A brief summary of the chapters and sections that we cover in Math 141

Chapter 1 Functions of Several Variables

Section 1 Examples of Functions of Several Variables

1. Many (most?) quantities depend on more than one other quantity. That is, most values are functions of several other variables (values that vary).
2. A multivariable function is one which has more than one input variable, for example $f(x, y)$ or $f(x, y, z)$, as opposed to a single variable function $f(x)$.
3. Sometimes we're interested in what combination of values of the variables would result in a particular fixed value of the function. These are level curves.

Section 2 Partial Derivatives

1. Derivatives tell us how one values changes as its input changes, or to be more precise, if its input value increase by one unit. For example, for $f(x)$, a function of one variable x, $\frac{df}{dx}(3) = 5$ means that if currently $x = 3$, then increasing x by one unit (from 3 to 4) would cause f to change (to *increase*) by 5.
2. In finding partial derivatives, think of one variable as "the variable" and the other variables as constants.
3. For a function $f(x, y)$ of two variables x and y, $\frac{\partial f}{\partial x}(3,7) = -6$ means that if currently $(x, y) = (3,7)$, then increasing x by one unit (from 3 to 4), but not changing y, would cause f to *decrease* by 6.
4. In general, for $f(x, y)$:
    - $\frac{\partial f}{\partial x}$ tells us how much f changes if x were to increase by 1 unit (and the value of y did not change)
    - $\frac{\partial f}{\partial y}$ tells us how much f changes if y were to increase by 1 unit (and the value of x did not change)
5. We can find second (and third and …) derivatives, including mixed derivatives, for example $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(\frac{\partial}{\partial y} f))$.
6. We can evaluate (find the value of) a derivative at a certain point, that is, at a certain value of (x, y).

Section 3 Maxima and Minima of Functions of Several Variables

1. For a function $f(x)$ of one variable, recall that if at a certain point $x = a$ we have $f'(a) = 0$ (that is, $\frac{df}{dx}(a) = 0$), then:
   - $f''(a) > 0$ (that is, $\frac{d^2f}{dx^2}(a) > 0$) $\Rightarrow f(x)$ has a local/rel. min at $x = a$.
   - $f''(a) < 0$ (that is, $\frac{d^2f}{dx^2}(a) < 0$) $\Rightarrow f(x)$ has a local/rel. max at $x = a$.
   - $f''(a) = 0$ (that is, $\frac{d^2f}{dx^2}(a) = 0$) $\Rightarrow$ we can't tell if there is min or max or neither at $x = a$.
2. For $f(x, y)$, if at $(x, y) = (a, b)$ we have $\frac{\partial f}{\partial x}(a, b) = 0$ and $\frac{\partial f(a,b)}{\partial y} = 0$, then where

$$D(x, y) = \begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{vmatrix} = \left(\frac{\partial^2 f}{\partial x^2}\right)\left(\frac{\partial^2 f}{\partial y^2}\right) - \left(\frac{\partial^2 f}{\partial x \partial y}\right)\left(\frac{\partial^2 f}{\partial y \partial x}\right) = \left(\frac{\partial^2 f}{\partial x^2}\right)\left(\frac{\partial^2 f}{\partial y^2}\right) - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$$

   we have
   - $D(a, b) > 0$, $\frac{\partial^2 f}{\partial x^2}(a, b) > 0 \Rightarrow f(x, y)$ has a local/rel. min at $(x, y) = (a, b)$.
   - $D(a, b) > 0$, $\frac{\partial^2 f}{\partial x^2}(a, b) < 0 \Rightarrow f(x, y)$ has a local/rel. max at $(x, y) = (a, b)$.
   - $D(a, b) < 0$, $\Rightarrow f(x, y)$ has a saddlepoint (neither a local/relative min or max) at $(x, y) = (a, b)$.
   - $D(a, b) = 0$, $\Rightarrow$ we can't tell if there is a min or max or saddlepoint at $(x, y) = (a, b)$.
3. There are lots of interesting and useful optimization problems.

