n choose k = $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

MA 262 – Probability and Statistics Notes

standard deviation $s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$

Preliminaries

Prior to beginning the course, a few introductory remarks are in order.

Definition of Statistics

Statistics is the collection, analysis, interpretation, and display of data. There are 2 branches of statistics:

- **Descriptive statistics** – the methods used to summarize and organize information regarding a population based a sample of that population. These include, but are not limited to, graphs, charts, tables, and calculations.
- **Inferential statistics** – the methods used to infer conclusions regarding a population based on the information drawn from a sample of that population. These include, but are not limited to, hypothesis tests, confidence intervals, and *p*-values.

Population vs. Sample

It is important to be able to distinguish a population and a sample when dealing with both probability and statistics. A **population** is the entire set of individuals, items, or objects under consideration in a statistical study.

A **sample** is a representative subset of a population, sharing the sample characteristics of that population, from which inferences can be drawn and then extrapolated to the population.

Example Suppose you wanted to determine the average height of an adult male in Milwaukee County. One way to do this is to obtain the height of every adult male in the county. (This would be the population.) Clearly, this is impractical. Instead, a representative sample is drawn from the population, usually much smaller in number than the population, and the average height of that sample is determined. If the sample is truly representative, the average height obtained from the sample can be used as a very reasonable estimate of the average height of the population.

How do we go about choosing a "representative" sample from a population if the purpose is to estimate some characteristic(s) of that population? Suppose we consider the previous example. Would it seem reasonable to base our estimate on a sample obtained from those leaving the Milwaukee Bucks locker room after a tryout session? Would it seem reasonable to base the estimate on a sample obtained from standing outside the jockey's locker room at the track?

The problem with each of these samples is that the individuals they are measuring tend to share a common characteristic, a characteristic that is not representative of the population and that can severely affect the population characteristic trying to be estimated.

For a sample to have the highest chance of being truly representative of a sample, it must be *randomly* selected. That is, there can be no bias inherent in the selection process. Such a sample is known as a **random sample**.

Data

The word "data" is plural. Therefore it is necessary to say "The data are …" and not "The data is …" (The exception, of course, is when referring to a character in *Star Trek: The Next Generation.*) The singular form of data is "datum."

Types of Data

Nominal data (also known as Categorical data)– Different data values can be placed into categories but the categories themselves imply no specific order.

Example Men/women, Republicans/Democrats, and Igneous/Sedimentary/Metamorphic.

Ordinal data – Different data can be placed in logical order, but arithmetic is not meaningful.

Example – Restaurant (movie) ratings. A 4-star restaurant cannot be assumed to be twice as good as a 2-star restaurant. (Letter grades also.)

Interval data – These data are such that the difference between 2 data values is meaningful.

Example – The difference between 65 degrees and 70 degrees is the same as the difference between -10 degrees and – 5 degrees.

Ratio data – These data have the same properties of interval data, but zero is clearly defined. When a value of zero is encountered, none of the variable is present.

Example – Heights and weights are ratio data, while most temperatures (C and F) are interval data. (Kelvin temperatures are ratio data.) When working with ratio data, but not interval data, you can consider the ratio of two measurements. A weight of 4 grams is twice a weight of 2 grams, because weight is ratio data. A temperature of 100 degrees C is not twice as hot as 50 degrees C, because temperature C is not ratio data.

Measurements of Central Position

There are several ways to determine a "typical" data point. (In many circumstances, a data point in the middle of the set of data is considered to be a typical data point. Hence the reference to "central position.")

Mean – The average of a set of data. That is, if there are n data points and if x_i refers to the i^{th} data value, then the mean $\bar{x}$ of the data set is

$\bar{x}$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Example Find the mean of the data values 7, 5, and 12.

$= \frac{24}{3} = 8$

Median – The middle piece of data when the data values are placed in ascending order. Therefore, for the purposes of determining the median of a set of n pieces of data, x_1 represents the smallest data value and x_n represents the largest piece of data. If n is an ***odd*** positive integer the median $\tilde{x}$ is the middle piece of ordered data. That is, $\tilde{x} = x_{(n+1)/2}$. If n is an ***even*** positive integer the median $\tilde{x}$ is the average of the middle 2 pieces of ordered data. That is, $\tilde{x} = \frac{x_{n/2} + x_{(n/2)+1}}{2}$.

$\tilde{x}$

Example Find the median of the data values 7, 5, and 12.

5, 7, 12

$= 7$

Example Find the median of the data values 7, 5, 12, and 6.

5, 6, 7, 12

mean

= 6.5

Measure(s) of Dispersion

Knowledge of the mean of a set of data is not sufficient in described the set of data. For example, the following data sets each have the same mean:

We need a measurement of how spread out the data are.

