Unit 4
The Bernoulli and Binomial Distributions

"If you believe in miracles, head for the Keno lounge"

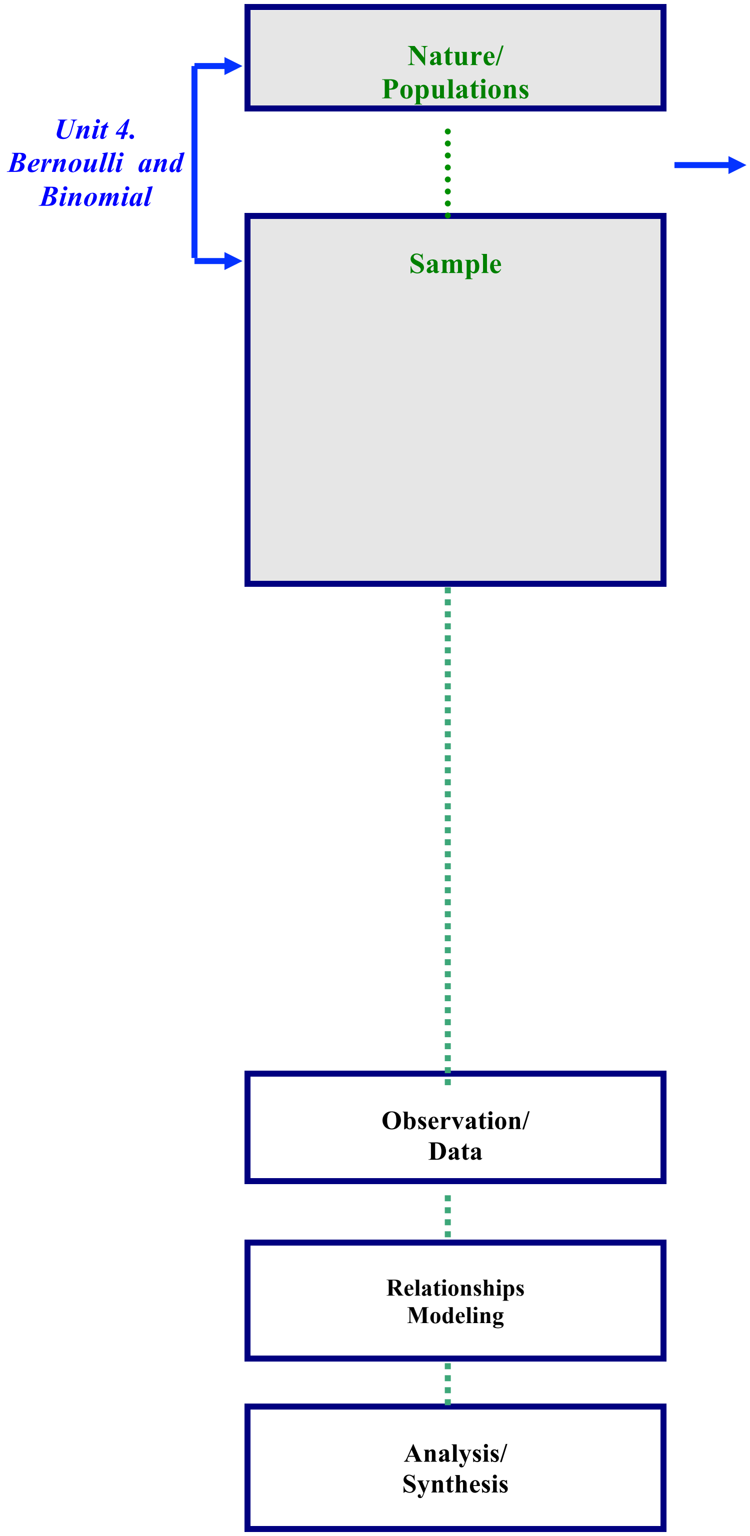
- Jimmy the Greek

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 model the chance occurrence of "success/failure" outcomes. They are also the basis of logistic regression which is used to explore (through modeling) the possibly multiple predictors of "success/failure" outcomes.

