Stat 515: Introduction to Statistics

Chapter 4

Random Variable

- **Random Variable –** a numerical measurement of the outcome of a random phenomena
  - Capital letters refer to the random variable
  - Lower case letters refer to specific realizations

- Recall our definitions of Discrete and Continuous quantitative variables from before

Random Variable

- **Discrete Example:** Number of goals in an EPL soccer match
  - We refer to the number of goals in an EPL soccer match as X, until we have a concrete observation
  - x=2 goals is a realization – a concrete observation

Random Variable

- **Continuous Example:** Height of Americans
  - We refer to the Height of Americans as X, until we have a concrete observation
  - x=72 inches is a realization – a concrete observation

Discrete Distributions!

- **Probability Distribution –** a summary of all possible outcomes of a random phenomena along with their probabilities
  - **Example 1**: Number of goals scored in an EPL game
  - **Example 2&3:** Number of red lights on your way to work
  - **Example 4:** Number of free throws made

Random Variable: Discrete

- The possible outcomes must be countable
  - Remember quantitative discrete variables from before
- We have a **valid** discrete probability distribution if
  1. Our outcomes are discrete (countable)
  2. All the probabilities are valid
     - $0 \leq P(x) \leq 1 \; for \; all \; outcomes \; x$
  3. We've accounted for all possible outcomes
     - $\sum P(x) = 1$

Example 1: Discrete Distributions

- **Example**: number of goals scored in an EPL soccer match
- The number of goals is countable
- All probabilities are between 0 and 1
- $\sum P(x) = .0711 + .1974 + .2158 + .1842 + .1658 + .1026 + .0447 + .0105 + .0026 + .0053 = 1$

X = # of Goals	P(x) = Probability
0	.0711
1	.1974
2	.2158
3	.1842
4	.1658
5	.1026
6	.0447
7	.0105
8	.0026
9	.0053
TOTAL	1

Example 2: Discrete Distribution

- Example: Number of red lights on the way to work (there are only three red lights on your way to work – this means you can catch 0,1,2 or 3 lights on your way to work.)

X = Number of lights	P(x) = Probability
0	.10
1	.10
2	.10
3	.40

- The number of goals is countable
- All probabilities are between 0 and 1
- $\sum P(x) = .10 + .10 + .10 + .40 = .70$
- Since $\sum P(x) = .70 \neq 1$ we **do not** have a valid Discrete Dist.

Example 3 Discrete Distributions Route 2

- Example: Number of red lights on the way to work (there are only three red lights on your way to work – this means you can catch 0,1,2 or 3 lights on your way to work.)

X = Number of lights	P(x) = Probability
0	.40
1	.30
2	.20
3	.10

- The number of goals is countable
- All probabilities are between 0 and 1
- $\sum P(x) = .40 + .30 + .20 + .10 = 1$

Example 3 Discrete Distributions Route 2

- Example: Number of red lights on the way to work (there are only three red lights on your way to work – this means you can catch 0,1,2 or 3 lights on your way to work.)

X = Number of lights	P(x) = Probability
0	.20
1	.30
2	.10
3	.40

- The number of goals is countable
- All probabilities are between 0 and 1
- $\sum P(x) = .20 + .30 + .10 + .40 = 1$

Example 4: Discrete Distribution

- **Example:** Number of free throws made by a basketball player in 2 tries

X = Number Made	P(x) = Probability
0	.40
1	.40
2	.20

- The number of goals is countable
- All probabilities are between 0 and 1
- $\sum P(x) = .40 + .40 + .20 = 1$

The Mean of a Discrete Distribution

- The **mean** of a probability distribution represents the average of a large number of observed values. [Remember: in the long run]

- We denote this with the Greek letter as below

$$\mu_x = E(X) = Expected\ value\ of\ x = \sum xP(x)$$

Example 1: Discrete Distributions

- **Example**: # of goals scored in an EPL soccer match
- $\mu_x = E(x) = \sum x * P(x) = 0 + .1974 + .4316 + .5526 + .6632 + .5130 + .2682 + .0735 + .0208 + .0477 = 2.768$

X = # of Goals	P(X)	X*P(X)
0	.0711	0*.0711=0
1	.1974	1*.1974=.1974
2	.2158	2*.2158=.4316
3	.1842	3*.1842=.5526
4	.1658	4*.1658=.6632
5	.1026	5*.1026=.5130
6	.0447	6*.0447=.2682
7	.0105	7*.0105=.0735
8	.0026	8*.0026=.0208
9	.0053	9*.0053=.0477
TOTAL	1	2.768

Example 1: Discrete Distributions

- **Example**: # of goals scored in an EPL soccer match
- $\mu_x = E(x) = \sum x * P(x) = 0 + .1974 + .4316 + .5526 + .6632 + .5130 + .2682 + .0735 + .0208 + .0477 = 2.768$
- We like to write the interpretation in reasonable terms
  - "**On average, we expect** between two and three goals in an EPL soccer match"

Example 2 Discrete Distributions Comparing Routes: Route 1

X = Number of lights	P(X)	X*P(X)
0	.40	0*.40=0
1	.30	1*.30=.30
2	.20	2*.20=.40
3	.10	3*.10=.30

- $E(X) = \sum xP(x) = 0 + .3 + .4 + .3 = 1$
- **“On average, we expect** that Route 1 will result in hitting one red light”

Example 3 Discrete Distributions Comparing Routes: Route 2

X = Number of lights	P(X)	X*P(X)
0	.20	0*.20=0
1	.30	1*.30=.30
2	.10	2*.10=.20
3	.40	3*.40=1.20

- $E(X) = \sum xP(x) = 0 + .30 + .20 + 1.20 = 1.7$
- **"On average, we expect** that Route 2 will result in hitting between one and two red lights"

Example 2&3 Discrete Distributions Comparing Routes

- Route 1
  - $E(X) = \sum xP(x) = 1$
- Route 2
  - $E(X) = \sum xP(x) = 1.7$

- Route 2 will result in more lights **on average**

Example 4: Discrete Distribution

- **Example:** Number of free throws made by a basketball player in 2 tries

X = Number Made	P(x) = Probability	x*P(x)
0	.40	0*.40 = 0
1	.40	1*.40 = .40
2	.20	2*.20 = .40

- $\mu_x = E(x)$
$= \sum x * P(x) = 0 + .40 + .40 = .80$
- **"On average, we expect** between zero and one free throw in two tries"

The Variance of a Discrete Distribution

- The **variance** of a probability distribution represents the spread of observed values. It is calculated by finding the expected **squared distance from the mean**
- We denote this with the Greek letter as below

$$\sigma_x^2 = E[(X - \mu)^2] = \sum (x - \mu)^2 * P(x)$$

The Standard Deviation of a Discrete Distribution

- The **standard deviation** of a probability distribution represents the spread of observed values. It is calculated by finding the square root of the variance.

- We denote this with the Greek letter as below

$$\sigma_x = \sqrt{{\sigma_x}^2} = \sqrt{\sum (x-\mu)^2 * P(x)}$$

Example 1: Discrete Distributions

- **Example**: # of goals scored in an EPL soccer match

X = # Goals	P(x)	$(X-\mu)^2$	$(X-\mu)^2 * \mathrm{P(X)}$
0	.0711	$(0-2.768)^2 = 7.6618$	.5448
1	.1974	$(1-2.768)^2 = 3.1258$	.6170
2	.2158	$(2-2.768)^2 = .5898$	.1273
3	.1842	$(3-2.768)^2 = .0538$	.0099
4	.1658	$(4-2.768)^2 = 1.5178$	.2517
5	.1026	$(5-2.768)^2 = 4.9818$	.5111
6	.0447	$(6-2.768)^2 = 10.4458$	.4669
7	.0105	$(7-2.768)^2 = 17.9098$	.1881
8	.0026	$(8-2.768)^2 = 27.3738$	.0712
9	.0053	$(9-2.768)^2 = 38.8378$	.2058

Example 1: Discrete Distributions

- $\sigma_x^2 = \sum(x-\mu)^2 * P(x) = .5448 + .6170 + .1273 + .0099 + .2517 + .5111 + .4669 + .1881 + .0712 + 20.56 = 2.9938$

- $\sigma_x \ = \sqrt{2.9938} = 1.7303$

Example 2 Discrete Distributions Comparing Routes: Route 1

X = # of lights	P(x)	$(X-\mu)^2$	$(X-\mu)^2 * \mathrm{P(X)}$
0	.40	$(0-1)^2 = 1$	.40
1	.30	$(1-1)^2 = 0$	0
2	.20	$(2-1)^2 = 1$	.20
3	.10	$(3-1)^2 = 4$	.40

- $\sigma_x^2 = \sum(x-\mu)^2 * P(x) = .4 + .2 + .4 = 1$
- $\sigma_x = \sqrt{1} = 1$

Example 3 Discrete Distributions Comparing Routes: Route 2

X = # of lights	P(x)	$(X-\mu)^2$	$(X-\mu)^2 * \mathrm{P(X)}$
0	.20	$(0-1)^2 = 1$	.20
1	.30	$(1-1)^2 = 0$	0
2	.10	$(2-1)^2 = 1$	.10
3	.40	$(3-1)^2 = 4$	1.60