Data Set	Data Values				
1	100	0	0	0	0
2	10	15	20	25	30
3	20	20	20	20	20
4	5	12	24	27	32

Each data set has a mean of $\bar{x} = 20$, and yet the data sets have decidedly different appearances. **Data Set 1** has 4 identical data values with no spacing between them and 1 considerably larger value spaced very far from the others. **Data Set 2** has evenly spaced data values. **Data Set 3** has no spacing between the data values. **Data Set 4** seems to have randomly spaced data values.

The difference in the data sets is how the data values are spaced, or *dispersed*, from each other. The most common measure of dispersion is the **standard deviation**, which is *an approximation to the average distance of a data point from the mean of the data.*

Again, let x_i represent the i^{th} data value, let x_i^2 represent the square of the i^{th} data value, let $\sum_{i=1}^{n} x_i$ represent the sum of the data values, and let $\sum_{i=1}^{n} x_i^2$ represent the sum of the squares of the data values. Then the formula for determining the standard deviation s of a set of data is

Familiarize →

$$s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$$

This formula can also be written as

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

The numerator in this equivalent form of the formula for the standard deviation, i.e.,

$$\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

will be used in a formula at the end of the course when considering analysis of variance (ANO-VA).

Example Find the standard deviation of the data values 7, 5, and 12.

x	x^2
7	49
5	25
12	144
24	218

$$\frac{3(218) - (24)^2}{3 \cdot 2} = \frac{654 - 576}{6} = s = \sqrt{13}$$

---

p = probability

① $p \geq 0$ (non-negative)

② $\sum_{i=1} p_i = 1$ } $0 \leq p \leq 1$ (probability is between 0 and 1)

Random variable: Fair coin tossed 3 times

X = number of times coin lands heads up.

$X = 0, 1, 2, 3$

The square of the standard deviation is known as the variance. That is, the variance s^2 is

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

In this course, the variance will only refer to the square of the standard deviation and have no further interpretation. However, the variance has a very useful interpretation in the field of regression analysis, which might be discussed in other courses.

Although the standard deviation (and the related variance) is the only measure of dispersion that will be discussed now, there are other measures of dispersion. These include, but are not limited to, the range, percentiles, and z-scores. (z-scores will be discussed later.)

Introduction to Probability

We often observe or measure the values of variables whose specific individual outcome(s) at any particular time cannot be predicted in advance. Such variables are known as **random variables** and such observations are known as **random experiments**. As a definition, a random variable is a function that assigns a number to each of the possible outcomes of a random experiment.

For example, a coin on which one side is designated "heads" and the other side is designated "tails" is tossed 4 times. Let the random variable be the number of heads observed. If we let an upper case letter represent the random variable and a lower case letter represent the values the random variable can assume, then we can say

$$X = \textit{number of heads observed}, \; x = 0, 1, 2, 3, 4$$

On any succession of 4 tosses, there is no way to predict how many heads will be observed. That is, the exact value of X is random. Hence, X is a random variable.

Any specific result of a random experiment is known as an **outcome**. For example, if a coin is tossed 2 times, and if H represents a head and T represents a tail, then one outcome is HT. A set of outcomes is known as an **event**. For example, if a coin is tossed 2 times, the event of obtaining exactly 1 head is the set of outcomes HT and TH. The **sample space S** of an experiment is the set of all possible outcomes. If a coin is tossed 2 times, the sample space is HH, HT, TH, and TT.

We will define the word "**probability**" as a measure of the "likelihood of the occurrence of a stated event." If the event is A, then we will use the notation **Pr (A)** to represent "the probability of the occurrence of event A."

There are essentially 3 ways of assigning probabilities to events.

1. **Subjective** – The probabilities are based on subjective feelings or belief. This is the method used by many gamblers.

2. **Empirical** – Experiments are performed repeatedly in an attempt to infer associated probabilities. Probabilities are then usually assigned based on the relative frequencies of the occurrences of the outcomes. For example, suppose you have a coin which you feel is biased. Two hundred tosses of this coin result in the observance of 78 heads. You might then conclude, empirically, that

$$Pr(heads) = \frac{78}{200} = 0.39$$

3. **Axiomatic** – Probability models are studied in a rigorous mathematical framework that provides rules that must always be obeyed. This is the approach that will be used in this course and that is used in any scientific field.

Axioms are statements that are taken to be self-evident and that require no proof. The axioms that we will assume are as follows:

1. $Pr(A) \geq 0$ (Probabilities *cannot* be negative)
2. $\sum Pr(A) = 1$ (The sum of all related probabilities *must* be one.)