Table of Contents

1. Unit Roadmap



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

The <u>"frequentist"</u> view of probability says that ***probability is the relative frequency in an indefinitely large number of trials***. In this framework, a <u>probability distribution model</u> 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 models of patterns of occurrence of events that are of interest in public health; eg - mortality, disease, and exposure.

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

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

Before we begin - Here is a little recap of some key ideas.

A **variable** is a quantity (or characteristic) that may vary from object to object ***(typically, the object is a person on whom we are making an observation)***.

A **random variable** denotes a quantity (or characteristic) that varies from object to object according to some sort of random process. ***Chance!***

Remember:

- We represent the **name** of the random variable with **capital letters** (X, Y, etc)
- We represent the **values** of the random variable with **lower case letters** (x, y, etc.)

A **discrete random variable** is one for which there is only a ***finite*** set of possible values (***"finite" is nice, because it means that, in principle, we can list them out)***.

A **probability distribution** is a ***model*** that links the possible values of a random variable with the likelihood (chances) of their occurrence. The model might take the form of a **table** or an **equation**. (***tables versus equations will be clear in the examples that follow)***

A discrete probability distribution is associated with a discrete random variable. To complete the definition of a discrete probability distribution, we need 2 things:

1st - a listing of all the possible random variable values (check that it is exhaustive) and
2nd - the listing of the associated likelihoods (probabilities) of occurrence (these will be either numbers between 0 and 1 or an equation that yields numbers 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, meaning that we can no longer make tidy listings of all possibilities.

Example of a Discrete Random Variable - 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

Example, continued - Probability Distribution of X= Gender of a randomly selected student

1st – Listing of all possible values Value of the Random Variable X is x =	2nd – Listing of associated likelihoods (probabilities) Probability function: Pr [X = x] =
0 = male **1 = female** ↑ ***Note that this roster exhausts all possibilities.***	**Pr [X = 0] = 0.53** **Pr [X = 1] = 0.47** ↑ ***Note that the sum of these individual probabilities, because the sum is taken over all possibilities, is 100% or 1.00.***

Some useful terminology -

1. 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 collection of all possible values of the random variable.

2. A ***probability distribution*** defines the relationship between the outcomes and their likelihood of occurrence.

3. ***Putting it all together …*** To ***define a probability distribution***, we make an assumption (the probability model) and use this to assign likelihoods.

4. Statistical Expectation for Discrete Random Variables

Statistical expectation was introduced in Appendix 2 of Unit 2 *Introduction to Probability*, pp 54-55. The following paragraph is excerpted from that appendix.

An Example to Motivate the Ideas:
Suppose you stop at a convenience store on your way home and play the lottery.

Suppose further that the specifics of the lottery game that you are playing are the following–

$1 is won with probability = 0.50
$5 is won with probability = 0.25
$10 is won with probability = 0.15
$25 is won with probability = 0.10 ***(wow, the chances are 100% that you'll win something!)***

Equipped with this information, you can calculate for yourself the average winning or, put another way, what the winning is likely to be in the long run.

Average winning ("what the winning is likely to be in the long run")

= [$1](probability of a $1 ticket)
+ [$5](probability of a $5 ticket)
+ [$10](probability of a $10 ticket)
+ [$25](probability of a $25 ticket)

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

= $5.75

$5.75? Problem? How can the average be $5.75 if the specifics of the lottery are such that the possible awards are in the amounts of $1, $5. $10, or $25 only? ANSWER: Averages do not have to be possible outcomes.

Other names for this intuition are

- ♣ Expected winnings
- ♣ "Long range average"
- ♣ Statistical expectation!

The calculated average, $5.75, is useful even if it does not correspond to a possible winning.
One way to appreciate this is to imagine that you are in the budget office of the State of Massachusetts. This lottery game was your brain-child.