Section 4 Lagrange Multipliers and Constrained Optimization

1. Sometimes we want to optimize (maximize or minimize) a function when there is some constraint (restriction) on what values its input values (its variables) can be.
2. Sometimes we can simply solve for one variable in terms of the others within the constraint and substitute this into the objective (the function to maximize or minimize) and optimize this function as described in Section 3. But sometimes we can't. This is where the Method of Lagrange Multipliers comes in.
3. In the <u>Method of Lagrange Multipliers</u>, the basic steps are:
    1. Create the auxiliary function $F(x, y, \lambda) = f(x, y) + \lambda g(x, y)$ from the <u>objective</u> $f(x, y)$ and the <u>constraint</u> $g(x, y)$ functions.
    2. Find $\frac{\partial F}{\partial x}$, $\frac{\partial F}{\partial y}$ and $\frac{\partial F}{\partial \lambda}$.
    3. Solve for λ in both $\frac{\partial F}{\partial x} = 0$ and $\frac{\partial F}{\partial y} = 0$.
    4. Set these two values of λ equal and use them to find a relationship between x and y: solve for one variable in terms of the other.
    5. Substitute this relationship of one of the variables in terms of the other into $\frac{\partial F}{\partial \lambda} = 0$ (which is simply the original constraint equation), and solve for that variable. Use that value to find the value of the other variable.
    6. Bonus: the value of λ tells us how much the objective would change if the constraint value were increased by one unit.

    When dealing with three (or more) variables $x, y, z,$ the steps are essentially the same, but Step 4 in particular is more complex.

Section 5 The Method of Least Squares

1. Given any data, we can find the line $y = Ax + B$ (I'm more used to the notation $y = mx + b,$ and maybe you are, too, but we'll follow the book's notation) that best fits the data.
2. This line is called the <u>best fit</u> line or <u>least squares</u> line. The process of finding and using this line is sometimes called <u>linear regression</u>. We can also find all sorts of other types of functions (other than straight lines) to fit a given set of data, but that is beyond what we are doing in this class.
3. The slope A and $y-$intercept B give us useful information from our data points (x, y) about how y is related to x.
4. We can use the line $y = Ax + B$ to:
    - Estimate the value of y that would result from a particular value x;
    - Estimate the value of x that would be needed to produce a particular y.

Section 6 Double Integrals

The ideas in this section are useful, but the book doesn't get into any real-world problems, so we'll not cover this section.

Chapter 2 Matrices

Section 1 Systems of Linear Equations with Unique Solutions

1. A system of linear equations is simply a collection of linear equations. A matrix is a way to organize and work with the numbers from the linear equations. (Actually, matrices are far more useful and powerful than merely a way to work with the numbers from the equations.)
2. Each equation is a condition or restriction. Each unknown (variable) is some value that you need to determine. As we will see in Section 2, for some system of equations there is more than one solution, and for some systems of equations there is no solution.
3. Gauss-Jordan (or Gaussian) elimination: get 1's on diagonal and 0's everywhere else. This is equivalent to eliminating variables in equations. An elementary row operation is a particular step we take as part of this process. These are summarized on pages 51 – 53:
   - Interchange any two equations.
   - Multiply an equation by a nonzero number.
   - Change an equation by adding to it a multiple of another equation.

   This process put the matrix into diagonal form.
4. Check your solution: just plug (substitute) the values you found back into the original equations.
5. How to recognize mistakes: *working backwards*, plug the values you found into the equations one step at a time. Where the values you found no longer satisfy the equation is where you made your mistake in Gaussian elimination.
6. The standard form equation (for two unknowns) is $ax + by = c$. It describes a relationship between the *unknowns* x and y, or it might describe a relationship between two *variables* x and y.
7. Technology is generally used in real-life settings, and can be used to check your homework. There are online tools you can use, and the book gives some information on using Excel or http://www.wolframalpha.com/. Calculators can also be used, but computers are computationally far more powerful and visually vastly superior.

Section 2 General Systems of Linear Equations

1. Gaussian elimination with <u>pivoting</u> is a more organized and more predictable way to do Gaussian elimination:
    - Transform the given entry into a one.
    - Transform all other entries in the same column into zeros.