- $\sigma_x^2 = \sum(x-\mu)^2 * P(x) = .2 + 0 + .1 + 1.6 = 1.9$
- $\sigma_x \;= \sqrt{1.9} = 1.3784$

Example 2&3 Discrete Distributions Comparing Routes

- Route 1
  - $E(X) = \sum xP(x) = 1$
  - $\sigma_x \; = 1$
- Route 2
  - $E(X) = \sum xP(x) = 1.7$
  - $\sigma_x \; = 1.3784$

- Route 2 will result in more lights **on average**
- Route 2 also has a larger spread

Example 4: Discrete Distribution

- **Example:** Number of free throws made by a basketball player in 2 tries

X = # made	P(x)	$(X-\mu)^2$	$(X-\mu)^2 \ast \mathrm{P(X)}$
0	.40	$(0-1)^2 = 1$	.40
1	.40	$(1-1)^2 = 0$	0
2	.20	$(2-1)^2 = 1$	.20

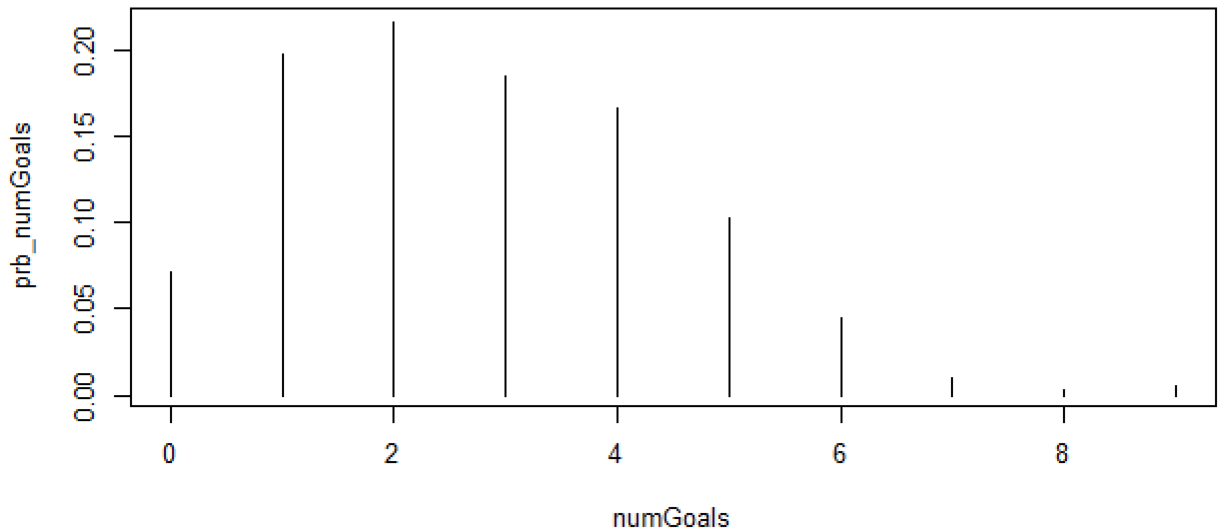
- $\sigma_x^2 = \sum(x-\mu)^2 \ast P(x) = .4 + 0 + .2 = .60$
- $\sigma_x \ = \sqrt{.60} = 77.46$

Discrete Distribution

- Note that for all of these variables we have found the mean and standard deviation
- Knowing these values we can look at a graph of the distribution, x vs. P(x), and use either Chebyshev's Rule or the Empirical Rule depending on its shape

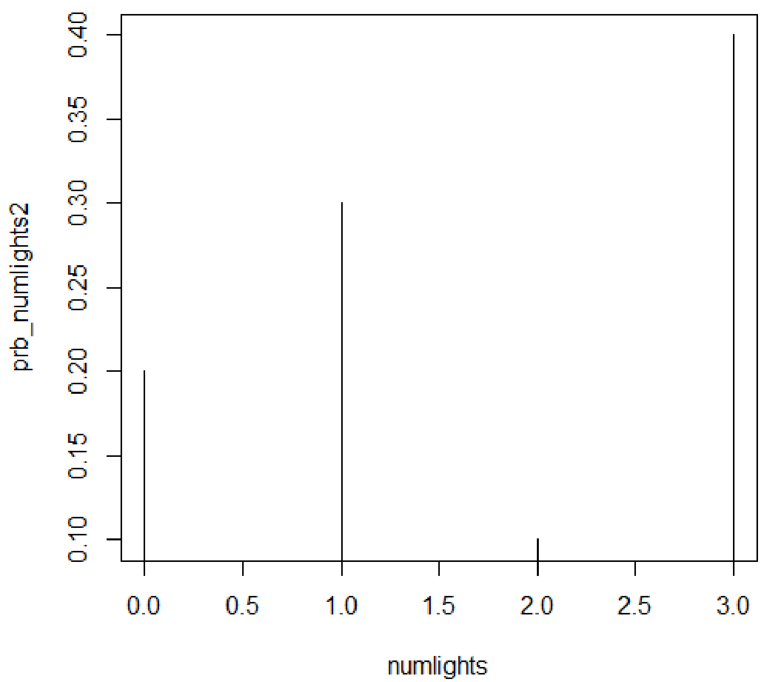
Example 1

- This is bell-shaped but it is a bit skewed so we would use Chebyshev's rule in this case



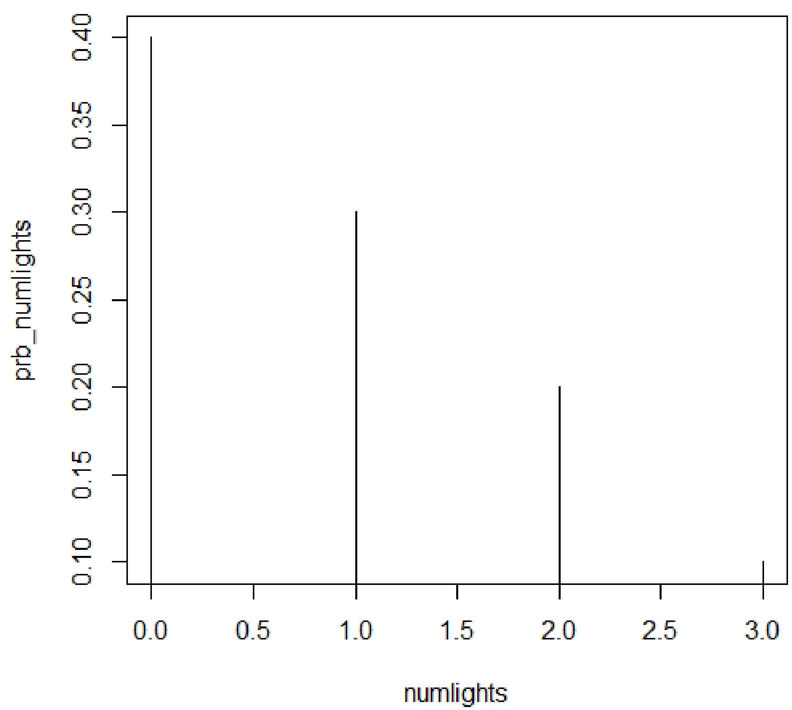
Example 2

- This is not bell-shaped so we would use Chebyshev's rule in this case



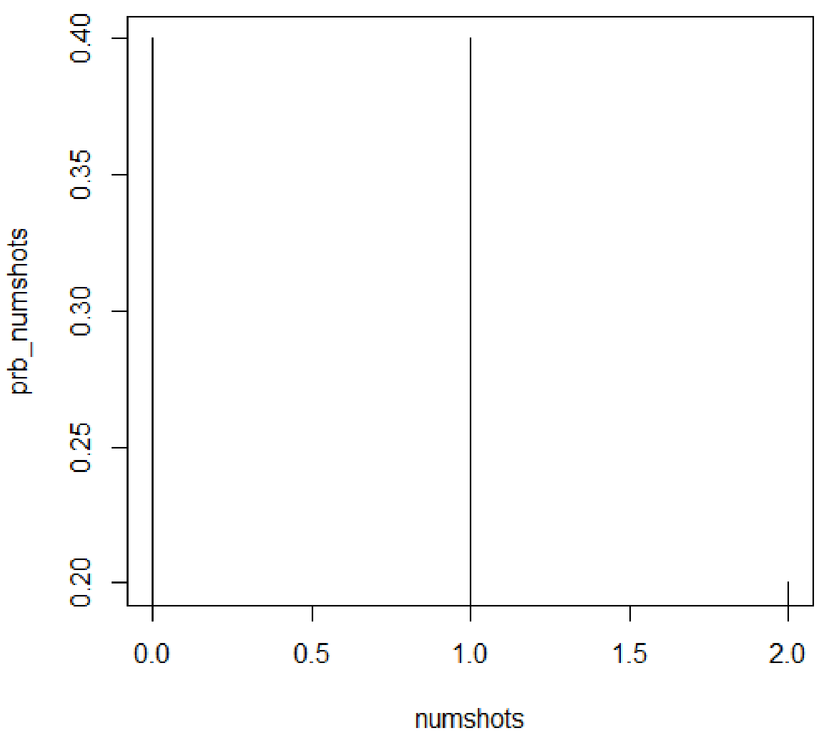
Example 3

- This is not bell-shaped so we would use Chebyshev's rule in this case



Example 4

- This is not bell-shaped so we would use Chebyshev's rule in this case



Discrete Distribution

```
numGoals<-0:9
prb_numGoals<-c()
plot(numGoals,prb_numGoals,type="h")

numlights<-0:3
prb_numlights<-c(.4,.3,.2,.1)
prb_numlights2<-c(.2,.3,.1,.4)
plot(numlights,prb_numlights,type="h")
plot(numlights,prb_numlights2,type="h")

numshots<-0:2
prb_numshots<-c(.4,.4,.2)
plot(numshots,prb_numshots,type="h")
```