The figure $5.75 represents what the State of Massachusetts can expect to have to pay out ***on average***. ***Still with me? At this point a bell might go off in your head, saying: well then, if the state hopes to make money then it had better hike up the cost to purchase a lottery ticket. Stay tuned.***

The statistical expectation ("long range average", "in the long run", "on balance") is a weighted sum of all the possible winnings, where each possible winning is given a "weight" that is equal to its "chances of occurrence" (likelihood, probability").

[Statistical expectation = $5.75]
= [$1 winning] (chances, or probability, of this winning =0.50) +
[$5 winning] (chances, or probability, of this winning =0.25) +
[$10 winning] (chances, or probability, of this winning = 0.15) +
[$25 winning]](chances, or probability, of this winning = 0.10)

We have what we need to define the statistical expectation of a discrete random variable. ***Note – we'll have to modify this slightly when we come to random variables that are continuous.***

Statistical Expectation
Discrete Random Variable X

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

Value of X, x =	P[X = x] =
$ 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]=μ . When X is discrete, it is equal to the weighted sum of all the possible values x, using weights equal to associated probabilities of occurrence Pr[X=x]

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

Example, continued - In the "likely winnings" example, μ = $5.75

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

Example, continued – Now suppose that, what we really want to know is about profit! That is, 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.

So, really, a lottery ticked purchaser can expect to win a -$9.25. Put another way, he or she can expect to lose $9.25. Here's how it works.

[Statistical expectation of amount won = -$9.25]

= [$1 winning - $15 cost] (percent of the time this winning occurs=0.50) +
[$5 winning - $15 cost] (percent of the time this winning occurs =0.25) +
[$10 winning - $15 cost] (percent of the time this winning occurs = 0.15) +
[$25 winning] - $15 cost](percent of the time this winning occurs = 0.10)

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

Value of Y, y =	**P[Y=y] =**
$ 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 $E[Y]=\mu_Y$*

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

Guessing you might not bother to buy a lottery ticket on your way home ….

5. The Variance of a Random Variable 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 μ = \$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 σ^2 for the first time, we understood this to be a measure of the variability of individual values in a population.

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

Discrete Random Variables
Variance σ^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:

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

The variance of a random variable X is the statistical expectation of the random variable $[X-\mu]^2$ is written as Var[X]=σ^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 Pr[X=x]

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

Example, continued - In the "likely winnings" example, σ^2 = 51.19 dollars *squared.*

6. The Bernoulli Distribution

The Bernoulli Distribution is an example of a discrete probability distribution. It is often the probability model that is used for the analysis of proportions and rates. In BIOSTATS 640, we'll learn about another distribution that is used for the analysis of rates, the poisson distribution.

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 **π and $(1-\pi)$** as our placeholder for the associated probabilities
  π = Pr[Z=1] **eg – This is the probability of "heads" and is equal to .5 when the coin is fair**
  $(1-\pi)$ = Pr[Z=0]

Bernoulli Distribution (π) ("Bernoulli Trial")

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

Value of Z =	P[Z = z] =
1	π
0	$(1-\pi)$

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

(2) σ^2 = Variance = 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.

Mean (μ) and Variance (σ^2) of a Bernoulli Distribution

Mean of Z = μ = π

μ = the mean of Z (the statistical expectation of Z) is represented as E[Z].

E[Z] = π because the following is true:

$$\mu = \text{E[Z]} = \sum_{\text{All possible z}} \text{[z]Probability[Z=z]}$$

$$= [0]\Pr[Z=0] + [1]\Pr[Z=1]$$

$$= [0](1-\pi) + [1](\pi)$$

$$= \pi$$

Variance of Z = σ^2 = (π)(1-π)

The variance of Z is Var[Z] = E[(Z – (EZ)^2].

