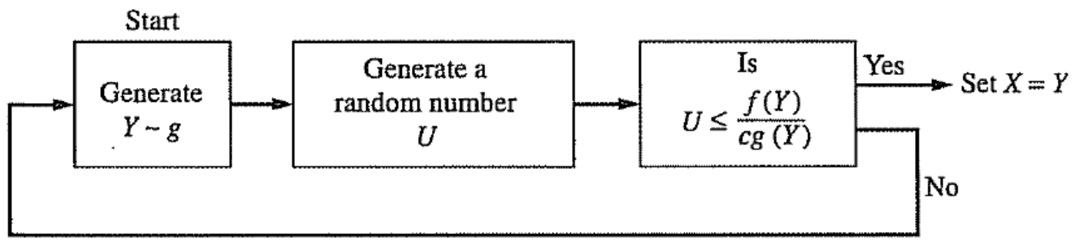


Figure 5.1. The rejection method for simulating a random variable X having density function f.

GENERATING THE FIRST T TIME UNITS OF A POISSON PROCESS

Generating the First T Time Units of a Poisson Process with Rate λ

STEP 1: $t = 0, I = 0$.
STEP 2: Generate a random number U.
STEP 3: $t = t - \frac{1}{\lambda}\log U$. If $t > T$, stop.
STEP 4: $I = I + 1$, $S(I) = t$.
STEP 5: Go to Step 2.

THINNING ALGORITHM FOR GENERATING THE FIRST T TIME UNITS OF A NONHOMOGENEOUS POISSON PROCESS

Generating the First T Time Units of a Nonhomogeneous Poisson Process

STEP 1: $t = 0, I = 0$.
STEP 2: Generate a random number U.
STEP 3: $t = t - \frac{1}{\lambda}\log U$. If $t > T$, stop.
STEP 4: Generate a random number U.
STEP 5: If $U \leq \lambda(t)/\lambda$, set $I = I + 1$, $S(I) = t$.
STEP 6: Go to Step 2.

EXERCISES

1) Use the inverse transform method to generate a random variable having distribution function $F(x) = \frac{x^2+x}{2}$. [Ross, Simulation, Exercise 5.3]

2) A casualty insurance company has 1000 policyholders, each of whom will independently present a claim in the next month with probability 0.05. Assuming that the amounts of the claims made are independent exponential random variables with mean $800, use simulation to estimate the probability that the sum of these claims exceeds $50,000. [Ross, Simulation, Exercise 5.10]

3) Use the rejection method with an exponential density having rate λ to generate a random variable having density function $f(x) = \frac{1}{2}x^2e^{-x}$. [Based on, Ross, Simulation, Exercise 5.16]

4) Write a program that generates the first T time units of a Poisson process having rate λ. [Ross, Simulation, Exercise 5.22]

5) Write a program that uses the thinning algorithm to generate the first ten time units of a nonhomogeneous Poisson process with intensity function $\lambda(t) = 3 + \frac{4}{t+1}$. [Ross, Simulation, Exercise 5.25]

6) Buses arrive at a sporting event according to a Poisson process with rate 5 per hour. Each bus is equally likely to contain either 20, 21, …, or 40 fans with the numbers in the different buses

independent. Write a simulation of the arrival of fans to the event as a function of time. [Ross, Simulation, Exercise 5.24]

REFERENCES

S. Ross, *Simulation 4th Edition*, Chapter 5

Rubinstein, *Simulation and the Monte Carlo Method*

Wackerly et al, *Mathematical Statistics with Applications 7th Edition*, Chapter 4

