

## Unit 4

### The Bernoulli and Binomial Distributions

*“If you believe in miracles, head for the Keno lounge”*

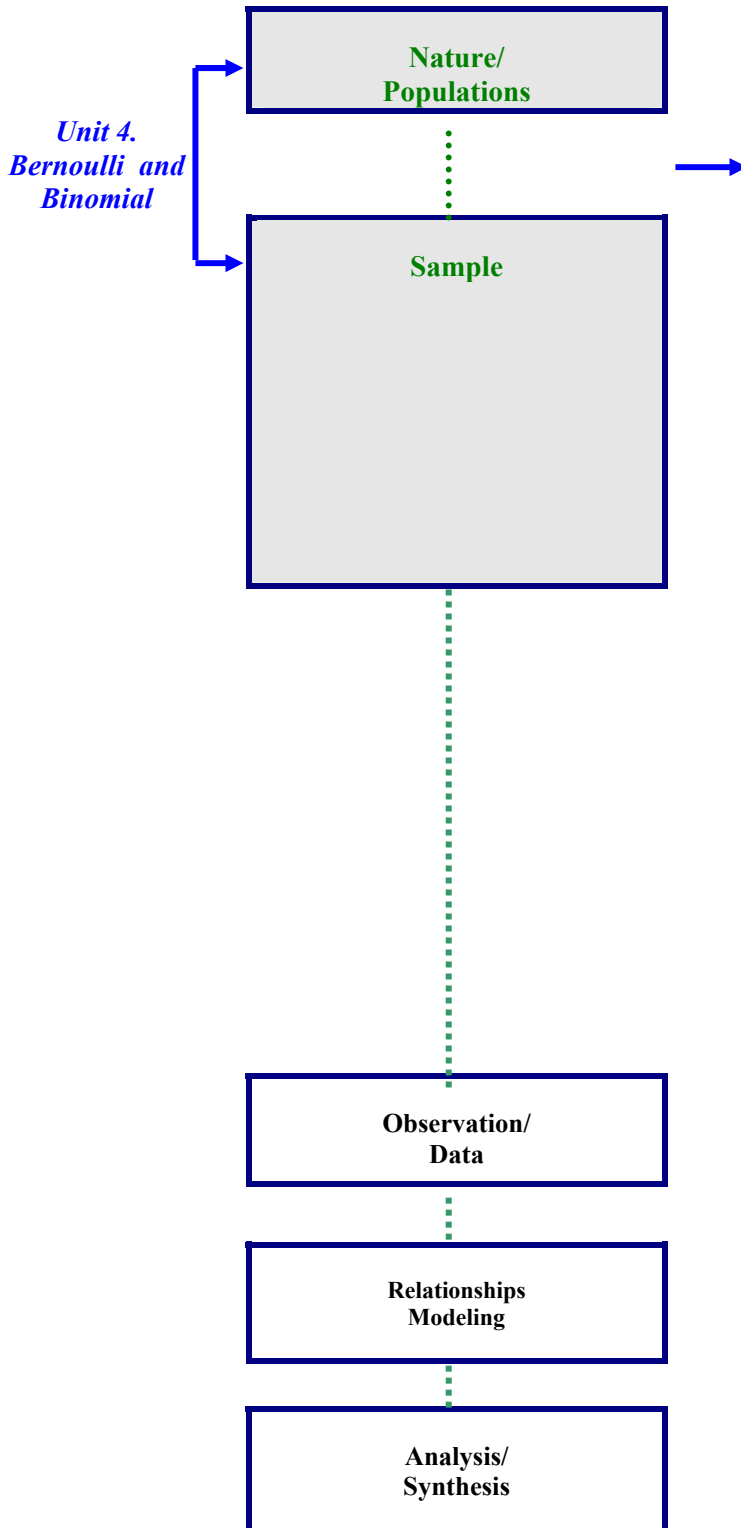
*- Jimmy the Greek*

The Amherst Regional High School provides flu vaccinations to a random sample of 200 students. How many will develop the flu? A new treatment for stage IV melanoma is given to 75 cases. How many will survive two or more years? In a sample of 300 cases of uterine cancer, how many have a history of IUD use? The number of “events” in each of these scenarios is a random variable that is modeled well using a Binomial probability distribution. When the number of trials is just one, the probability model is called a Bernoulli trial.