Var[Z] = π(1-π) because the following is true:

$$\text{Var[Z]} = \text{E}[(Z-\pi)^2] = \sum_{\text{All possible z}} [(z-\pi)^2]\text{Probability[Z=z]}$$

$$= [(0-\pi)^2]\Pr[Z=0] + [(1-\pi)^2]\Pr[Z=1]$$

$$= [\pi^2](1-\pi) + [(1-\pi)^2](\pi)$$

$$= \pi(1-\pi)[\pi + (1-\pi)]$$

$$= \pi(1-\pi)$$

7. Introduction to Factorials and Combinatorials

Why are we doing this?

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.

Example -

Birthday party: 5 Guests, 2 Hats - ***who gets to wear a hat?***

A birthday party has **5 guests (n=5)**: Bill, Ed, Sarah, John, and Alice. There are just **2 hats (x=2)**.
From among 5 guests, how many ways are there to choose 2 guests to wear a hat? This is an example of:

- Sampling 2 items without replacement from a collection of 5
- "5 choose 2" which is written more generally as "n choose x"
  "5 choose 2" is written with the special notation $\binom{5}{2}$
- "n choose x" is written with the special notation $\binom{n}{x}$

The answer is: there are 10 ways (hat):

1	Bill	Ed	Sarah	John	Alice
2	Bill	Ed	Sarah	John	Alice
3	Bill	Ed	Sarah	John	Alice
4	Bill	Ed	Sarah	John	Alice
5	Bill	Ed	Sarah	John	Alice
6	Bill	Ed	Sarah	John	Alice
7	Bill	Ed	Sarah	John	Alice
8	Bill	Ed	Sarah	John	Alice
9	Bill	Ed	Sarah	John	Alice
10	Bill	Ed	Sarah	John	Alice

- **Is there a way we can obtain # ways = 10 without having to write them all out?**
- **Yes, and it makes use of 2 tools: factorial and combinatorial**

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) ⋯ (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) … (2)(1)

0! = 1 ***by convention***

Motivating the Combinatorial – Example, continued

A "combinatorial" is the name given to the solution to the question "how many ways can you choose x from n?" The solution itself makes use of factorials.

Five guests are at a birthday party. How many ways can you choose 2 to wear a hat?

- We've seen that one "way" that satisfies "2 hats and 3 non-hats" occurs when the first 2 guests (Bob and Ed) wear a hat and the remaining 3 do not wear a hat:

  BOB ED SARAH JOHN ALICE

- Another "way" that satisfies "2 hats and 3 non-hats" occurs when the last 2 guests (John and Alice) wear a hat and the first 3 do not wear a hat:

  BOB ED SARAH JOHN ALICE

- So now - what is the total number of outcomes that satisfy "2 hats and 3 non-hats"? 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:* All 10 are shown on the previous page.

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 $_nC_x$. Another is $\binom{n}{x}$**

Combinatorial

The number of ways to choose x items from n (without regard to order) is:

$$ {}_nC_x = \binom{n}{x} = \frac{n!}{x!\,(n-x)!} $$

Note -

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

Eureka! Choosing x is the same as leaving (n-x) behind.

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

8. The Binomial Distribution

When is the Binomial Distribution Used? -

The binomial distribution is used to answer questions of the form: "What is the probability that, in **n** independent success/failure trials with probability of success equal to **π**, the result is **x** events of success?"

What are n, π, and X in the Binomial Distribution?

n = number of independent trials **(eg – the number of coin tosses performed, n=20)**
π = Probability[individual trial yields "success" **(eg – probability[single coin lands heads] = ½)**
x = number of events of success that is obtained **(eg – x=12 "heads")**

Example (details on the solution below)-

What is the probability that, among n=100 vaccinated for flu, with subsequent probability of flu π = .04, that x=13 will suffer flu? Answer: Pr[X=13] for Binomial (n=100, π=.04) = .00014 using online calculator http://vassarstats.net/binomialX.html and setting n=100, k=x=13 a and p =π = .04

The solution for binomial distribution probabilities makes use of the **combinatorial and factorial tools**. Introduced on pp 13-16.

Binomial Distribution (n, π)
(n independent Bernoulli Trials)

A random variable X is said to follow a Binomial (n, π) distribution if it is the sum of n independent Bernoulli Distribution trials each with probability of "success" = π.