A Special Discrete Distribution: **The Binomial Distribution**

- We look at a categorical variable with two outcomes
  - We consider one a success and zero a failure

x		P(x)
Success (denoted as 1)	This is what we're interested in, even if it isn't particularly successful in the sense of the English word	p = Probability of a 'success'
Failure (denoted as 0)	This is the other case – what we aren't interested in ,even if it isn't particularly a failure in the sense of the English word	q = Probability of a 'failure' = 1- p

The Binomial Distribution

- **The Binomial Distribution Assumptions**
  1. It consists of **n trials** with **binary output**
     - They are denoted 1 or 0, or success and failure
  2. The probability of success on each trial is the same
     - The trials are **identical**
  3. The outcome of one trial does not affect the outcome of another trial
     - The trials are **independent**
  4. The binomial random variable x is the number of times we see a success in n trials

The Binomial Distribution: Notation

- **n =** the number of trials
- **p =** the probability of success for any given trial (this will be the same for every trial)
- **q =** the probability of failure for any given trial
  - By complement rule: q = 1 - p
- **X =** the number of successes for n trials
- **X** is the random variable, **n** and **p** are parameters; **x** will be the observation

Binomial Formula

- $P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
- Recall: n! = n*(n-1)*(n-2)*…*2*1
  - Examples
    - 5! = 5*4*3*2*1=120
    - 0!=1
    - 5!/3!= 5*4

Binomial Calculations in R

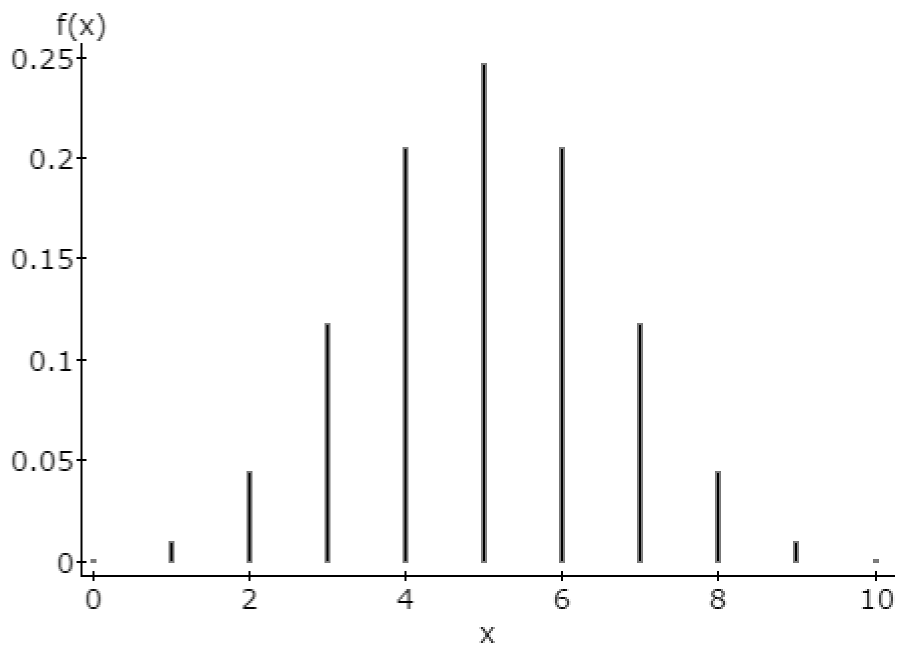
- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$ = dbinom(x,n,p)

- $P(X \leq x) = \mathrm{P(X = x) + P(X = x - 1) + \cdots + P(X = 0) = pbinom(n, p, x)}$

- $P(X > x) = 1 - \mathrm{P(X \leq x)} = 1 - \mathrm{pbinom(x, n, p)}$

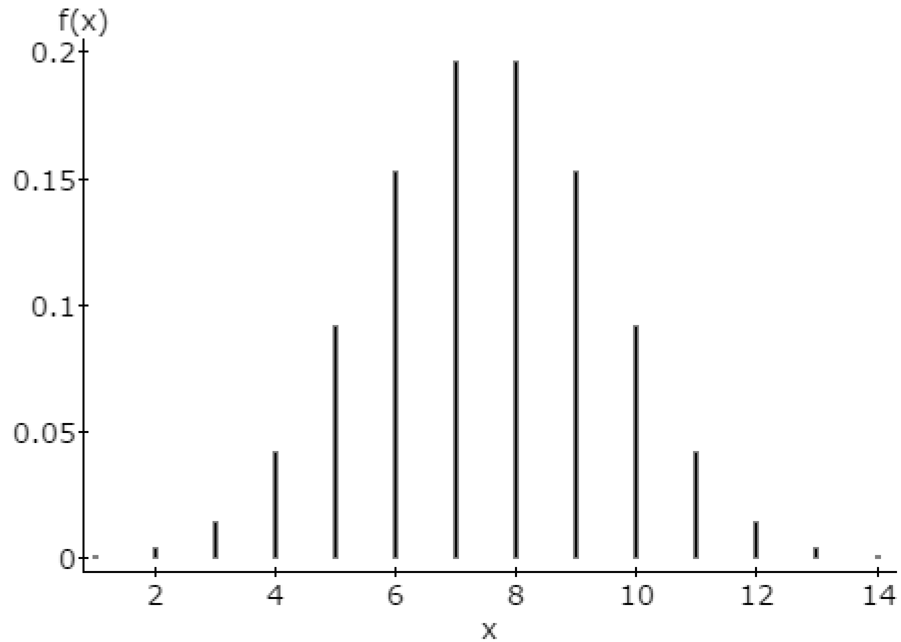
Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=10, p=.5 : Bell shaped, but there's empty space



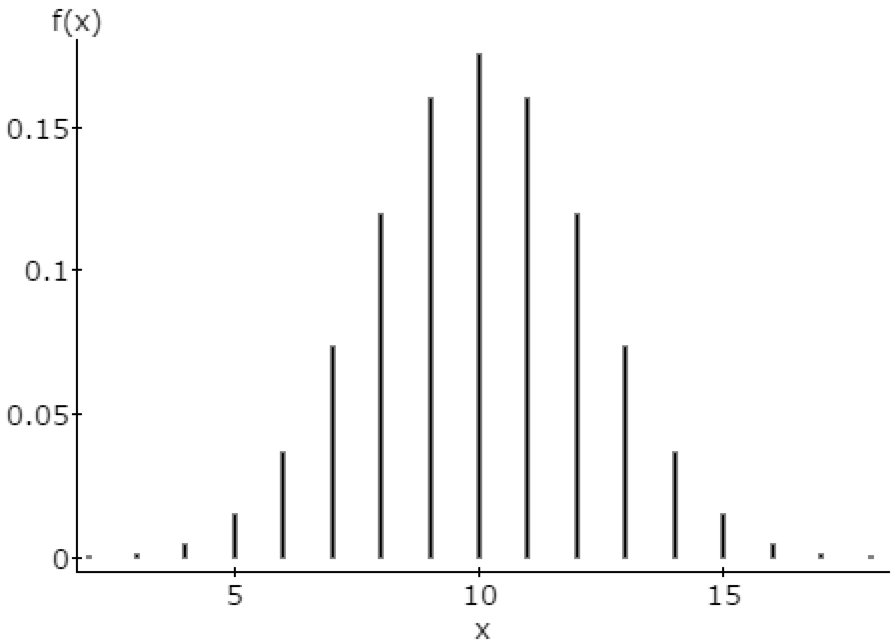
Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=15, p=.5 : Bell shaped, but there's still empty space