The combination of these two properties implies that

$$0 \leq Pr(A) \leq 1$$

We have already discussed the nature of a random variable. There are 2 types of random variables:

Summation
1. A **discrete random variable** is one whose values are finitely countable. Put another way, all of the values of a discrete random variable can be put into a one-to-one correspondence with a subset of the positive integers. When the values of a discrete random variable are placed in ascending order, no other value(s) of the discrete random variable lie between any two consecutive listed values.

integral
2. A **continuous random variable** is one whose values are not finitely countable. Put another way, the values of a continuous random variable cannot be put into a one-to-one correspondence with the positive integers. Between any two continuous random variable values lie an infinite number of other values of that random variable.

Perhaps the best way to distinguish between discrete and continuous random variables is to use the following, which is an oversimplified statement but one which is often useful:

Discrete random variables ***count*** *things. Continuous random variables* ***measure*** *things.*

Entire courses exist in the study of probability. We will only be studying a few basic applications necessary to the study of statistics.

Discrete Probability Distributions

A **probability distribution funtion**, also known as a **probability mass function (pmf)**, is a way of assigning probabilities to values of the random variable. In the case of discrete random variables, the probability distribution is often given in tabular form. However, the probability distributions of continuous random variables cannot be given in tabular form. Instead, they are given in terms of a formula.

Even though a formula can often be given for a discrete probability distribution, most times discrete probability distributions are presented in tabular form. When placed in tabular form, a discrete probability distribution is a table of the values the random variable can assume along with the associated probability for each value.

Example A fair pair of dice is rolled. Assume that the dice are distinguishable in some way. Construct the appropriate probability distribution.

one is green
one is red
X = number of spots on upper faces
x = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Pr(x = k) = Probability that x is k.
probability of rolling a 7 != $\frac{1}{11}$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Example Determine a pmf that generates the probability of the previous example.

$$Pr(x) = \frac{6 - |x - 7|}{36}$$

x	Pr(x)
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

Generating pmfs might not be easy to accomplish when a list of outcomes or events needs to be generated. Often the actual list of outcomes or events is not important but rather *how many* of them there are. We need a quick way of counting outcomes and events without the necessity of listing them. One such convenient method is through use of the binomial coefficient.

Definition The *binomial coefficient* $\binom{n}{k}$ represents the number of ways in which k objects can be selected, or chosen, from a group of n objects that is at least as large as the group of k objects. $\binom{n}{k}$ is read as "*n choose k*." The number the binomial coefficient represents is

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

where

$$k! = k(k-1)(k-2)\cdots(3)(2)(1), \quad k \equiv \textit{non-negative integer}$$

$0! = 1$? true

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

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

Let $n = 1$

$\frac{1}{1} = 0!$

Example *List* the ways in which 4 tosses of a coin result in 2 heads.

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{(4)(3)(2)(1)}{(2)(1)(2)(1)} = \frac{12}{2} = 6$$

HHTT
THHT
TTHH
HTHT
THTH
HTTH

6 ways

Example Use the binomial coefficient to determine the ***number of ways*** in which 4 tosses of a coin can result in 2 heads.

$$\binom{10}{4} = \frac{(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)}{(4)(3)(2)(1)(6)(5)(4)(3)(2)(1)}$$

210 ways

Example Determine the value of the constant k such that $f(x) = k\binom{7}{5-x}$, $x = 0, 1, 2, 3$ can serve as a pmf. — probability mass function

$f(x) = k\binom{7}{5-x}$ $x: 0, 1, 2, 3$

$f(0) = k\binom{7}{5} = 21k$

$f(1) = k\binom{7}{4} = 35k$

$f(2) = k\binom{7}{3} = 35k$

$f(3) = k\binom{7}{2} = 21k$

It <u>must</u> be that $21k + 35k + 35k + 21k = 1$

$70k + 42k = 1$

$112k = 1$

$k = \frac{1}{112}$

$f(0) = \frac{21}{112}$

$f(1) = \frac{35}{112}$

$f(2) = \frac{35}{112}$

$f(3) = \frac{21}{112}$

x	f(x)
0	$\frac{21}{112}$
1	$\frac{35}{112}$
2	$\frac{35}{112}$
3	$\frac{21}{112}$

Definition A **probability distribution** is (usually) a tabular listing of the values of the *discrete* random variable and the associated probabilities.

Example For the pmf just determined, give the probability distribution.

Example Find another binomial coefficient numerically equal to $\binom{5717}{3275}$. (Notice that the problem is not to evaluate this binomial coefficient but rather to find a binomial coefficient of the form $\binom{a}{b}$ such that $\binom{a}{b} = \binom{5717}{3275}$.)