   The process is described in detail on page 61, with more information on page 63.
2. For systems of linear equations, there is always 0 or 1 or ∞ solutions.
3. Generally:
    - $number\ of\ equations\ >\ number\ of\ unknowns\ \Rightarrow\ 0\ solution$
    - $number\ of\ equations\ =\ number\ of\ unknowns\ \Rightarrow\ 1\ solution$
    - $number\ of\ equations\ <\ number\ of\ unknowns\ \Rightarrow \infty\ solutions$

   But there can be exceptions.
4. How to recognize each case of 0 or 1 or ∞ solutions, after Gaussian elimination:
    - A row of *all* 0's can simply be ignored (the equation on that row was a redundant restriction/relationship between the unknowns).
    - A row of all 0's on the left (in the coefficient part of the matrix) along with any non-zero value on the right (the "right hand side" part of the matrix) means there is no solution, no matter what other numbers there are in the matrix.
    - If we end up with fewer non-zero rows than there are columns, then (assuming that there are no rows with all zeros on the left and something non-zero on the right, that is, assuming that there is a solution) there are ∞ solutions, and the free variables in the solution correspond to the non-pivot columns (the columns without a pivot 1 in them).

   See page 63.

Section 3 Arithmetic Operations on Matrices

1. The product of two matrices has real-life meaning, which depends on the meanings of the two matrices being multiplied.
2. The dimension $m \times n$ of a matrix A means A has m rows and n columns. A matrix is square if $m = n$, that is, it has the same number of rows as columns. It turns out that square matrices are pretty important. For example, $number\ of\ equations = number\ of\ unknowns$ usually means exactly one solution.)
3. A system of equations can be written in matrix notation $AX = B$ (this is more commonly written as $A\vec{x} = \vec{b}$ or $\mathrm{A}\mathbf{x} = \mathbf{b}$, but we'll follow this book's notation). In this case A is the coefficient matrix (the matrix of coefficients from the equations), X is the unknown (one-column) matrix (also called a vector) with the unknown values, and B is the (one-column) matrix of right hand side values.
4. Technology is generally used to do the work, but for this class use it to check the work that you'll do by hand. It's like how in learning arithmetic, you first learn how to do things by hand, but ultimately you use calculators or computers to do the work.
5. The identity matrix I is sort of the matrix version of the number 1: for any matrix A we have $AI = A$ and $IA = A$. ("Identity" is related to "identical," meaning "the same as" or "not different" or "unchanged," so multiplying a matrix by the identity matrix leaves that matrix unchanged.) So it turns out that in doing Gaussian elimination in Sections 1 and 2, we go from $[A|B]$ to $[I|X]$ where X is the solution to $AX = B$.

Section 4 The Inverse of a Square Matrix

1. The inverse of a matrix A^{-1} is the matrix such that $AA^{-1} = I$ and $A^{-1}A = I$. This is just like for a number a we have $a \cdot a^{-1} = 1$ and $a^{-1} \cdot a = 1$; that is, $a \cdot \frac{1}{a} = 1$ and $\frac{1}{a} \cdot a = 1$. Only square (number of rows = number of columns) matrices might have an inverse, but not all square matrices have inverses.
2. The inverse is quite useful: given the system of equations $AX = B$, multiplying both sides (on the left) gives us the solution $X = A^{-1}B$:
$$\begin{aligned} AX &= B \\ A^{-1}AX &= A^{-1}B \\ IX &= A^{-1}B \\ X &= A^{-1}B \end{aligned}$$
3. For 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $A^{-1} = \frac{1}{D}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ where determinant
$$D = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$
4. If the determinant $\neq 0$, then A has an inverse and the system of equations $AX = B$ has a unique (one and only one) solution $X = A^{-1}B$. If the determinant $= 0$, then there is no A^{-1} and $AX = B$ will either have no solution or infinite solutions, depending on right hand side B.

Section 5 The Gauss-Jordan Method for Calculating Inverses

1. There is not a nice, simple formula for finding A^{-1} for matrices A that are larger than 2×2.
2. To find A^{-1}, we do Gaussian elimination (row reduction) $[A|I] \to \cdots \to [I|A^{-1}]$.
3. If A cannot be transformed to I, then A has no inverse. A must be square to possibly have an inverse, but being square doesn't guarantee that A has an inverse.