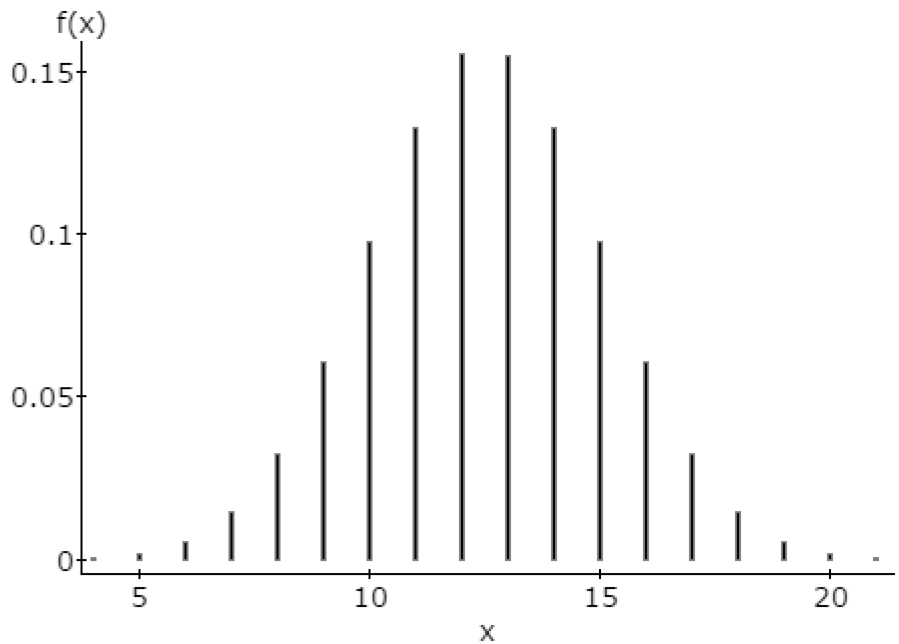
Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=20, p=.5 : Bell shaped, but there's still empty space



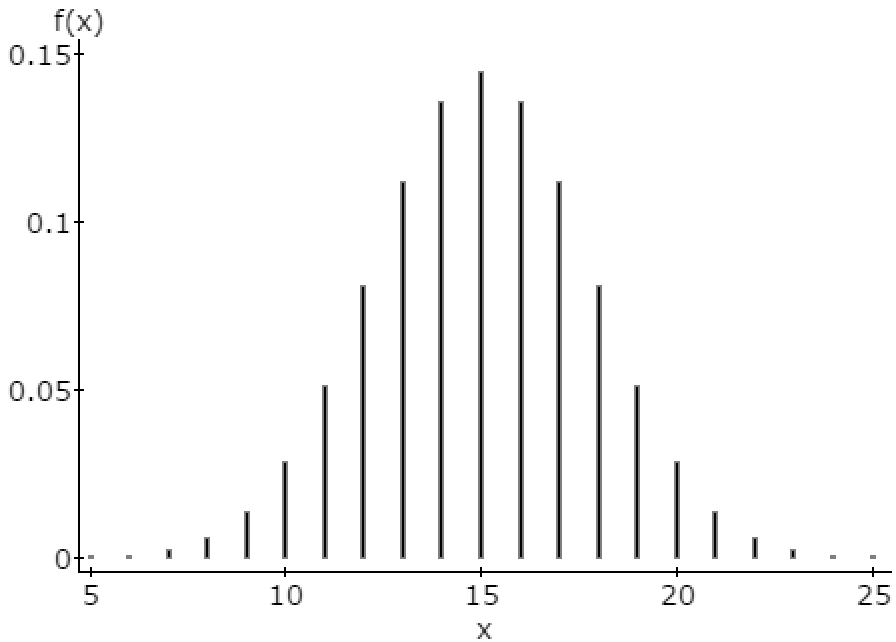
Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=25, p=.5 : Bell shaped, but there's still empty space



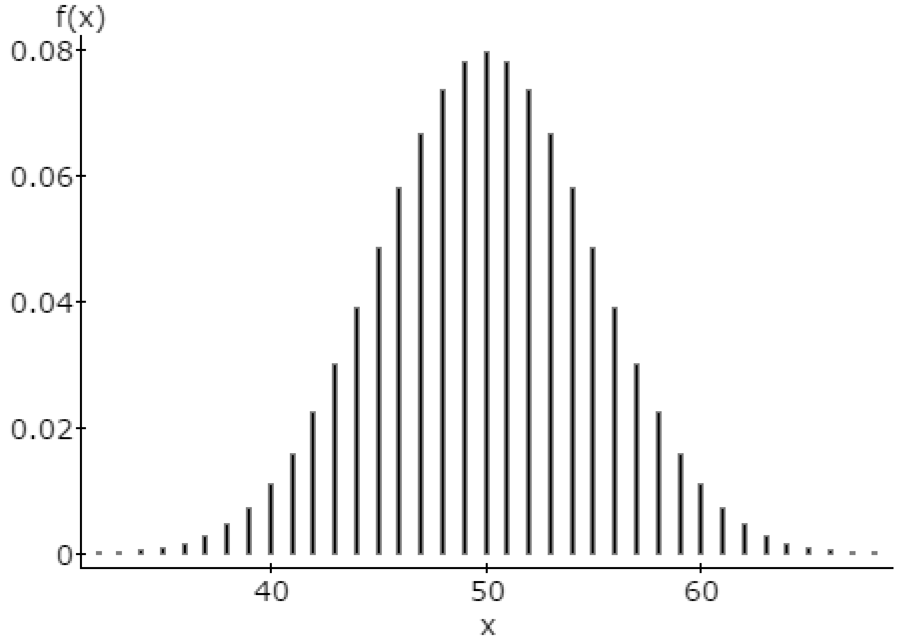
Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=30, p=.5 : Bell shaped, but there's still empty space



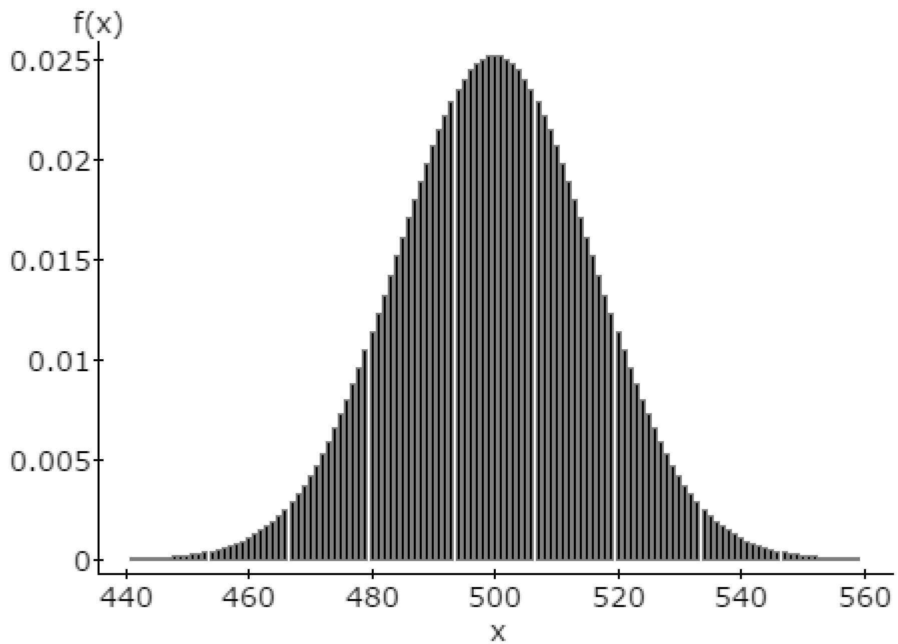
Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=100, p=.5 : Bell shaped, but there's still empty space



Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
  - n=1000, p=.5 : Bell shaped, and space is negligible



Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

  - We will say that the binomial is bell-shaped if
    $n * p \geq 15 \; AND \; n * (1-p) \geq 15$

  - We will say that the binomial is not bell-shaped if either
    $n * p < 15 \; OR \; n * (1-p) < 15$

Shape of Binomial for Graphs

n	p	n*p	n*(1-p)	Bell-Shaped?
10	.5	10*.5 = 5 $<$ 15	10*(1-.5) = 5 $<$ 15	No
15	.5	15*.5 = 7.5 $<$ 15	15*(1-.5) = 7.5 $<$ 15	No
20	.5	20*.5 = 10 $<$ 15	20*(1-.5) = 10 $<$ 15	No
25	.5	25*.5 = 12.5 $<$ 15	25*(1-.5) = 12.5 $<$ 15	No
30	.5	30*.5 = 15 $\geq$ 15	30*(1-.5) = 15$\geq$ 15	**Yes**
100	.5	100*.5 = 50 $\geq$ 15	100*(1-.5) = 50$\geq$ 15	**Yes**
1000	.5	1000*.5 = 500$\geq$ 15	1000*(1-.5) = 500$\geq$ 15	**Yes**

Shape of More Complicated Binomials

n	p	n*p	n*(1-p)	Bell-Shaped?
10	.25	10*.25 = 2.5 $<$ 15	10*(1-.25) = 7.5 $<$ 15	No
15	.25	15*.25 = 3.75 $<$ 15	15*(1-.25) = 11.25 $<$ 15	No
20	.25	20*.25 = 5 $<$ 15	20*(1-.25) = 15$\geq$ 15	No
25	.25	25*.25 = 6.25 $<$ 15	25*(1-.25) = 18.75 $\geq$ 15	No
30	.25	30*.25 = 7.5 $<$ 15	30*(1-.25) = 22.5 $\geq$ 15	No
100	.25	100*.25 = 25 $\geq$ 15	100*(1-.25) = 75 $\geq$ 15	**Yes**
1000	.25	1000*.25 = 250$\geq$ 15	1000*(1-.25) = 750 $\geq$ 15	**Yes**

Shape of a binomial

- For fixed p, as the sample size increases the probability distribution of X becomes bell shaped.
  - We consider n to be large enough when
  - $n * p > 15 \; AND \; \mathrm{n} * (1 - p) \geq 10$
  - This will be very important as we transition to inferential statistics.