Value of X =	P[X = x] =
0	$(1-\pi)^n$
1	$n\pi(1-\pi)^{n-1}$
…	…
x	$\binom{n}{x}\pi^x(1-\pi)^{n-x}$
…	…
n	π^n

(1) μ = Mean = E[X] = Statistical Expectation of X
$\mu = n\pi$

(2) σ^2 = Variance = Var[X] = E[$(X-\mu)^2$] = Statistical Expectation of $(X-\mu)^2$
$\sigma^2 = n\pi(1-\pi)$

Binomial Probability Distribution
Formula for Calculating Probabilities

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

What is the probability that, among n independent Bernoulli trials, each with probability of success = π, x events of "success" occur?

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

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

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

The n Bernoulli trials are Z_1 Z_2 … Z_n

- **Each Z_i has possible values of 1 ("success") or 0 ("failure")**
- **Pr [$Z_i = 1$] = π and**
  Pr [$Z_i = 0$] = $(1-\pi)$ for i=1, 2, …, n

The Binomial random variable is $X = Z_1 + Z_2 + \ldots + Z_n$. X is distributed Binomial(n, π)

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

For X ~ Binomial (n, π), 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, …, n

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

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

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

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.

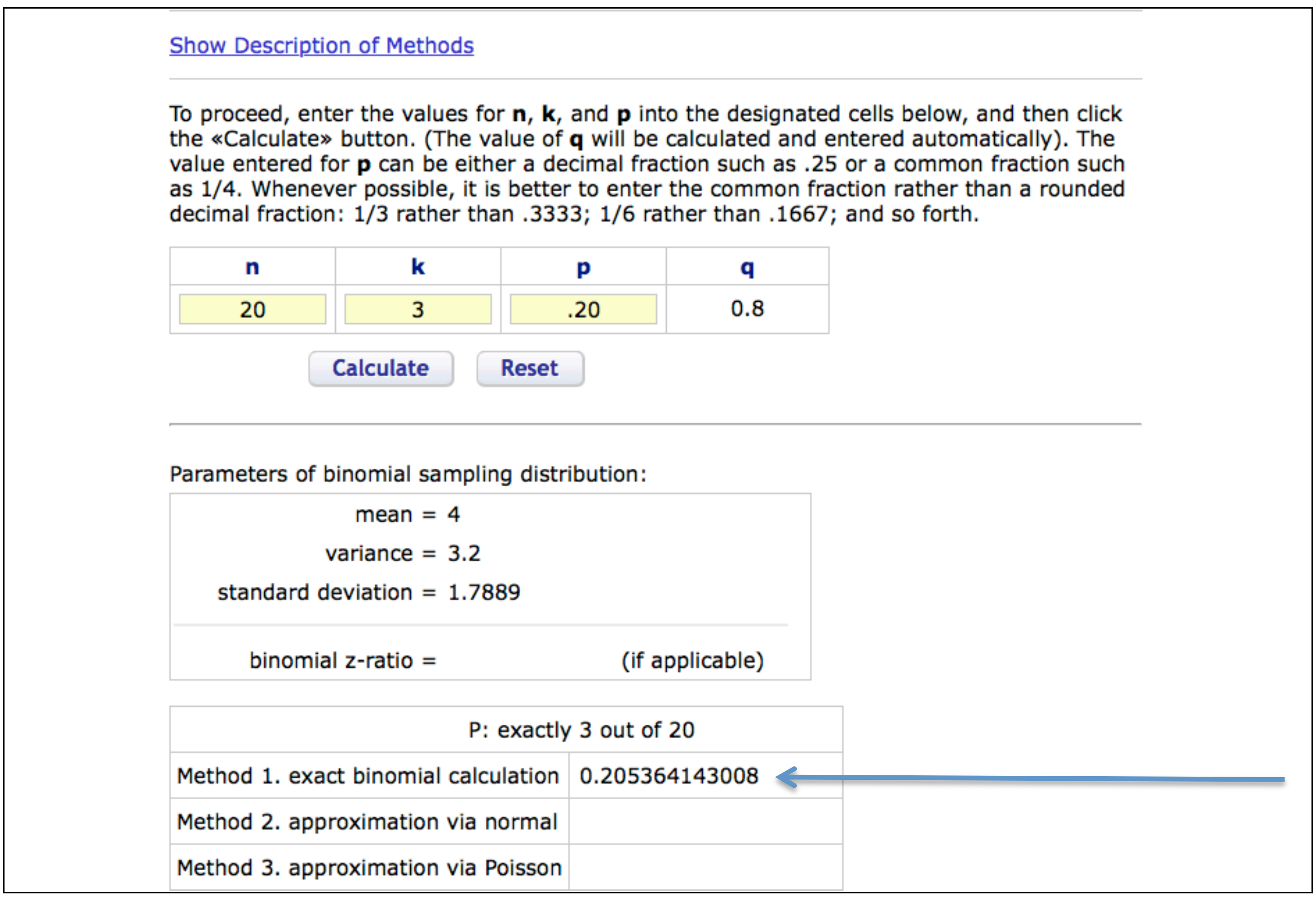
The "event" is an outcome of either "5" or "6"
Thus, Probability [event] = π = .20
"20 spins" says that the number of trials is n = 20
Thus, X is distributed Binomial(n=20, π=.20)