***The Bernoulli and Binomial probability distributions are used to describe the chance occurrence of “success/failure” outcomes. They are also the basis of logistic regression which is used to identify the possibly multiple predictors of “success/failure” outcomes.***



# 1. Unit Roadmap



**Data are the measurements of our observations of nature. This unit focuses on nominal data that are binary.** Previously, we learned that data can be of several types – nominal, ordinal, quantitative, etc. *A nominal variable* that is *binary* or *dichotomous* has exactly two possible values. **Examples are vital status (alive/dead), exposure (yes/no), tumor remission(yes/no), etc.**

The “frequentist” view of probability says that *probability is the relative frequency in an indefinitely large number of trials*. In this framework, a probability distribution model is a model of chance. It describes the way that probability is distributed among the possible values that a random variable can take on.

**The Bernoulli and Binomial probability distribution models** are often very good descriptions of patterns of occurrence of events that are of interest in public health; eg - mortality, disease, and exposure.



## 2. Learning Objectives

**When you have finished this unit, you should be able to:**

- Explain the “frequentist” approach to probability.
- Define a discrete probability distribution.
- Explain statistical expectation for a discrete random variable..
- Define the Bernoulli probability distribution model.
- Explain how to “count the # ways” using the tools of factorials and combinatorials.
- Define the Binomial probability distribution model.
- Calculate binomial probabilities.

### 3. Introduction to Discrete Probability Distributions

A **discrete** probability distribution is defined by (i) a listing of all the possible random variable values, together with (ii) their associated probabilities of occurrence.

- The listing of possible random variable outcomes must comprise **ALL** the possibilities (be exhaustive)
- Each possibility has a **likelihood** of occurrence (“chances of occurrence”) that is a number somewhere between 0 and 1.
- *Looking ahead* ... We’ll have to refine these notions when we come to speaking about **continuous** distributions because, in those situations, the number of possible outcomes is infinite!.

#### Example: Gender of a randomly selected student

- We’ll use **capital X** as our placeholder for the random variable name:

$X$  = Gender of randomly selected student from the population of students at a University

- We’ll use **small x** as our placeholder for a value of the random variable  $X$ :  
 $x = 0$  if gender of the selected student is male  
 $x = 1$  if gender of the selected student is female

Value of the Random Variable X is x =	Probability that X has value x is $\Pr [ X = x ] =$
<p>0 = male 1 = female</p> <p style="text-align: center;">↑</p> <p><i>Note that this roster exhausts all possibilities.</i></p>	<p>0.53 0.47</p> <p style="text-align: center;">↑</p> <p><i>Note that the sum of these individual probabilities, because the sum is taken over all possibilities, is 100% or 1.00.</i></p>

**Some useful terminology -**

- For discrete random variables, a probability model is the set of assumptions used to assign probabilities to each outcome in the sample space.  
  
The sample space is the universe, or collection, of all possible outcomes.
- A probability distribution defines the relationship between the outcomes and their likelihood of occurrence.
- To define a probability distribution, we make an assumption (the probability model) and use this to assign likelihoods.



## 4. Statistical Expectation

Statistical expectation was introduced in Appendix 2 of Unit 2 *Introduction to Probability*, pp 54-55.

**A variety of wordings might provide a clearer feel for statistical expectation.**

- **Statistical expectation** is the “**long range average**”. Think of this as “what you can expect in the long run”. The statistical expectation of what the state of Massachusetts will pay out is the long range average of the payouts taken over all possible individual payouts.
- **Statistical expectation** represents an “**on balance**”, even if “**on balance**” is not actually possible.