Example Determine the value of the constant k so that $f(x) = k\binom{2}{x}\binom{3}{3-x}$, $x = 0, 1, 2$ can serve as a pmf.

$$f(x) = k\binom{2}{x}\binom{3}{3-x},\ x = 0, 1, 2$$

$$f(0) = k\binom{2}{0}\binom{3}{3} = k$$

$$f(1) = k\binom{2}{1}\binom{3}{2} = 6k$$

$$f(2) = k\binom{2}{2}\binom{3}{1} = 3k$$

$$k + 6k + 3k = 1$$

$$10k = 1$$

$$k = \frac{1}{10}$$

$$f(0) = \frac{1}{10}$$

$$f(1) = \frac{6}{10}$$

$$f(2) = \frac{3}{10}$$

Example What is the probability distribution for the previous example?

x	$f(x)$
0	$\frac{1}{10}$
1	$\frac{6}{10}$
2	$\frac{3}{10}$

Example For the previous probability distribution, what is $Pr(X \le 1)$?

$$Pr(X \le 1) = f(0) + f(1)$$
$$= Pr(X=0) + Pr(X=1) =$$
$$= \frac{1}{10} + \frac{6}{10} = \frac{7}{10}$$

Example Determine the value of the constant k so that $f(x) = \frac{k - x^2}{10}$, $x = 0, 1, 2, 3, 4$ can serve as a probability distribution.

$$f(x) = \frac{k - x^2}{10}, \quad x = 0, 1, 2, 3, 4$$

$$f(0) = \frac{k}{10} = \frac{k}{10}$$

$$f(1) = \frac{k-1}{10} = \frac{k-1}{10}$$

$$f(2) = \frac{k-4}{10} = \frac{k-4}{10}$$

$$f(3) = \frac{k-9}{10} = \frac{k-9}{10}$$

$$f(4) = \frac{k-16}{10} = \frac{k-16}{10}$$

So $\frac{k}{10} + \frac{k-1}{10} + \frac{k-4}{10} + \frac{k-9}{10} + \frac{k-16}{10} = 1$

$$\frac{5k - 30}{10} = 1$$

$$k = 8$$

But $f(3) = -\frac{1}{10}$, $f(4) = -\frac{8}{10}$

Probabilities cannot be negative!

So k does not exist!

$\int 3x\,dx = 3\int x\,dx = 3\cdot\frac{x^2}{2} = \frac{3x^2}{2}$

Mathematical Expectation

Originally, the concept of mathematical expectation arose with respect to games of chance. If a player became involved in a game of chance, how much could that person expect to win (or more likely, lose) when playing the game? However, statisticians have considerably extended the meaning of expectation since then.

Consider a fair coin tossed 4 times. How many heads to you *expect* to see? Since heads should come up half of the time, you expect to see 2 heads in 4 tosses, since half of the tosses should be heads and half of the tosses should be tails. Does this mean that every time you toss a fair coin 4 times that you will observe 2 heads? Of course it doesn't, but what does it mean? It means that, over time with an extremely high number of repetitions of the experiment, you will *average* 2 heads in 4 tosses.

The **expected value** of the random variable X, denoted as $E(X)$, represents the mean of a probability distribution. In particular, for a discrete random variable, the ***mean*** can be determined by

$$\mu = E(X) = \sum_x \left[x \cdot f(x)\right]$$

In addition, the ***variance*** of a discrete random variable can be determined by

$$\sigma^2 = E\left[(x-\mu)^2\right] = \sum_x \left[(x-\mu)^2 \cdot f(x)\right]$$

We can further define the expected value of a function of the random variable X:

$$E(g(x)) = \sum_x \left[g(x) f(x)\right]$$

$E[aX] = aE[X]$

$E[2] = 2 \quad E[C] = C$

$\int x^2 dx \neq \int x\,dx \cdot \int x\,dx$

$E[x^2] \neq E[x] \cdot E[x]$

Example We have already determined the probability distribution for tossing a fair coin four times. Use this information to determine the expected number of heads observed.

$X \equiv$ number of heads $(x = 0, 1, 2, 3, 4)$

X	$P_r(X)$
0	1/16
1	6/16
2	4/16
3	4/16
4	1/16

$E[x^2] = \sum_x x^2 P_r x$

$= \frac{4 + 24 + 36 + 16}{16} = \frac{80}{16} = 5$

TTTT — 1/16

HTTT, THTT, TTHT, TTTH — 4/16

THHH, HTHH, HHTH, HHHT — 4/16

HHTT, THHT, TTHH, HTHT, THTH, HTTH — 6/16

HHHH — 1/16

Example Determine the variance in the number of heads observed in 4 tosses of a fair coin.

Theorem If a and b are constants, $E(aX+b)=aE(X)+b$.