Section 6 Input-Output Analysis

1. There are many real world problems in which matrices are very helpful in organizing information and solving various types of problems that lead to the problem $AX = B$. In this section is an interesting and important use of matrix inverses, important enough to win a Nobel Prize in 1973.
2. The problem we wish to solve is: how much to produce if

   Amount produced − Amount consumed = Demand
3. That is, where X is the amount of stuff we produce, D ("demand") is the amount we want to end up with, and A is the <u>input-out</u> ("<u>consumption</u>") <u>matrix</u>, solve for X in $X - AX = D$:
   $$X - AX = D \Rightarrow (I - A)X = D \Rightarrow X = (I - A)^{-1}D.$$
4. The i-th column of matrix A is the amount *consumed* if *producing* one unit of product i. The i-th column of matrix $(I - A)^{-1}$ is the amount needed *to produce* if we want to *end up* with one unit of product i.

Chapter 3 Sets and Counting

Section 1 Sets

1. The union $A \cup B$ of two sets is the collection of all items in *either* A or B (or both).
2. The intersection $A \cap B$ of two sets if the collection of all items in *both* A and B.
3. Set B is a subset of A ($B \subseteq A$) if everything in B is also in A.
4. The complement A' of set A is everything that is not in A but that could be; that is, everything that is not in A but that is in the universal set U.
5. The empty set $\emptyset$ is the set that has nothing in it. It's like the set version the number 0.
6. For any set S we have $S \cup S' = U$ and $S \cap S' = \emptyset$.

Section 2 A Fundamental Principle of Counting

1. The number of things in the union of two sets is the number of things in the one set plus the number in the other set minus the part that is counted twice (the part that is in both sets). That is,
$$n(A \cup B) = n(A) + n(B) - n(A \cap B).$$
This is called the inclusion-exclusion principle.
2. If A and B are mutually exclusive (that is, if $A \cap B = \emptyset$, so that $n(A \cap B) = 0$), then $n(A \cup B) = n(A) + n(B)$.
3. De Morgan's Laws:
$(S \cup T)' = S' \cap T'$
$(S \cap T)' = S' \cup T'$
In words:
Everything not in (S or T) = everything (not in S) and (not in T).
Everything not in (S and T) = everything (not in S) or (not in T).
Some intuition: *or* is the opposite of *and*.

Section 3 Venn Diagrams and Counting

1. Venn Diagrams are a nice way to visualize how sets are related to each other. They are useful for two sets or for three sets. For four or more sets (as we'll see later), they are not as useful.
2. The inclusion-exclusion principle for three sets is
$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$
and for four sets it's
$$\begin{aligned} n(A \cup B \cup C \cup D) = {} & n(A) + n(B) + n(C) + n(D) \\ & - n(A \cap B) - n(A \cap C) - n(A \cap D) - n(B \cap C) - n(B \cap D) - n(C \cap D) \\ & + n(A \cap B \cap C) + n(A \cap B \cap D) + n(A \cap C \cap D) + n(B \cap C \cap D) \\ & - n(A \cap B \cap C \cap D) \end{aligned}$$

Crazy fun, eh? Hopefully you can see the pattern here.

Section 4 The Multiplication Principle

1. The multiplication principle: the total number of ways to do a task consisting of multiple choices is the product of the number of ways there is to perform each choice. See the more detailed definition on page 135.

Section 5 Permutations and Combinations

1. From the dictionary: "permute" means "to order."
2. Permutation: choose some items in a particular order ("order matters").
   Combination: choose some items in no particular order ("order doesn't matter").
3. How permutations and combinations are related:

   The number of ways to *choose* and *order* r items from n =
   The number of ways to *choose* r items from n $*$ The number of ways to *order* r items.

   That is, $P(n,r) = C(n,r) * r!$
4. Formulas: $P(n,r) = n(n-1)(n-2)\cdots(n-r+1)$. (There are r factors in this product.) This is simply the multiplication principle choosing r things from n in which you don't get to choose the same thing twice.
5. The formula for $P(n,r)$ is useful mostly to give us a formula for combinations:

$$C(n,k) = \frac{P(n,r)}{r!} = \frac{\frac{n!}{(n-r)!}}{r!} = \frac{n!}{r!\,(n-r)!}\ .$$

Section 6 Further Counting Techniques

1. There is nothing new in this section, just more different counting problems.

(We will not cover Section 7.)