**IF**

\$1 has a probability of occurrence = 0.50  
 \$5 has a probability of occurrence = 0.25  
 \$10 has a probability of occurrence = 0.15 and  
 \$25 has a probability of occurrence = 0.10

**THEN** in the long run, or “**on balance**”, the expected winning is \$5.75 because

$$\$5.75 = [\$1](0.50) + [\$5](0.25) + [\$10](0.15) + [\$25](0.10)$$

*Notice that the “**on balance**” dollar amount of \$5.75 is not an actual possible winning*

**In the long run, what can the State of Massachusetts expect to pay out on average?**

The answer is a weighted sum of the possible winnings, where the “weights” are the associated “chances of occurrence”.

[ **Statistical expectation** = \$5.75 ]

$$= [\text{\$1 winning}] (\text{percent of the time this winning occurs} = 0.50) +$$

$$[\text{\$5 winning}] (\text{percent of the time this winning occurs} = 0.25) +$$

$$[\text{\$10 winning}] (\text{percent of the time this winning occurs} = 0.15) +$$

$$[\text{\$25 winning}] (\text{percent of the time this winning occurs} = 0.10)$$

You can replace the word **statistical expectation** with *net result, long range average, in the long run, or, on balance.*

**Statistical Expectation**  
**Discrete Random Variable X**

For a discrete random variable X (e.g. winning in lottery)  
Having probability distribution as follows:

<u>Value of X, x =</u>	<u>P[X = x] =</u>
\$ 1	0.50
\$ 5	0.25
\$10	0.15
\$25	0.10

The statistical expectation of the random variable X is written as  **$E[X] = \mu$** . When X is discrete, it is calculated as the weighted sum of all the possible values x, using weights equal to associated probabilities of occurrence  $\text{Pr}[X=x]$

$$E[X] = \mu = \sum_{\text{all possible } X=x} [x]P(X = x)$$

In the “likely winnings” example,  $\mu = \$5.75$

**We can calculate the statistical expectation of other things, too.**

**Example** – Suppose we want to know how much we can expect to win or lose, by taking into account the cost of the purchase of the lottery ticket.

Suppose a lottery ticket costs \$15 to purchase

We can expect to win a -\$9.25. Put another way, we can expect to lose \$9.25. Here’s how it works.

[ **Statistical expectation of amount won = -\$9.25** ]

$$= [\$1 \text{ winning} - \$15 \text{ cost}] (\text{percent of the time this winning occurs} = 0.50) +$$

$$[\$5 \text{ winning} - \$15 \text{ cost}] (\text{percent of the time this winning occurs} = 0.25) +$$

$$[\$10 \text{ winning} - \$15 \text{ cost}] (\text{percent of the time this winning occurs} = 0.15) +$$

$$[\$25 \text{ winning}] - \$15 \text{ cost} ](\text{percent of the time this winning occurs} = 0.10)$$

**Statistical Expectation**  
**Discrete Random Variable  $Y = [X-15]$**

<u>Value of Y, y =</u>	<u>P[Y=y] =</u>
\$ 1 - \$15 = -\$14	0.50
\$ 5 - \$15 = -\$10	0.25
\$10 - \$15 = -\$5	0.15
\$25 - \$15 = +\$10	0.10

The realization of the loss random variable Y has *statistical expectation*  $E[Y]=\mu_Y$

$$\mu_Y = \sum_{\text{all possible } Y=y} [y] P(Y=y) = - \$9.25$$

## 5. The Population Variance is a Statistical Expectation

### Example, continued - One play of the Massachusetts State Lottery.

- **The random variable X** is the “winnings”. X possible values x=\$1, \$5, \$10, and \$25.
- **The statistical expectation of X is  $\mu = \$5.75$** . Recall that this figure is what the state of Massachusetts can expect to pay out, on average, in the long run.
- **What about the variability in X?** In learning about population variance  $\sigma^2$  for the first time, we understood this to be a measure of the variability of individual values in a population.