$$\Pr[X=3] = \binom{20}{3}[.20]^3[1-.20]^{20-3}$$

$$= \binom{20}{3}[.20]^3[.80]^{17}$$

$$= .2054$$

VassarStats Online Calculator for Pr[X=3]: Set n=20, k=3 and p=π=.20
http://vassarstats.net/binomialX.html **After entering settings, click CALCULATE.**


Show Description of Methods

To proceed, enter the values for **n**, **k**, and **p** into the designated cells below, and then click the «Calculate» button. (The value of **q** will be calculated and entered automatically). The value entered for **p** can be either a decimal fraction such as .25 or a common fraction such as 1/4. Whenever possible, it is better to enter the common fraction rather than a rounded decimal fraction: 1/3 rather than .3333; 1/6 rather than .1667; and so forth.

n	k	p	q
20	3	.20	0.8

Calculate Reset

Parameters of binomial sampling distribution:

mean =	4
variance =	3.2
standard deviation =	1.7889
binomial z-ratio =	(if applicable)

P: exactly 3 out of 20	
Method 1. exact binomial calculation	0.205364143008
Method 2. approximation via normal	
Method 3. approximation via Poisson	

Solution for #2.

The "event" is an outcome of either "7" or "8" or "9"
Thus, Pr[event] = π = .30
As before, n = 20
Thus, X is distributed Binomial(n=20, π=.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"***