Section 8 Multinomial Coefficients and Partitions

1. Three equivalents notations for combinations:

$$\binom{n}{r} = {}_nC_r = C(n,r) = \frac{n!}{r!\,(n-r)!}\,.$$

2. In choosing k items from n total, there are also $n-k$ items that we are not choosing. So in choosing r items from n it's like we are dividing the n items into two groups of sizes k and $n-k$, which we can write

$$\binom{n}{r,n-k} = \frac{n!}{r!\,(n-r)!}\,.$$

3. More generally, the number of ways that we can partition (divide) n items into groups of sizes $n_1, n_2, \dots, n_k$ (where $n_1 + n_2 + \cdots + n_k = n$) is

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!\ n_2!\ \cdots\ n_k!}\ .$$

Chapter 4 Probability

Section 1 Experiments, Outcomes, Sample Spaces and Events

1. Four new ideas

Term	Meaning	Example
Experiment	An activity with an observable result	Roll 2 dice
Outcome	A possible result of an experiment	A 3 and a 6
Sample space	The set of all possible outcomes, denoted S	All 36 outcomes
Event	A particular type/collection of outcomes	Sum of 9 (3/6, 4/5, 5/4, 6/3)

2. The same rules for sets apply to outcomes, sample spaces and events, which makes sense since these are all sets (this is why the chapter on sets came before this chapter). Two events are mutually exclusive if they cannot both occur, that is, if there is no overlap between the two events.
3. Notation: events are sometimes denoted E or F, or sometimes A or B, but you can really use any letter you want (this is like how variables are often called x or y, but you can use any letter).

Section 2 Assignment of Probabilities

1. Probabilities are decimal values, like .75, which is the same as 75% (75 per cent literally means 75 per 100, that is, 75/100; that is, .75).
2. Property 1: The probability of any event is between 0 (an impossible event) and 1 (a certain event).
3. Property 2: The sum of the probabilities for all of the possible outcomes is always 1. That is, there is always a 100% chance that one of the possible outcomes will occur (duh!). Example: when you roll a single standard die, there is a probability of 1 that you will roll a 1 or 2 or 3 or 4 or 5 or 6.
4. Useful observation: $\Pr(E) + \Pr(E') = 1$, which means $\Pr(E) = 1 - \Pr(E')$.
5. Addition Principle: if an event can happen in multiple ways, then the probability of that event is the sum of the probabilities of the individual outcomes. Example: roll a single die, and the probability of rolling an odd number is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6}$.
6. Inclusion-Exclusion Principle: For events E and F, $\Pr(E \cup F) = \Pr(E) + \Pr(F) - \Pr(E \cap F)$. In words, $\Pr(E \text{ or } F) = \Pr(E) + \Pr(F) - \Pr(E \text{ and } F)$.
7. Odds: If $\Pr(E) = p$, then the odds of E occurring are p to $1-p$, that is, $p : 1-p$, that is, $\frac{p}{1-p}$. If the odds of E occurring are a to b (that is, $a{:}b$, that is, $\frac{a}{b}$), then the probability of E occurring is $\frac{a}{a+b}$.
   Example: the probability of the Lakers will win their next game is .25 $\Leftrightarrow$ the odds the Lakers will *win* their next game are 1 to 3 ($\Leftrightarrow$ the odds the Lakers will *lose* are 3 to 1).

Section 3 Calculating Probabilities of Events

1. Much of probability is very intuitive. The simplest rule of probability is:

$$\Pr(E) = \frac{number\ of\ ways\ E\ can\ occur}{total\ number\ of\ outcomes}$$

2. As seen in the previous section, $\Pr(E) = 1 - \Pr(E')$. Sometimes to find $\Pr(E)$ it is actually easier to find $\Pr(E')$, sort of a "back door" approach. One instance of this is when E can occur in many ways while E' only occurs in a few ways.