The population variance  $\sigma^2$  of a random variable X is the statistical expectation of the quantity  $[X - \mu]^2$

**Discrete Random Variables**  
**Variance  $\sigma^2 =$  Statistical Expectation of  $[X-\mu]^2$**   
 **$= E[X-\mu]^2$**

For a discrete random variable X (e.g. winning in lottery)  
 Having probability distribution as follows:

<u>Value of <math>[X-\mu]^2 =</math></u>	<u><math>P[X = x] =</math></u>
$[1 - 5.75]^2 = 22.56$	<b>0.50</b>
$[5 - 5.75]^2 = 0.56$	<b>0.25</b>
$[10 - 5.75]^2 = 18.06$	<b>0.15</b>
$[25 - 5.75]^2 = 370.56$	<b>0.10</b>

The variance of a random variable X is the statistical expectation of the random variable  $[X-\mu]^2$  is written as  **$\text{Var}[X]=\sigma^2$**  When X is discrete, it is calculated as the weighted sum of all the possible values  $[x-\mu]^2$ , using weights equal to associated probabilities of occurrence  $\text{Pr}[X=x]$

$$\sigma^2 = E\left[(X-\mu)^2\right] = \sum_{\text{all possible } X=x} [(x-\mu)^2] P(X=x)$$

In the “likely winnings” example,  $\sigma^2 = 51.19$  dollars *squared*.

## 6. The Bernoulli Distribution

The Bernoulli Distribution is an example of a **discrete** probability distribution. It is an appropriate tool for the analysis of proportions and rates.

### Example – The fair coin toss.

- We'll use **capital Z** as our placeholder for the random variable name here:

$Z$  = Face of coin toss

- We'll use **small z** as our placeholder for a value of the random variable  $Z$ :

$z = 1$  if “heads”

$z = 0$  if “tails”

- We'll use  **$\pi$  and  $(1-\pi)$**  as our placeholder for the associated probabilities

$\pi = \text{Pr}[Z=1]$

eg – This is the probability of “heads” and is equal to .5 when the coin is fair

$(1-\pi) = \text{Pr}[Z=0]$

### Bernoulli Distribution ( $\pi$ ) (“Bernoulli Trial”)

A random variable  $Z$  is said to have a **Bernoulli Distribution** if it takes on the value 1 with probability  $\pi$  and takes on the value 0 with probability  $(1-\pi)$ .

Value of  $Z$  =

1

0

$P[Z = z] =$

$\pi$

$(1 - \pi)$

(1)  **$\mu$  = Mean** =  $E[Z]$  = Statistical Expectation of  $Z$

$\mu = \pi$

(2)  **$\sigma^2$  = Variance** =  $\text{Var}[Z] = E[(Z-\mu)^2]$  = Statistical Expectation of  $(Z-\mu)^2$

$\sigma^2 = \pi(1 - \pi)$

A Bernoulli Distribution is used to model the outcome of a SINGLE “event” trial

Eg – mortality, MI, etc.

Nature ——— Population/ Sample ——— Observation/ Data ——— Relationships/ Modeling ——— Analysis/ Synthesis

## Mean ( $\mu$ ) and Variance ( $\sigma^2$ ) of a Bernoulli Distribution

### Mean of $Z = \mu = \pi$

The mean of  $Z$  is represented as  $E[Z]$ .

$E[Z] = \pi$  because the following is true:

$$\begin{aligned} E[Z] &= \sum_{\text{All possible } z} [z] \text{Probability}[Z = z] \\ &= [0] \text{Pr}[Z=0] + [1] \text{Pr}[Z=1] \\ &= [0](1 - \pi) + [1](\pi) \\ &= \pi \end{aligned}$$

### Variance of $Z = \sigma^2 = (\pi)(1-\pi)$

The variance of  $Z$  is  $\text{Var}[Z] = E[(Z - (EZ))^2]$ .