Proof *Using the definition of expectation of a discrete random variable, we find that:*

$$E(aX+b)=\sum_x (ax+b)f(x)=\sum_x axf(x)+\sum_x bf(x)=a\underbrace{\sum_x xf(x)}_{=E(X)}+b\underbrace{\sum_x f(x)}_{=1}=aE(X)+b$$

Notice that as corollaries, we have shown the following when a and b are constants and X is a discrete random variable:

Corollary 1 $E(aX)=aE(X)$

Corollary 2 $E(b)=b$

Example Use the results of the previous theorem to prove the following with respect to the variance:

$$\sigma^2=E\left[(x-\mu)^2\right]=\sum_x\left[(x-\mu)^2\cdot f(x)\right]=E(X^2)-\mu^2$$

Example Use this alternate formula for the variance to determine the variance in the number of heads observed in 4 tosses of a fair coin.

Example A box containing 50 memory chips for a computer has 5 of those chips being defective. If 5 chips are chosen at random, how many defective chips can be expected to be chosen? What is the variance in the number of defective chips selected?

$\binom{50}{5}$ probability $= \dfrac{\binom{5}{x}\binom{45}{5-x}}{\binom{50}{5}}$

$\binom{5}{x}\binom{45}{5-x}$

$\mu = E[x] = \frac{1}{2}$

$\sigma^2 = E(x^2) - \mu^2$

$E[x]^2 = \sigma^2 = \frac{65}{98} - \left(\frac{1}{2}\right)^2 = \frac{81}{196} = \sqrt{\frac{81}{196}} = \frac{9}{14}$

Definition The expected value of the random variable function $g(X)$, where X has probability function $f(x)$ is defined to be

$$E[g(X)] = \sum_x g(x)f(x)$$

over all possible values of X if X is a discrete random variable

Example Let X be a random variable with the following probability distribution:

x	-3	6	9
$f(x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Find the mean and variance of X and the mean and variance of $g(X) = (2X+1)^2$.

Find μ and σ^2

$\mu = E[x] = (-3)(\frac{1}{6}) + 6(\frac{1}{2}) + 9(\frac{1}{3})$

$E[x]^2 = 9[\frac{1}{6}] + 36(\frac{1}{2}) + 81(\frac{1}{3}) - (\frac{11}{2})^2 = \frac{65}{4}$

Let $g(x) = 2x+1$

$g(x)$	x	$Pr(x)$
25	-3	$\frac{1}{6}$
169	6	$\frac{1}{2}$
361	9	$\frac{1}{3}$

$E[g(x)] = 25(\frac{1}{6}) + 169(\frac{1}{2}) + 361(\frac{1}{3}) = 209$

$\sigma^2 = 25^2(\frac{1}{6}) + (169)^2(\frac{1}{2}) + (361)^2(\frac{1}{3}) - (209)^2 = 14144$

Binomial Distribution and Binomial Probabilities

The Binomial Distribution

The binomial distribution is one of the two specific discrete distributions we will study. The distinguishing characteristic about an occurrence that might be modeled by the binomial distribution is that it allows for only 2 possible outcomes. These outcomes are generally labeled as "success" and "failure." Examples of events for which the binomial distribution might be appropriate include the following:

1. Determining the probability that at least 40 heads will occur in the next 75 tosses of a fair (or even of a biased) coin.
2. Determining the probability that in a class of 40 kindergarten children no more than 6 will contract chicken pox during the school year.
3. Determining the probability that a new medical test will be able to detect the presence of a disease in at least 25 of 30 people known to have the disease.

Each performance of the experiment is known as a **trial**. For a binomial experiment, these trials have the following characteristics:

1. Each trial can result in only 2 possible outcomes: "success" and "failure."
2. The probability of a success remains constant from trial to trial.
3. The number of trials is a finite constant.
4. The trials are independent. That is, the outcome of any trial has no effect on the outcome of any other trial.

Trials in which all of the above characteristics are true are known as **Bernoulli trials**.

Consider tossing a coin 4 times. With each toss, the side facing up (head or tail) is observed and recorded. How many ways can a head show up twice? Realizing that the tosses are independent, and that there is an order to the occurrence of the heads, we can list all of the outcomes of 4 tosses in which exactly 2 of them are heads. (We are willing to construct this list because the list is relatively short.) These outcomes are:

HHTT HTHT HTTH THHT TTHH THTH

If the coin were tossed 100 times, listing all of the outcomes satisfying the requirement of the number of occurrences of n heads, $n \le 100$, would be impractical due to the amount of time and patience that would be needed. What is needed is a formula that would indicate the number of ways k successes can occur in n trials. This number is known as a **binomial coefficient**. The formula representing the binomial coefficient, that is, the formula indicating the number of ways k successes can occur in n trials, is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots(2)(1)}$$

Binomial Parameters

The binomial distribution applies to populations. Therefore, instead of using the sample statistics $\bar{x}$ and s, we use the population parameters μ and σ. (Notice that Greek letters are used for population parameters while English letters are used for sample statistics.)