Section 4 Conditional Probabilities and Independence

1. Conditional probability: the probability that one event will be true or will occur given that some other event is true or has occurred.
2. Since

$$\Pr(E) \quad = \frac{number\ of\ ways\ E\ can\ occur}{total\ number\ of\ outcomes}$$

then we have

$$\Pr(E|F) = \frac{number\ of\ ways\ E\ can\ occur\ \ \text{where F is true}}{total\ number\ of\ outcomes\ \ \ \text{where F is true}}$$

$$= \frac{number\ of\ ways\ E\ and\ F\ can\ occur}{number\ of\ ways\ F\ can\ occur} = \frac{n(E \cap F)}{n(F)}$$

which (where S is the entire sample space) also leads to

$$\Pr(E|F) = \frac{n(E \cap F)/n(S)}{n(F)/n(S)} = \frac{\Pr(E \cap F)}{\Pr(F)}$$

So to summarize, there are two equivalent formulas for $\Pr(E|F)$:

$$\Pr(E|F) = \frac{n(E \cap F)}{n(F)} = \frac{\Pr(E \cap F)}{\Pr(F)}\ .$$

3. The above also leads to the formula:

$$\Pr(E \cap F) = \Pr(F) \cdot \Pr(E|F)\ .$$

Reversing the roles of E and F gives a second version of this formula:

$$\Pr(E \cap F) = \Pr(E) \cdot \Pr(F|E)\ .$$

4. Events E and F are said to be independent if knowing that one is true or has occurred does not affect whether the other is true or will occur. There are three definitions/results of independence:

$$\Pr(E|F) = \Pr(E)$$

$$\Pr(F|E) = \Pr(F)$$

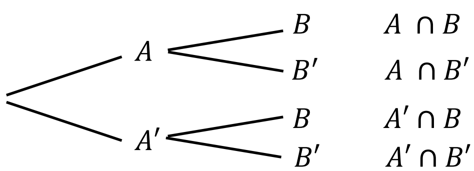
$$\Pr(E \cap F) = \Pr(E) \cdot \Pr(F)$$

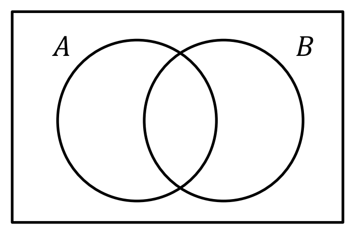
Section 5 Tree Diagrams

1. Tree diagrams help us visualize and organize all of the possible outcomes.
2. Some terms/ideas that arise in medical (or similar) testing are given on page 212. In the context of testing positive (+) or negative (-) and having the condition (C) for which you are testing or not (C'):

Sensitivity	$\Pr(+|C)$
Specificity	$\Pr(-|C')$
Positive predictive value	$\Pr(C|+)$
Negative predictive value	$\Pr(C'|-)$

   I'm not too interested in your memorizing these terms, but we will be interested in what they mean in looking at the effectiveness of medical testing. The closer these are to 1 (i.e. 100%), the more reliable the test.
3. Given two events A and B, there are two ways A could occur: when B occurs, or not. That is,
   $$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B').$$
   Similarly,
   $$\Pr(B) = \Pr(A \cap B) + \Pr(A' \cap B).$$
   This can be visualized with a tree diagram or Venn Diagram.





Section 6 Bayes' Theorem, Natural Frequencies

1. Since $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B|A)$ and $\Pr(A \cap B) = \Pr(B) \cdot \Pr(A|B)$, then we have
   $$\Pr(B|A) = \frac{\Pr(\mathrm{B} \cap \mathrm{A})}{\Pr(A)} = \frac{\Pr(\mathrm{B}) \cdot \Pr(\mathrm{A}|\mathrm{B})}{\Pr(A)} = \frac{\Pr(\mathrm{B}) \cdot \Pr(\mathrm{A}|\mathrm{B})}{\Pr(\mathrm{B}) \cdot \Pr(\mathrm{A}|\mathrm{B}) + \Pr(\mathrm{B}') \cdot \Pr(\mathrm{A}|\mathrm{B}')}.$$
   This is Bayes' Theorem.
2. The generalized version of Bayes' Theorem is given on page 219.
3. The natural frequency of an outcome is simply how much of the population would be in each outcome given a particular population size. For example, see Example 3 and Figure 3 on pages 218 – 219.