$\text{Var}[Z] = \pi(1-\pi)$  because the following is true:

$$\begin{aligned} \text{Var}[Z] &= E[(Z - \pi)^2] = \sum_{\text{All possible } z} [(z - \pi)^2] \text{Probability}[Z = z] \\ &= [(0 - \pi)^2] \text{Pr}[Z = 0] + [(1 - \pi)^2] \text{Pr}[Z = 1] \\ &= [\pi^2](1 - \pi) + [(1 - \pi)^2](\pi) \\ &= \pi(1 - \pi)[\pi + (1 - \pi)] \\ &= \pi(1 - \pi) \end{aligned}$$

## 7. Introduction to Factorials and Combinatorials

### From 1 Trial to Many Trials -

When we do a SINGLE trial of event/non-event occurrence, this is a Bernoulli trial.

When we do SEVERAL trials of event/non-event occurrence, this is a Binomial random variable. We need to understand factorials and combinatorials in order to understand the Binomial distribution

### Preliminary – Introduction to the **factorial**

The “factorial” is just a shorthand that saves us from having to write out in longhand multiplications of the form (3)(2)(1) or (5)(4)(3)(2)(1) or (10)(9)(8)(7)(6)(5)(4)(3)(2)(1) ... well, you get the idea..

- **Notation**: “n factorial” is written **n!**
- **Definition**:  $n! = (n)(n-1)(n-2) \cdots (3)(2)(1)$
- **Example** -  $3! = (3)(2)(1) = 6$
- **Example** -  $8! = (8)(7)(6)(5)(4)(3)(2)(1) = 40,320$
- **Definition**:  $0! = 1$

**Factorial**

$$n! = (n)(n-1)(n-2) \dots (2)(1)$$

$$0! = 1 \quad \textit{by convention}$$

### Motivating the Combinatorial – Example

A box contains 1 red marble and nine green marbles. Five draws are made at random *with replacement*. In how many ways can you get 2 reds and 3 greens?

- One outcome that satisfies “2 reds and 3 greens” occurs when the first 2 draws are both red and the remaining three draws are all green:

R R G G G

- Another outcome that satisfies “2 reds and 3 greens” occurs when the first and last draws are red and the middle draws are green:

R G G G R

- So now - what is the total number of outcomes that satisfy “2 reds and 3 greens”? The answer is a **combinatorial**. Here, it is solved as “**five choose 2**” and is equal to:

$$\text{"5 choose 2" ways} = \binom{5}{2} = \frac{5!}{2! 3!} = \frac{(5)(4)(3)(2)(1)}{(2)(1)(3)(2)(1)} = 10$$

- Check: If you wanted to, you could check this result for yourself by “listing all the ways” to obtain 2 red and 3 green. I’ve given you 2 (“RRGGG” and “RGGGR”). There are 8 more.

**The Combinatorial**

- **Question:** *“How many ways can we choose “x” from “n?” Another wording of this question is – “what is the number of combinations of n items that can be formed by taking them x at a time?”*

- **Notation:** One notation for this is  ${}_n C_x$  . Another is  $\binom{n}{x}$

**Combinatorial**

*The number of ways to take x items from n without order is:*

$${}_n C_x = \binom{n}{x} = \frac{n!}{x! (n-x)!}$$

*Note -*

$$\binom{n}{x} = \binom{n}{n-x}$$

$$\binom{n}{0} = \binom{n}{n} = 1$$



## Binomial Formula

The **binomial formula** is the binomial distribution probability that you use to calculate a binomial probabilities of the form:

*What is the probability that  $n$  independent Bernoulli trials, each with probability of success  $= \pi$  yields a total of  $x$  events of “success”?*

The probability of obtaining exactly  $x$  events of success in  $n$  independent trials, each with the same probability of event success equal to  $\pi$ :

$$\Pr[X=x] = \binom{n}{x} \pi^x (1-\pi)^{n-x} = \left[ \frac{n!}{x! (n-x)!} \right] \pi^x (1-\pi)^{n-x}$$