What Sample Size Do I Need?

- Say we have that the probability of a success is .45, i.e. p=.45. What sample size would we need to have to say that the binomial is bell-shaped?

$$np \geq 15$$
$$n(.45) \geq 15$$
$$n \geq \frac{15}{.45}$$
$$n \geq 33.3333$$

<u>AND</u>

$$n(1-p) \geq 15$$
$$n(1-.45) \geq 15$$
$$n(.55) \geq 15$$
$$n \geq \frac{15}{.55}$$
$$n \geq 27.2727$$

- So, in order for both to be bigger than or equal to 15 we would need $n \geq 34$

Binomial Experiment – Example 1

- The two New England natives who founded Portland Oregon, Asa Lovejoy of Boston and Francis Pettygrove of Portland, Maine, both wanted to name the new city after their respective hometowns
- They decided to make the decision based on a best two-out-of-three coin toss.
- Let’s say Pettygrove chose heads

Binomial Experiment – Example 1

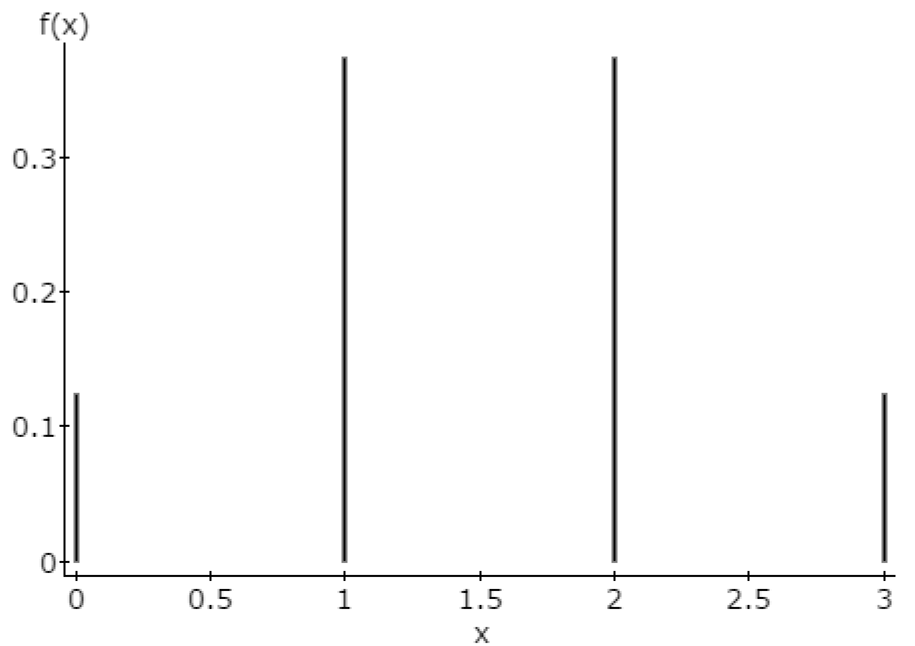
- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - **n = 3**
  - p = .50
  - q = 1 - p = 1 - .50 =.50
- **Trials are identical** – we flip the same coin each time
- **Trials are independent** as the outcome of one trial doesn’t affect another

Binomial Experiment – Example 1

- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - **n = 3**
  - p = .50
  - q = 1 - p = 1 - .50 =.50
  - $np = 3 * .5 = 1.5 < 15$ and $n(1 - p) = 3 * (1 - .5) = 1.5 < 15$
  - Because $np < 15 \ and \ n(1 - p) < 15$ we cannot say that the binomial is bell-shaped

Binomial Experiment – Example 1

- A fair one-cent piece is flipped three times
  - Because $np < 15 \ and \ n(1-p) < 15$ we cannot say that the binomial is bell-shaped



Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - n = 3, p = .50, q =.50
  - Find the probability that there are exactly 2 heads

$$\mathrm{P(X = 2)} = \frac{n!}{x!\,(n-x)!}p^x q^{n-x}$$

$$= \frac{3!}{2!\,(3-2)!}(.5)^2(.5)^{3-2} = \frac{3!}{2! * 1!}(.5)^2(.5)^1$$

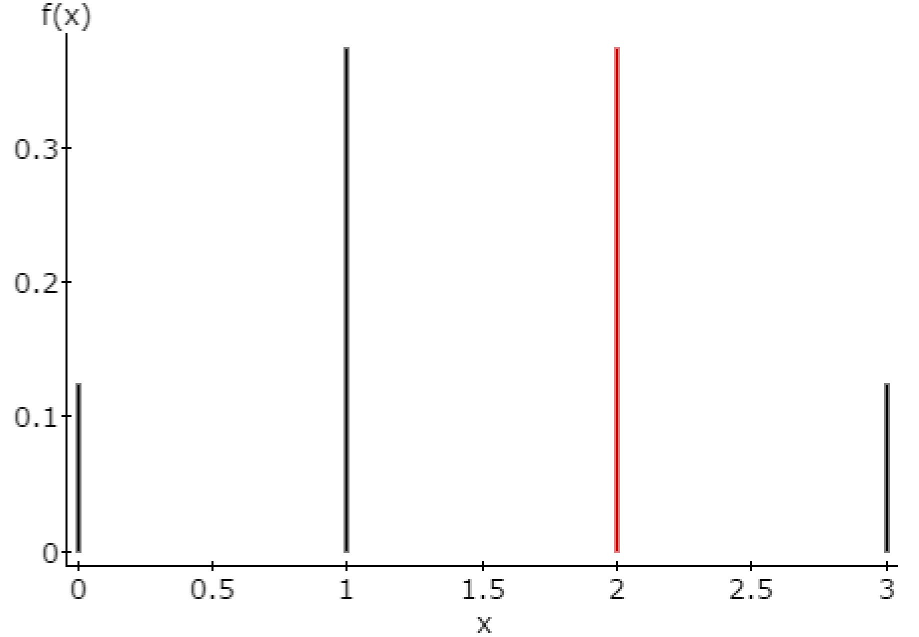
$$= \frac{3*2*1}{(2*1)*(1)}(.25)(.5)$$

$$= .375 = dbinom(2,3,.5)$$

Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times

$$P(X = 2) = .375$$



Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - n = 3, p = .50, q =.50
  - Find the probability that there are exactly 3 heads

$$\mathrm{P(X = 3)} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$$= \frac{3!}{3!\,(3-3)!} (.5)^3 (.5)^{3-3} = \frac{3!}{3! * 0!} (.5)^3 (.5)^0$$

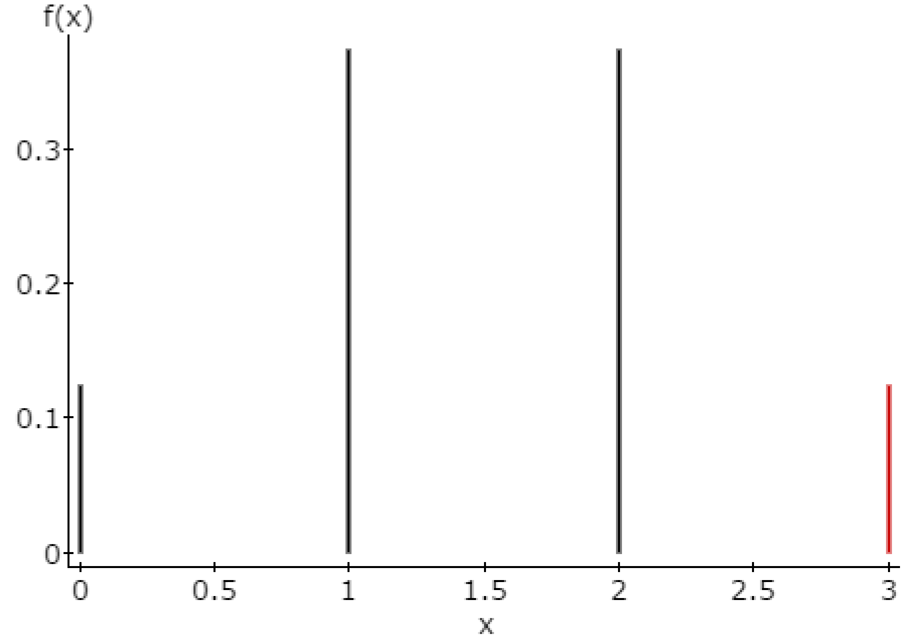
$$= \frac{3*2*1}{3*2*1} (.125)(1)$$

$$= .125 = dbinom(3,3,.5)$$

Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times

$$P(X = 3) = .125$$



Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - n = 3, p = .50, q =.50
  - Find the probability that Pettygrove wins
    - i.e. Find the probability that there are at least 2 heads