We will not cover Section 7.

Chapter 5 Probability and Statistics

Section 1 Visual Representations of Data

1. A picture is worth 1000 words. Visuals give us information quickly and simply, but often they cannot give us as precise information as a table of actual values.
2. Of the different ways to visualize data, the histogram will be of most interest to us in this class.

Section 2 Frequency and Probability Distributions

1. A distribution consists of (1) what the possible outcomes are and (2) *how often* each outcome occurred.
2. A frequency distribution consists of (1) what the possible outcomes are and (2) *how many times* each outcome occurred.
3. A relative frequency distribution consists of (1) what the possible outcomes are and (2) *what fraction of the time* each outcome occurred. This is more useful than the frequency distribution.
4. Area in a relative frequency distribution $=$ probability. See Figure 4 on page 247 as an example.
5. A random variable in an experiment is a value that (1) is unknown or unpredictable ("random") and (2) varies from observation to observation ("variable"). If we know the distribution of a random variable X, then we can find the distribution of any variation of X as well, such as $X^2 + 3$. We won't really get into why we care about this in this class.

Section 3 Binomial Trials

1. "Bi" = two and "nomial" = number, so "binomial" literally means "two numbers." A binomial trial is simply an experiment with exactly two possible outcomes. A few examples:
   a. Trial: shoot free throw. Outcomes: make or miss.
   b. Trial: gender of class members. Outcomes: male or female.
   c. Trial: employment status. Outcomes: employed or not.
2. Notation for binomial trials:
   a. p is probability of success (success if one of the two outcomes, failure is the other outcome—it doesn't matter which outcome you call success and which you call failure, as long as you are consistent once you have decided which is which).
   b. q is the probability of failure. Note that $q = 1 - p$.
   c. n is the number of trials (e.g. the number of shots you take, the number of class members, or the number of American adults).
   d. X is the number of successes (e.g. the number of shots you make, the number of class members who are male, or the number of American adults who are employed).
3. So the probability of having k successes in a total of n binomial trials which has a probability of success of p is

$$Pr(X = k) = C(n, k)p^k(1 - p)^{n-k} = \binom{n}{k} p^k(1 - p)^{n-k} = \binom{n}{k} p^k q^{n-k} .$$

$p^k(1 - p)^{n-k}$ is the probability of one particular outcome of k successes and $n - k$ failures, and $C(n, k)\binom{n}{k}$ is the number of ways of having k of the n trials be successes. So in finding $\Pr(X = k)$, you are adding $p^k(1 - p)^{n-k}$ up $\binom{n}{k}$ times.

Section 4 The Mean

1. The <u>mean</u> is just another name for average. If you are finding the mean of a bunch of numbers $x_1, x_2, \dots, x_r$ which occur with frequencies $f_1, f_2, \dots, f_n$, then the mean $\bar{x}$ of these numbers is

$$\bar{x} = \frac{\overbrace{x_1 + x_1 + \cdots + x_1}^{f_1 \text{ times}} + \overbrace{x_2 + x_2 + \cdots + x_2}^{f_2 \text{ times}} + \cdots + \overbrace{x_r + x_r + \cdots + x_r}^{f_r \text{ times}}}{f_1 + f_2 + \cdots + f_r} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_r f_r}{n}$$

where $n = f_1 + f_2 + \cdots + f_r$.

2. Another way to write this is

$$\bar{x} = x_1\left(\frac{f_1}{n}\right) + x_2\left(\frac{f_2}{n}\right) + \cdots + x_r\left(\frac{f_r}{n}\right)$$

So to find the mean, we find a <u>weighted average</u>, where the weighting for each value is simply what fraction of the time $\frac{f_i}{n}$ each outcome x_i occurred.

3. If we have all values from a given population, then the population (of size N) mean μ (a Greek "m," pronounced "$my\bar{u}$") is

$$\mu = x_1\left(\frac{f_1}{N}\right) + x_2\left(\frac{f_2}{N}\right) + \cdots + x_r\left(\frac{f_r}{N}\right).$$

4. If we have a probability p_i for each outcome x_i occurring, then the expected mean, also called
the <u>expected value</u> of random variable X, is

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_r p_r \,.$$

(In the book they use N rather than r for the number of possible outcomes.) Note that the expected value is not necessarily one of the values that can occur, so in this sense "expected value" is a misnomer. It's simply on average what we expect to happen.