**Another Look**  
**Binomial Distribution (n,  $\pi$ )**  
**X = sum of n independent Bernoulli( $\pi$ ) Trials Z**

The n **Bernoulli** trials are  $Z_1 Z_2 \dots Z_n$

- Each  $Z_i$  has possible values of 1 (“success”) or 0 (“failure”)
- $\Pr [ Z_i = 1 ] = \pi$  and  
 $\Pr [ Z_i = 0 ] = (1-\pi)$  for  $i=1, 2, \dots, n$

The **Binomial** random variable is  $X = Z_1 + Z_2 + \dots + Z_n$ . X is distributed Binomial(n,  $\pi$ )

$$X = \sum_{i=1}^{i=n} Z_i$$

For  $X \sim \text{Binomial}(n, \pi)$ , the probability that  $X = x$  is given by the binomial formula:

$$\text{Probability}[X=x] = \left[ \frac{n!}{x! (n-x)!} \right] \pi^x (1-\pi)^{n-x},$$

where X has possible values  $x = 0, 1, 2, \dots, n$

E [X] and the variance Var [X] is obtained by working with the  $Z_1 Z_2 \dots Z_n$

$$E [X] \text{ is actually is } E\left[ \sum_{i=1}^n Z_i \right] = n \pi$$

$$\text{Var} [X] \text{ is actually } \text{Var}\left[ \sum_{i=1}^n Z_i \right] = n \pi (1-\pi)$$

## 9. Calculation of Binomial Probabilities

A roulette wheel lands on each of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 with probability = .10. Write down the expression for the calculation of the following.

#1. The probability of “5 or 6” exactly 3 times in 20 spins.

#2. The probability of “digit greater than 6” at most 3 times in 20 spins.

**Solution for #1.**

The “event” is an outcome of either “5” or “6”

Thus, Probability [event] =  $\pi = .20$

“20 spins” says that the number of trials is  $n = 20$

Thus, X is distributed Binomial( $n=20, \pi=.20$ )

$$\begin{aligned}\Pr[X = 3] &= \binom{20}{3} [.20]^3 [1-.20]^{20-3} \\ &= \binom{20}{3} [.20]^3 [.80]^{17} \\ &=.2054\end{aligned}$$

**Solution for #2.**

The “event” is an outcome of either “7” or “8” or “9”

Thus, Pr[event] =  $\pi = .30$

As before,  $n = 20$

Thus, X is distributed Binomial( $n=20, \pi=.30$ )

*Translation: “At most 3 times” is the same as saying “3 times or 2 times or 1 time or 0 times” which is the same as saying “less than or equal to 3 times”*

$$\begin{aligned}\Pr[X \leq 3] &= \Pr[X = 0] + \Pr[X = 1] + \Pr[X = 2] + \Pr[X = 3] \\ &= \sum_{x=0}^3 \left\{ \binom{20}{x} \right\} [.30]^x [.70]^{20-x} \\ &= \binom{20}{0} [.30]^0 [.70]^{20} + \binom{20}{1} [.30]^1 [.70]^{19} + \binom{20}{2} [.30]^2 [.70]^{18} + \binom{20}{3} [.30]^3 [.70]^{17} \\ &=.10709\end{aligned}$$

## 10. Resources for the Binomial Distribution

*Note - To link directly to these resources, visit the PubHlth540 course web site ([www-unix.oit.umass.edu/~biep540w](http://www-unix.oit.umass.edu/~biep540w)). From the welcome page, click on **BERNOULLI AND BINOMIAL DISTRIBUTIONS** at left.*

### Additional Reading

- A 2 page lecture on the Binomial Distribution from University of North Carolina.  
<http://www.unc.edu/~knhigho/econ70/lec7/lec7.htm>

### Calculation of Binomial Probabilities

- Vassar Stats Exact Probability Calculator  
<http://faculty.vassar.edu/lowry/binomialX.html>