$$\mathrm{P}(\mathrm{X} \geq 2) = P(X = 2) + P(X = 3) = .375 + .125 = .5$$

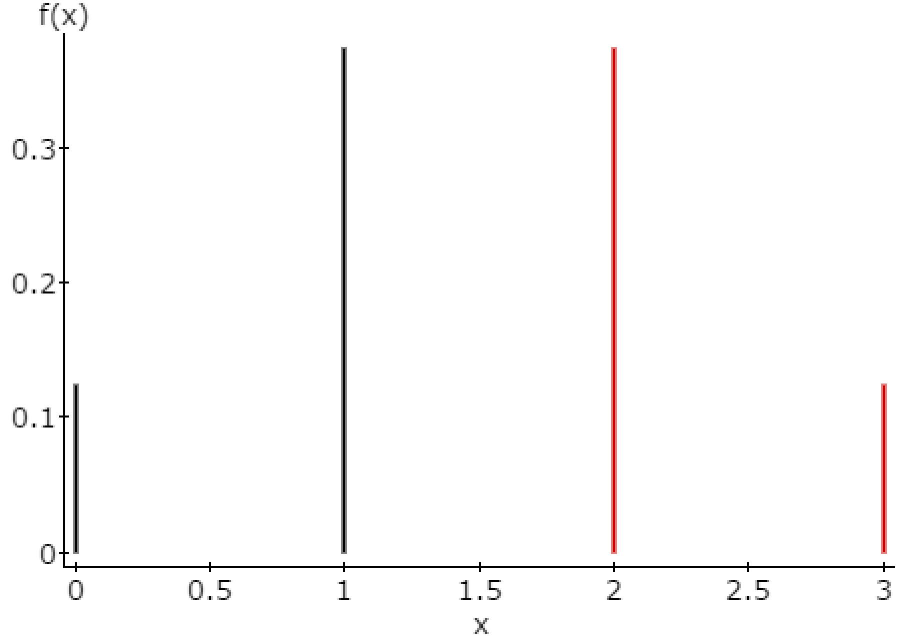
<u>OR Using Complement Rule</u>

$$\begin{aligned} \mathrm{P}(\mathrm{X} \geq 2) &= 1 - P(X < 2) \\ &= 1 - \big(P(X = 1) + P(X = 0)\big) \\ &= 1 - P(X \leq 1) \\ &= 1 - pbinom(1,3,.5) \end{aligned}$$

Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times

$$P(X \geq 2) = .5$$



Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - n = 3, p = .50, q =.50
  - Find the probability that Pettygrove loses(there are less than 2 heads)

$$\mathrm{P}(\mathrm{X} < 2) = 1 - \mathrm{P}(\mathrm{X} \geq 2)$$
$$= 1 - \big(P(X = 2) + P(X = 3)\big) = 1 - .5 = .5$$

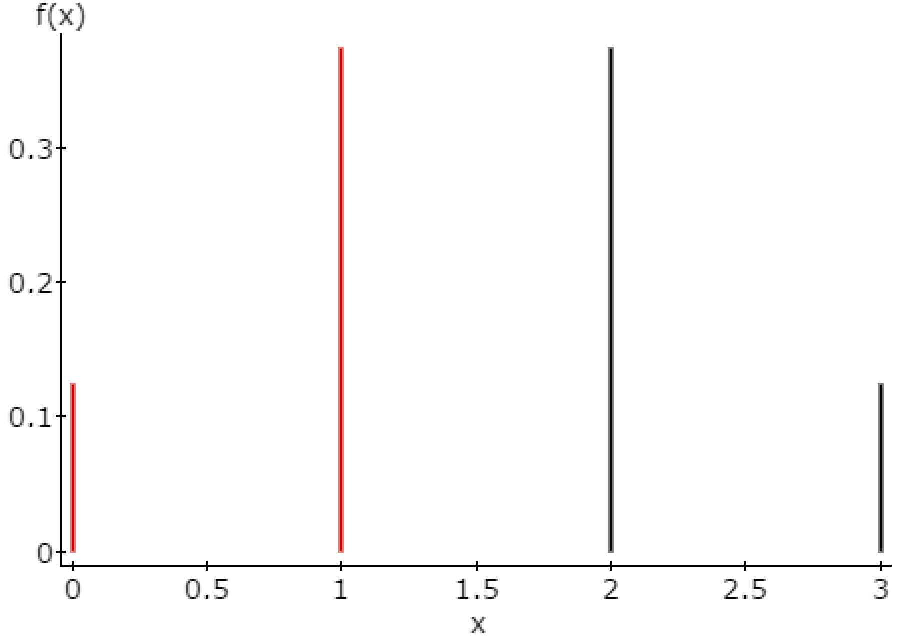
<u>OR Using Complement Rule</u>

$$P(X < 2) = P(X = 1) + P(X = 0) = P(X \leq 1)$$
$$= pbinom(1,3,.5) = .5$$

Binomial Experiment - Example 1

- A fair one-cent piece is flipped three times

$$P(X < 2) = .5$$



Binomial Experiment - Example 1

- The probability that Pettygrove wins $P(X \geq 2) = .5$
- The probability that Lovejoy wins
  $P(X < 2) = .5$
- We see that this is a **fair** game – they each have a 50% chance of winning
- So, why not just flip the coin once?

Binomial Experiment - Example 2

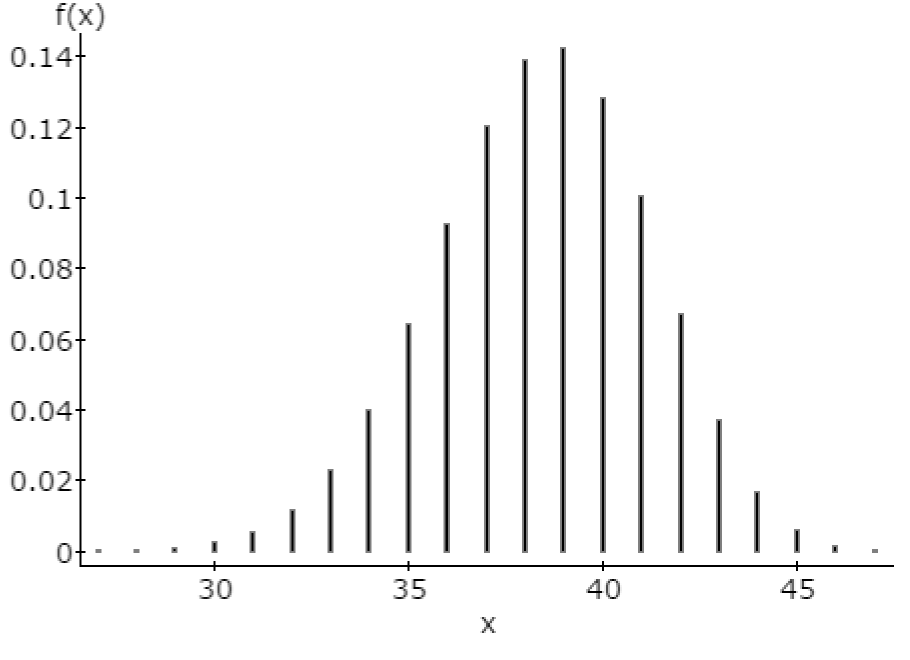
- After looking at some survey data you find that the probability that someone rates your attractiveness a two or higher is .80. Consider a class of 48 students.
- **n = 48**, p = 0.8, q = 1 - p = 1 – 0.8 = 0.2
- **Trials are independent** – one student's decision does not affect the others
- Let's go ahead and **assume identical trials** even though it can be argued that some people prefer different things

Binomial Experiment – Example 2

- Consider a class of 48 students.
  - Let X be the number of heads that occur
  - **n = 48**, p = 0.8, q = 1 - p = 1 – 0.8 = 0.2
  - $np = 48 * .8 = 38.4 \geq 15$ and $n(1 - p) = 48 * (1 - .8) = 9.6 < 15$
  - Because $n(1 - p) < 15$ we cannot say that the binomial is bell-shaped

Binomial Experiment – Example 2

- Consider a class of 48 students.
  - Because $n(1-p) < 15$ we cannot say that the binomial is bell-shaped