5. Notice how the above three formulas are basically the same thing. The first two are using the fraction of the time each outcome *did* occur and the third one is the fraction of the time each outcome *should* occur.
6. The expected value for a binomial random variable is simply $E(X) = np$, where n is the number of binomial trials and p is the probability of success.

Section 5 The Variance and Standard Deviation

1. The variance σ^2 measures how much variation there is in all of our outcomes. More precisely, it is a weighted average value of how different each outcome (squared) is from the mean. The weightings $\frac{f_i}{N}$, as before, correspond to how often each outcome x_i occurs.
$$\sigma^2 = (x_1 - \mu)^2 \left(\frac{f_1}{N}\right) + (x_2 - \mu)^2 \left(\frac{f_2}{N}\right) + \cdots + (x_r - \mu)^2 \left(\frac{f_r}{N}\right).$$
2. If we have a probability p_i for each outcome x_i occurring, then the variance is
$$\sigma^2 = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \cdots + (x_r - \mu)^2 p_r .$$
3. The standard deviation σ is simply the square root of the variance, that is, $\sigma = \sqrt{\sigma^2}$, and is often the value we're interested in (i.e. the reason we find the variance). It is approximately the average distance ("standard deviation") from each outcome to the mean.
4. The sample variance s^2 is if you have some portion (a sample) from a larger population. In this case you divide by the sample size n minus 1, as seen in the formulas on page 274.
5. Chebychev's Inequality on page 275 gives us a rough estimate for what portion of our outcomes/data will be in a certain range around the mean μ.
6. It turns out that the standard deviation of a binomial distribution is $\sigma = \sqrt{np(1-p)}$. This isn't that useful to us now, but it will be in Section 7.

Section 6 The Normal Distribution

1. Remember that a distribution consists of (1) the possible outcomes and (2) how much of the time each outcome occurs. Normal means *typical* or *common*. A normal distribution is a standard or typical distribution: the majority of values are around the mean, not many extreme values (far from the mean), and the distribution is balanced/symmetric about the mean. It's a bell curve.
2. Given the distribution of some data—that is, given its center (its mean) and how spread out it is (its standard deviation)—we can estimate what fraction of that data would be in a certain range. This is done by converting an actual value to a corresponding Z value (also called a standard value) $Z = \frac{X-\mu}{\sigma}$ and using Table 2 on page 283 (or in real life use technology, such as Excel), to determine what fraction of all values are in a certain range, as shown in more detail on page 286.
3. Note that
$$Z = \frac{X-\mu}{\sigma} \quad \Leftrightarrow \quad X = \mu + Z \cdot \sigma,$$
so Z just tells us how many standard deviations from the mean a particular value X is.
4. We can use the above to find percentiles: $x_p = \mu + z_p \cdot \sigma$, where z_p is the number standard deviations from the mean corresponding to pth percentile, x_p.
5. The first, second and third quartiles are the same as the 25th, 50th and 75th percentiles. The 50th percentile is simply the median, the value that is exactly in the middle of the data.
6. For normal distributions, a useful rule of thumb is that 68% (about 2/3) of all data are within 1 standard deviation of the mean, about 95% (most) are within 2 standard deviations, and 99.7% (nearly all) data are within 3 standard deviations of the mean.
7. So a good way of talking about extreme a particular value is to specify how many standard deviations above or below the mean that value is.

Section 7 Normal Approximation to the Binomial Distribution

1. We can treat a binomial distribution as if it is a normal distribution (the larger n is, the better a normal distribution approximates a binomial distribution). In particular, we can find a Z value
$$Z = \frac{X-\mu}{\sigma} = \frac{X-np}{\sqrt{np(1-p)}}$$
where X is the number of successes in n total trials.
2. This approximation is very accurate when both $np > 5$ and $n(1-p) > 5$, but the larger n is, the more accurate this approximation will be.