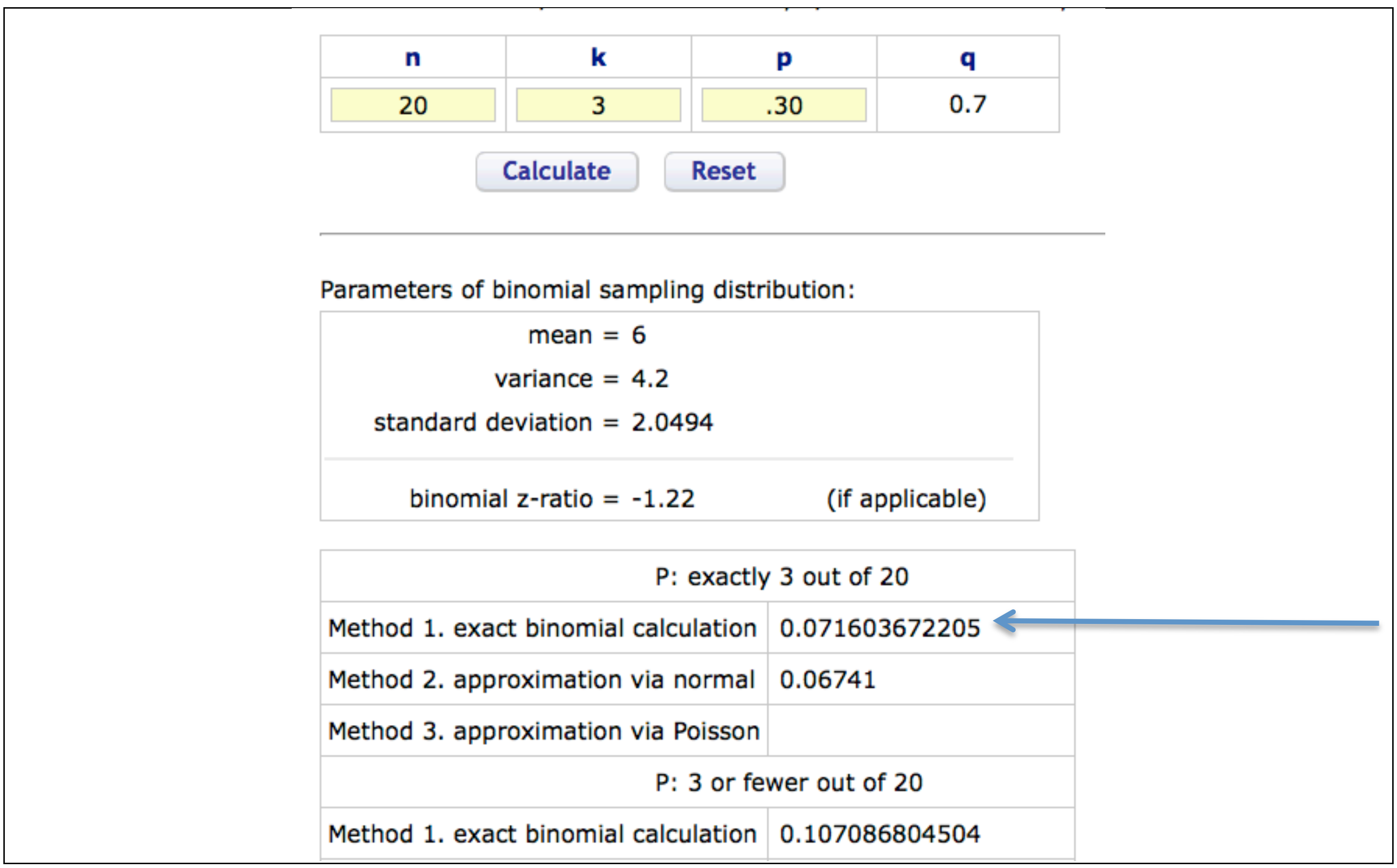
$$\Pr[X \le 3] = \Pr[X=0]+\Pr[X=1]+\Pr[X=2]+\Pr[X=3]$$

$$= \sum_{x=0}^{3} \left\{ \binom{20}{x} \right\} [.30]^x [.70]^{20-x}$$

$$= \binom{20}{0}[.30]^0[.70]^{20} + \binom{20}{1}[.30]^1[.70]^{19} + \binom{20}{2}[.30]^2[.70]^{18} + \binom{20}{3}[.30]^3[.70]^{17}$$

$$= .10709$$

VassarStats Online Calculator for Pr[X≤3]: Set n=20, k=3 and p=π=.30
http://vassarstats.net/binomialX.html **After entering settings, click CALCULATE**



n	k	p	q
20	3	.30	0.7

Calculate Reset

Parameters of binomial sampling distribution:

mean = 6
variance = 4.2
standard deviation = 2.0494

binomial z-ratio = -1.22 (if applicable)

P: exactly 3 out of 20	
Method 1. exact binomial calculation	0.071603672205
Method 2. approximation via normal	0.06741
Method 3. approximation via Poisson	
P: 3 or fewer out of 20	
Method 1. exact binomial calculation	0.107086804504