A binomial experiment in which n independent trials are performed on a random variable having probability of occurrence p has **mean**

$$\mu = np$$

and **standard deviation**

$$\sigma = \sqrt{np(1-p)}$$

The mean represents how many times one can expect the event having probability p to occur in the n trials. (In fact, another term for the population mean is the **expected value**.)

Example Suppose the probability that a diabetic man between the ages of 50 and 65 is overweight is 0.35. How many men in a group of 100 diabetic men in the requisite age group can be expected to be overweight? What is the variance of the number of overweight diabetic men in this group?

Example A biased coin $\left[\Pr(head) = 0.7\right]$ is tossed 21 times and the number of heads observed is recorded. Determine the expected number of observed heads and the standard deviation in the number of heads observed.

Binomial Probabilities

Suppose that a binomial random variable has probability of success equal to p, where $0 \le p \le 1$. If we let S represent the occurrence of a success, then we are saying that $Pr(S) = p$. This implies that the probability of a failure (F) is $1 - p$ (since a binomial experiment can only result in either a success or a failure). Equivalently, we say that $Pr(F) = 1 - p$.

Suppose we wish to determine the probability of obtaining exactly k successes in n trials of a binomial random variable having probability of success p. We realize that if there are n trials and k of these trials result in successes, the other $n - k$ trials must result in failures. Therefore, the required probability must be the product of 3 related numbers: the binomial coefficient (indicating in how many different ways the k successes can occur in the n trials – realize that this also indicates in how many ways the $n - k$ failures can occur in the n trials), the probability of k successes, and the probability of $n - k$ failures. This probability is given below:

↓ prob. of k successes

$$\Pr(k \text{ successes in n trials}) \equiv \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

binomial coef. ↗ ↑ prob of n-k failures

Notation To indicate the probability of observing k successes in a binomial process having n trials with probability p of success per trial is

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

Justification for the Binomial Probability Formula

The binomial probability formula given above is not intuitively obvious for the vast majority of people. An example might prove beneficial.

Example A fair die is rolled 5 times. What is the probability that exactly 2 of the rolls result in the upper face showing the number 6?

We first list all of the ways 2 rolls of 6 can be observed in 5 rolls of a fair die. Let 6 represent the event that a 6 was rolled and let X represent the event that a 6 was not rolled. The list is:

66XXX	*6X6XX*	*6XX6X*	*6XXX6*	*X66XX*
XX66X	*XXX66*	*X6X6X*	*XX6X6*	*X6XX6*

Notice that there are 10 ways in which 2 rolls of 6 occur in 5 rolls of a die.

Since the die is assumed fair, we know that:

$$Pr(6) = \frac{1}{6} \qquad Pr(X) = \frac{5}{6}$$

where X represents the probability of rolling anything other than a 6. Now consider the probability of rolling any one of the listed rolls. For example, suppose we look at the outcome listed as 66XXX. We must assume that the rolls are independent. That is, the result of the first roll has no effect on the result obtained on the second roll, etc. Then, the probability of getting these 5 rolls in the indicated order is

$$Pr(66XXX) = Pr(6)Pr(6)Pr(X)Pr(X)Pr(X) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$$

Analogously, we find that the probability for the 5-roll order 6X6XX is

$$Pr(6X6XX) = Pr(6)Pr(X)Pr(6)Pr(X)Pr(X) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$$

It should be no surprise that each of the 10 given orders in which 2 of the 5 rolls result in observing a 6 has the same probability. Therefore, the probability of rolling 2 6s in 5 rolls is

$$Pr(2 \text{ rolls of 6 in 5 total rolls}) = 10\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$$

However, recall that the binomial coefficient $\binom{n}{x}$ *is interpreted as the number of ways x successes can be obtained in n trials. For the die-rolling experiment, a "success" means obtaining a 6.*

So we need to know how many ways we can obtain 2 rolls of 6 in 5 rolls of the die. (Each of these ways has the same probability of occurrence.)

Therefore, the probability of obtaining 2 rolls of 6 out of 5 rolls of the die is

$$Pr(X = 2) = \underbrace{\binom{5}{2}}_{\substack{number\\of\ ways}} \underbrace{\left(\frac{1}{6}\right)^2}_{\substack{prob.of\\success}} \underbrace{\left(\frac{5}{6}\right)^3}_{\substack{prob.of\\failure}}$$

Generalizing, we should find it reasonable to believe that the probability of obtaining x successes out of n trials, where each success has a probability of p and each failure has a probability of 1 – p is given by

$$Pr(X = k) = \underbrace{\binom{n}{k}}_{\substack{number\\of\ ways}} p^k \underbrace{(1-p)}_{\substack{prob.of\\failure}}^{\overbrace{n-k}^{\#\ fail}}$$

Example Suppose it is known that in a given subpopulation, 40% of that subpopulation has O+ blood. What is the probability that in a group of 20 of these people:

a. exactly 10 have type O+ blood?

b. at least 10 have type O+ blood?

c. either 8, 9, or 10 have type O+ blood?