Binomial Experiment - Example 2

- Consider a class of 48 students.
- n = 48, p = 0.8, q = 1 - p = 1 – 0.8 = 0.2
- The probability that **exactly half** of the 48 students think you were at least a two out of ten

$$\mathrm{P(X = 24)} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$$= \frac{48!}{24!\,(48-24)!}(.8)^{24}(.2)^{48-24} = \frac{48!}{24!\,24!}(.8)^{24}\,(.2)^{24}$$

$$= .00000255 = dbinom(24, 48, .8)$$

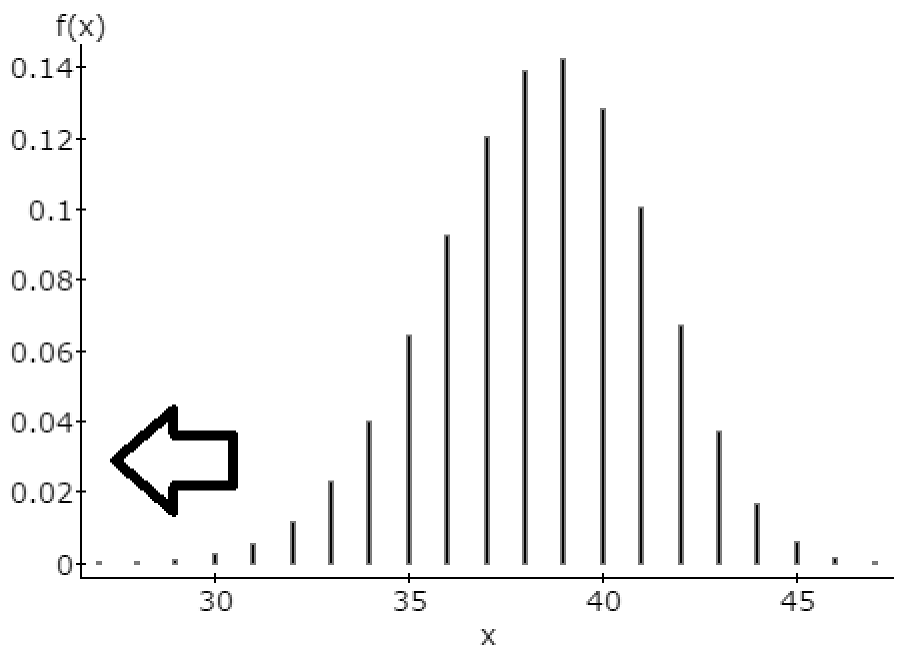
- This is an **almost impossible** event – we expect half of the class to think you were at least a two out of ten almost **0% of the time**

Binomial Experiment - Example 2

- Consider a class of 48 students.

$$P(X = 24) = .00000255$$

(Not visible because the probability is so small)



Binomial Experiment - Example 2

- Consider a class of 48 students.
- n = 48, p = 0.8, q = 1 - p = 1 – 0.8 = 0.2
- The probability that at least one of the students in your class think you were at least a two out of ten

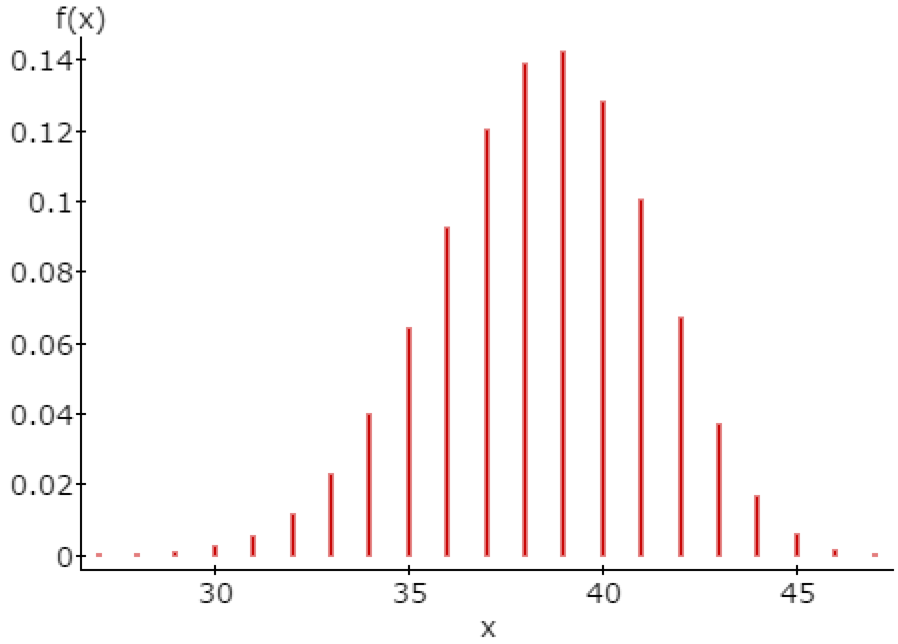
  $$\mathrm{P}(\mathrm{X} \geq 1) = P(X = 1) + P(X = 2) + \ldots P(X = 48)$$
  $$= 1 - P(X = 0) = 1 - dbinom(0,48,.8)$$
  $$= .999999999\ldots$$

- This is an **almost certain** event – we expect at least half of the class to think you were at least a two out of ten **more than 99% of the time**

Binomial Experiment - Example 2

- Consider a class of 48 students.

$$P(X \geq 1) = .999999999\ldots$$



Mean and Variance For A Binomial

- So far we have found probabilities for the binomial distribution. This gave us the ability to check the feasibility of certain outcomes or groups of outcomes.
- Here, we find what to expect!

- **Expected Value = E(X) = Mean = μ_x** $= n * p$
- **Variance =** $\sigma_x^2 = n * p * q$
- **Standard Deviation** $= \sigma_x = \sqrt{n * p * q}$

Binomial Experiment – Example 1

- A fair one-cent piece is flipped three times
  - Let X be the number of heads that occur
  - n = 3
  - p = .50
  - q = 1 - p = 1 - .50 =.50
- $\boldsymbol{Mean} = n * p = 3 * .50 = 1.50$
- **On average, we expect** between 1 and 2 heads in three flips

Binomial Experiment – Example 1

- A fair one-cent piece is flipped three times
  - Let X bet the number of heads that occur
  - n = 3
  - p = .50
  - q = 1 - p = 1 - .50 =.50
- $\boldsymbol{Standard\ Deviation} = \sqrt{3 * .50 * .50} = .75$

Binomial Experiment – Example 1

- A fair one-cent piece is flipped three times
  - Let X bet the number of heads that occur
  - n = 3, p = .50, q = .50
- $\boldsymbol{Mean} = n * p = 3 * .50 = 1.50$
- $\boldsymbol{Standard\ Deviation} = \sqrt{3 * .50 * .50} = .75$

Binomial Experiment - Example 2

- Considering a class of 48 students.
- n = 48
- p = 0.8
- q = 1 - p = 1 – 0.8 = 0.2
- $\boldsymbol{Mean} \; = \; n * p = 48 * 0.80 = 38$
- So, on average we expect about 38 of the 48 students to think you’re at least a two out of ten.

Binomial Experiment - Example 2

- Considering a class of 48 students.
- n = 48
- p = 0.8
- q = 1 - p = 1 – 0.8 = 0.2
- $\boldsymbol{Standard\ Deviation} = \sqrt{48 * .80 * .20} = 2.7713$

Binomial Experiment - Example 2

- Considering a class of 48 students.
- n = 48, p = 0.8, q = 1 - p = 1 – 0.8 = 0.2
- $\boldsymbol{Mean} \; = \; n * p = 48 * 0.80 = 38$
- $\boldsymbol{Standard\ Deviation} = \sqrt{48 * .80 * .20}$
  $= 2.7713$
- Since we cannot say this binomial is bell-shaped we cannot use the empirical rule but we can use Chebyshev’s Rule

A Special Discrete Distribution: **The Poisson Distribution**

- The Poisson random variable is for random variables that are counts
  - Number of traffic accidents at an intersection
  - Number of customers
  - etc

The Poisson Distribution

- **The Poisson Distribution Assumptions**
  1. It consists of **counting the number of times** a certain event occurs in a given amount of time or in a given area
  2. The probability an event occurs in a given unit of time or space is the same
  3. The number of events that occur in a given unit of time or space is independent of that in other units of time or space
  4. The mean is the expected number of events in each unit of time or space and is denoted by λ

The Poisson Distribution: Notation

- **X =** the number of times a certain event occurs in a given amount of time or in a given area
- λ **=** the expected number of times a certain event occurs in a given amount of time or in a given
- **X** is the random variable, λ is the parameter

Poisson Formula

- $P(X = x) = \frac{(\lambda^{x} e^{-\lambda})}{x!}$
- Recall: n! = n*(n-1)*(n-2)*…*2*1

- Mean = λ
- Variance = λ

Poisson Calculations in R

- $P(X = x) = \frac{(\lambda^{x} e^{-\lambda})}{x!}$ = dpois(x, λ)

- $P(X \leq x) = \mathrm{P}(\mathrm{X} = \mathrm{x}) + \mathrm{P}(\mathrm{X} = \mathrm{x} - 1) + \cdots + \mathrm{P}(\mathrm{X} = 0) = \mathrm{ppois}(\mathrm{x}, \lambda)$

- $P(X > x) = 1 - \mathrm{P}(\mathrm{X} \leq \mathrm{x}) = 1 - \mathrm{ppois}(x, \lambda)$

Example

- A study of the nesting of horse shoe crabs shows that the average number of satellites is 2.885 within a 50 foot radius of the nest.
  - Satellites are extramarital “boyfriends” of female horseshoe crabs
- Consider X=number of satellites within a 50 foot radius of the nest