Using the TI-89 to Obtain Binomial Probabilities

You must have Operating System 2.03 or higher. (To see if you have this operating system, press **F1**, then the **down arrow ▼** (or **up arrow ▲** if it will work on your calculator) until you come to **A: About**. Check Version number on screen.) You must also have the **Stats/List Editor** that contains the TI statistics package. If you don't have these, they can either be downloaded from the TI website (you need a special cable) or transferred from a calculator already containing this software.

To access the binomial probability tables, press **APPS** ⇒ **1: FlashApps** ⇒ **ENTER** ⇒ **Stats/List Editor** ⇒ **ENTER** ⇒ **ENTER** (assuming **main** folder is selected) ⇒ **ENTER** ⇒ **F5** (Distributions) ⇒ **▼** (down arrow) ⇒ **B: Binomial Pdf** (for probabilities with respect to one specific number of successes) or **C: Binomial Cdf** (for cumulative probabilities – that is, the probability of some range of successes).

Next, complete the table that appears on the screen.

- **Binomial Pdf**: In the top box, enter the number of trials. In the middle box, enter the probability p of a success. In the bottom box, enter the number of successes. Press **ENTER** (until the screen changes). The desired probability appears near the top of the box labeled **Binomial Pdf**. This probability follows the script **Pdf** and is given to 9 decimal places. Notice that the input values are also given.

- **Binomial Cdf:** In the top box, enter the number of trials. In the second box, enter the probability p of a success. In the third box, enter the minimum number of successes allowed. In the bottom box, enter the maximum number of successes allowed. Press **ENTER** (until the screen changes). The desired probability appears near the top of the box labeled **Binomial Cdf**. This probability follows the script **Cdf** and is given to 9 decimal places. Notice that the input values are also given.

You can use the TI-89 to verify the probabilities obtained in the previous examples.

Example Suppose a biased coin $\left[Pr(H) = 0.55\right]$ is tossed 15 times. Determine the following probabilities:

a. The probability of obtaining exactly 6 heads.

b. The probability of obtaining at most 6 heads.

c. The probability of obtaining more that 8 heads.

Example A biased pair of dice $\left[Pr(2)=0.1\right]$ is rolled 100 times. What is the probability that the number of rolls totaling 2 is within 2 standard deviations of the mean?

Example Consider the same pair of dice as in the previous example. Suppose that the probability that a roll of seven occurs with these dice is 0.25. What is the probability that in 63 rolls of these dice the number of rolls totaling 7 is within 1.5 standard deviations of the mean?

Poisson Distribution and Probabilities

The Poisson Distribution

A **Poisson process** is a process in which events are counted within some interval, such as time or area. For example, phone calls per hour or trees per acre. (The word "per" is often an indication of the applicability of a Poisson process.) Clearly, only an integer number of counts can occur. Implicit to being a Poisson process is that the number of counts (events) within an interval is relatively small.

The only parameter upon which the Poisson process depends is the mean number μ of counts per interval. Interestingly, the mean and variance of the Poisson distribution are identical. That is, for a Poisson process, $\sigma^2 = \mu$. Therefore, one of the indications of the presence of a Poisson process is a discrete distribution where the process variance is approximately the same as the process mean.

The formula for determining the probability of the occurrence of k events where the assumption of a Poisson process, with mean μ, is appropriate is

$$Pr\left(X = k\right) = \frac{e^{-\mu}\mu^k}{k!} \quad , \quad k = 0, 1, 2, \cdots$$

In the above formula, μ represents the average number of events observed per interval. Similarly, the probability of observing at least **a** events but no more than **b** events is

$$Pr\left(a \le x \le b\right) = \sum_{k=a}^{b} \frac{e^{-\mu}\mu^k}{k!}$$

Because of the unique association between the mean and variance of a Poisson distribution (recall that they are equal), a process can usually be modeled with a Poisson distribution whenever the sample mean and sample variance are approximately equal.

Using the TI-89 to Calculate Poisson Probabilities

To access the Poisson probability tables, press **APPS** ⇒ **1: FlashApps** ⇒ **ENTER** ⇒ **Stats/List Editor** ⇒ **ENTER** ⇒ **ENTER** (assuming **main** folder is selected) ⇒ **ENTER** ⇒ **F5** (Distributions) ⇒ ▼ (down arrow) ⇒ **D: Poisson Pdf** (for probabilities with respect to one specific number of successes) or **E: Poisson Cdf** (for cumulative probabilities – that is, the probability of some range of successes).

Next fill out the table that appears on the screen.

- **Poisson Pdf**: In the top box, enter the mean of the distribution. (Although this box asks for λ, many texts and applications use this symbol to represent the mean of the Poisson distribution.) In the bottom box, enter the number of occurrences you seek. Press **ENTER** (until the screen changes). The desired probability appears near the top of the box labeled **Poisson Pdf**. This probability follows the script **Pdf** and is given to 9 decimal places. Notice that the input values are also given.

- **Poisson Cdf:** In the top box, enter the mean of the distribution. In the middle box, enter the minimum number of occurrences allowed. In the bottom box, enter the maximum number of occurrences allowed. (If the upper value is infinity, either leave the box blank (infinity is the default for the upper limit), or directly enter ∞.) Press **ENTER** (until the screen changes). The desired probability appears near the top of the box labeled **Poisson Cdf**. This probability follows the script **Cdf** and is given to 9 decimal places. Notice that the input values are also given.

Example The Mount Palomar telescope detects an average of 5 comets per night Assume that the number of comets seen per night by this telescope follows a Poisson distribution.

a. How likely is it that the telescope will detect fewer comets tonight that it typically does?

b. How likely is it that the telescope will detect exactly 5 comets tomorrow night?

c. How likely is it that the telescope will detect at least 10 comets tomorrow night?

d. What is the probability that the number of comets detected tomorrow night will fall within 1.5 standard deviations of the mean?

Example The folks at the CCSD help desk have noticed that they typically receive 6 calls per hour during the academic day. Assuming a Poisson process, determine the following probabilities:

a. The next hour is a typical hour with respect to the number of phone requests they will receive.

b. The next hour will have fewer phone requests than expected.

c. The next hour will have at least 7 phone requests.

d. In excess of how many requests is the probability less than 5%?

Example Suppose the following data represent the number of deaths attributable to AIDS in a Nebraska county over the last 10 years:

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
AIDS-related deaths	22	17	3	9	5	8	14	5	4	7

Can a Poisson distribution be used to model this data set? Justify your decision statistically.

Poisson Approximation to the Binomial Distribution

Consider the following example:

Approximately 4 of every 250,000 meteorites found by astronomers with Earth-based telescopes have the potential to strike Earth and do significant damage. Of the next 1,000,000 meteorites found, what is the probability that at least 25 will have the potential to strike Earth and do significant damage?

This is clearly a binomial process. (Either the meteorite has the appropriate potential or it does not.) Therefore, the exact probability is

$$Pr(X \ge 25) = \sum_{k=25}^{1000000} \binom{1000000}{k} (0.000016)^k (0.999984)^{1000000-k}$$

Because of the large number of trials $(n = 1,000,000)$, calculation of this probability can be difficult. (The large number of trials will not appear in any published tables, and many calculators will not be able to evaluate large factorials such as would be needed here. Even the TI-89 gives a impossible probability.) By using EXCEL, we find the exact probability to be

$$Pr(X \ge 25) = \sum_{k=25}^{1000000} \binom{1000000}{k} (0.000016)^k (0.999984)^{1000000-k} = 0.0223$$

This presents the opportunity to demonstrate another use for the Poisson distribution:

> *The Poisson distribution can be used to approximate the binomial distribution when the number of trials in a binomial experiment is large and the probability of a success is small.*

Recall that the mean of the binomial distribution is $\mu = np$ and the variance is $\sigma^2 = np(1-p)$. If p is small (i.e., if the probability of a success is small), then the probability of a failure is large, since a failure has probability $1-p$. Therefore,

$$\sigma^2 = np\underbrace{(1-p)}_{\approx 1} \approx np$$

So, when the probability p of a success is small, the above observation shows that the variance is approximately equal to the mean. Whenever the mean of a discrete distribution is approximately equal to the variance of the distribution, the Poisson distribution will usually model the original discrete distribution very well. As a general guideline, the Poisson distribution can be used as an approximating distribution to the binomial distribution when $n \ge 100$ and $p \le 0.01$.

Example Verify that the Poisson distribution can be used to approximate the binomial distribution in the previous example, and then use the Poisson distribution to approximate the required probability.

Example Suppose a rare genetic defect presents in only 1 out of every 500,000 people. Assuming a sample size of 1,000,000, determine the following probabilities:

a. What is the probability that between 5 and 8 (inclusive) people have this genetic defect?

b. What is the probability that at least 8 people carry this gene?