Poisson Formula

- *Note:*
  - X=number of satellites within a 50 foot radius of the nest
  - $\lambda = 2.885$

- Mean = $\lambda = 2.885$
- Variance = $\lambda = 2.885$

Poisson Calculations in R

- *Note:*
  - X=number of satellites within a 50 foot radius of the nest
  - $\lambda = 2.885$
- $P(X = 0) = \frac{(\lambda^{\mathrm{x}} \mathrm{e}^{-\lambda})}{\mathrm{x}!}$ = dpois(0, 2.885) = .0559
- $P(X \leq 2) = \mathrm{P}(\mathrm{X} = 2) + \mathrm{P}(\mathrm{X} = 1) + \mathrm{P}(\mathrm{X} = 0) = \mathrm{ppois}(2, 2.885)$ = .4494
- $P(X > 5) = 1 - \mathrm{P}(\mathrm{X} \leq 5) = 1 - \mathrm{ppois}(5, 2.885)$=.0728

A Special Discrete Distribution: **The Hypergeometric Distribution**

- The Hypergeometric random variable is for the number of successes in n selections
- Similar to the binomial, we're interested in a success/failures
  - Here, trials **are not independent** because sampling is done **without replacement**
- Similar to the Poisson, we're interested in how many successes are in n trials (a count)

The Hypergeometric Distribution

- **The Hypergeometric Distribution Assumptions**
    1. It consists of randomly selecting n items without replacement from N items, consisting of r successes and (N-r) failures
    2. The random variable X is the number of successes among the n selected items

The Hypergeometric Distribution: Notation

- **X =** the number of successes in n trials of dependent trials done without replacement
- N = total number of items to choose from
- r = total number of success items in the N items
- n = the number of items selected
- **X** is the random variable, N, r, and n are parameters of the model

Hypergeometric Formula

- $P(X = x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$

- Mean = $\dfrac{nr}{N}$
- Variance = $\dfrac{r(N-r)n(N-n)}{N^2(N-1)}$

Hypergeometric Calculations in R

- $P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ = dhyper(x,n,N-r,r)

- $P(X \leq x) = \mathrm{P(X = x) + P(X = x - 1) + \cdots + P(X = 0)} =$ phyper(x,n,N−r,r)

- $P(X > x) = 1 - \mathrm{P(X \leq x)} = 1 -$ phyper(x,n,N−r,r)

Example

- Suppose we're playing poker - we randomly obtain 5 cards without replacement from an ordinary deck of 52 cards. What is the probability of getting exactly 3 hearts cards?
- **X =** the number of successes in n trials of dependent trials done without replacement
- N = 52 (total cards)
- r = 13 (total hearts cards)
- n = 5 (our hand)

Hypergeometric Formula

- $P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$
- Note:
  - N = 52 (total cards)
  - r = 13 (total hearts cards)
  - n = 5 (our hand)

- Mean = $\frac{nr}{N} = \frac{5*13}{52} = 1.25$
- Variance = $\frac{r(N-r)n(N-n)}{N^2(N-1)} = \frac{5*(52-13)*5*(52-5)}{52^2*(52-1)} = .3323$

Hypergeometric Calculations in R

- $P(X = 3) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ = dhyper(3,13,(52-13),5) = .0815

- $P(X \leq 3) = P(X = 3) + \mathrm{P}(\mathrm{X} = 2) + \mathrm{P}(\mathrm{X} = 1) + \mathrm{P}(\mathrm{X} = 0)$
  $=$ phyper(3,13,(52−13),5)=.9888

- $P(X > 3) = 1 - \mathrm{P}(\mathrm{X} \leq 3)$
  $= 1 -$ phyper(3,13,(52−13),5) = .0112

Summaries

Random Variable: Discrete

- The possible outcomes must be countable
  - Remember quantitative discrete variables from before
- We have a **valid** discrete probability distribution if
  1. Our outcomes are discrete (countable)
  2. All the probabilities are valid
     - $0 \leq P(x) \leq 1 \; for\ all\ outcomes\ x$
  3. We've accounted for all possible outcomes
     - $\sum P(x) = 1$

The Mean of a Discrete Distribution

- The **mean** of a probability distribution represents the average of a large number of observed values. [Remember: in the long run]

- We denote this with the Greek letter as below

$$\mu_x = E(X) = Expected\ value\ of\ x = \sum xP(x)$$

The Variance of a Discrete Distribution

- The **variance** of a probability distribution represents the spread of observed values. It is calculated by finding the expected **squared distance from the mean**
- We denote this with the Greek letter as below

$$\sigma_x^2 = E[(X-\mu)^2] = \sum (x-\mu)^2 * P(x)$$

The Standard Deviation of a Discrete Distribution

- The **standard deviation** of a probability distribution represents the spread of observed values. It is calculated by finding the square root of the variance.

- We denote this with the Greek letter as below

$$\sigma_x = \sqrt{\sigma_x{}^2} = \sqrt{\sum (x-\mu)^2 * P(x)}$$

The Binomial Distribution

- **The Binomial Distribution Assumptions**
  1. It consists of **n trials** with **binary output**
     - They are denoted 1 or 0, or success and failure
  2. The probability of success on each trial is the same
     - The trials are **identical**
  3. The outcome of one trial does not affect the outcome of another trial
     - The trials are **independent**
  4. The binomial random variable x is the number of times we see a success in n trials

The Binomial Distribution: Notation

- **n =** the number of trials
- **p =** the probability of success for any given trial (this will be the same for every trial)
- **q =** the probability of failure for any given trial
    - By complement rule: $q = 1 - p$
- **X =** the number of successes for n trials
- **X** is the random variable, **n** and **p** are parameters; **x** will be the observation

Binomial Formula

- $P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$
- Recall: n! = n*(n-1)*(n-2)*…*2*1
  - Examples
    - 5! = 5*4*3*2*1=120
    - 0!=1
    - 5!/3!= 5*4

Binomial Calculations in R

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$ = dbinom(x,n,p)

- $P(X \leq x) = \mathrm{P(X = x) + P(X = x - 1) + \cdots + P(X = 0) = pbinom(n, p, x)}$

- $P(X > x) = 1 - \mathrm{P(X \leq x) = 1 - pbinom(x, n, p)}$

Shape of Binomial

- $P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

  - We will say that the binomial is bell-shaped if $n * p \geq 15\ AND\ n * (1 - p) \geq 15$

  - We will say that the binomial is not bell-shaped if either $n * p < 15\ OR\ n * (1 - p) < 15$

The Poisson Distribution

- **The Poisson Distribution Assumptions**
  1. It consists of **counting the number of times** a certain event occurs in a given amount of time or in a given area
  2. The probability an even occurs in a given unit of time or space is the same
  3. The number of events that occur in a given unit of time or space is independent of that in other units of time or space
  4. The mean is the expected number of events in each unit of time or space and is denoted by λ

The Poisson Distribution: Notation

- **X =** the number of times a certain event occurs in a given amount of time or in a given area
- λ **=** the expected number of times a certain event occurs in a given amount of time or in a given
- **X** is the random variable, λ is the parameter

Poisson Formula

- $P(X = x) = \frac{(\lambda^{x} e^{-\lambda})}{x!}$
- Recall: n! = n*(n-1)*(n-2)*…*2*1

- Mean = λ
- Variance = λ

Poisson Calculations in R

- $P(X = x) = \frac{(\lambda^{\mathrm{x}} \mathrm{e}^{-\lambda})}{\mathrm{x}!}$ = dpois(x, λ)

- $P(X \leq x) = \mathrm{P}(\mathrm{X} = \mathrm{x}) + \mathrm{P}(\mathrm{X} = \mathrm{x} - 1) + \cdots + \mathrm{P}(\mathrm{X} = 0) = \mathrm{ppois}(\mathrm{x}, \lambda)$

- $P(X > x) = 1 - \mathrm{P}(\mathrm{X} \leq \mathrm{x}) = 1 - \mathrm{ppois}(x, \lambda)$

The Hypergeometric Distribution

- **The Hypergeometric Distribution Assumptions**
  1. It consists of randomly selecting n items without replacement from N items, consisting of r successes and (N-r) failures
  2. The random variable X is the number of successes among the n selected items

The Hypergeometric Distribution: Notation

- **X =** the number of successes in n trials of dependent trials done without replacement
- N = total number of items to choose from
- r = total number of success items in the N items
- n = the number of items selected
- **X** is the random variable, N, r, and n are parameters of the model

Hypergeometric Formula

- $P(X = x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$

- Mean = $\dfrac{nr}{N}$
- Variance = $\dfrac{r(N-r)n(N-n)}{N^2(N-1)}$

Hypergeometric Calculations in R

- $P(X = x) = \frac{\binom{n}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ = dhyper(x,r,N-r,n)

- $P(X \leq x) = \mathrm{P(X = x) + P(X = x - 1) + \cdots + P(X = 0)} =$ phyper(x,r,N−r,n)

- $P(X > x) = 1 - \mathrm{P(X \leq x)} = 1 -$ phyper(x,r,N−r,n)