Proportions

How can we <u>estimate</u> a population <u>proportion</u> using a random sample?

<u>Reminder 1:</u> The key to estimation and hypothesis testing is understanding the sampling distribution

<u>Reminder 2:</u> CTL means that even if the variable is not normally distributed, if **n** is *large enough* the sampling distribution will be normally distributed.

Bernoulli Trials:

- ***Random** process with **only two mutually exclusive outcomes***
  - Coin toss: heads versus tails
  - Contest: Win or lose
  - General: one is called a success, one is called a failure

- *The probability, **p**, of success is the same in every trial*

- *The trials are **independent-** the outcome of any particular trial has no influence on any other trial*

- **Binomial Random Variable**
  - Repeat a Bernoulli trial, with probability of success ***p***, to get a Binomial Random Variable
  - **X** is the number of successes in a fixed number, *n*, of repeated Bernoulli trials
    - Example: P(*X* = *k*), where X represents the number of heads in two coin flips so *k* = 0,1,2

k = # of successes	0	1	2
P*(X=k)*	0.25	0.50	0.25

- **Binomial Distribution:**
  - Describes the probability of a given number of '**successes**', which have a ***p*** probability, from a fixed number of independent trials, ***n***

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

- **<u>Binomial Distribution:</u>**

  - Describes the probability of a given number of 'successes', which have a *p* probability, from a fixed number of independent trials, *n*

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

Binomial Coefficient

- <u>The Binomial coefficient:</u>

  It counts all the unique unordered sequences of getting k successes in n trials. *ie. how many ways are there of getting k successes?*

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

Where n! = n x (n-1) x (n-2) x … x1
Also: **0! = 1 and 1! = 1**

- <u>The Binomial coefficient:</u>

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

<u>Example:</u> How many ways can 2 letters be chosen from the set {A B C D}?

- <u>The Binomial coefficient:</u>

  It counts all the unique unordered sequences of getting *k* successes in *n* trials.

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

<u>Example:</u> How many ways can the 2 letters be chosen from the set {A B C D}?

<u>Answer:</u> 4 choose 2 = 4!/(2!2!) = 4x3x2x1/{(2x1)(2x1)} = 6

<u>AB</u> <u>AC</u> <u>AD</u> <u>BC</u> <u>BD</u> <u>CD</u>

You might notice that this treats <u>AB</u> the same as <u>BA</u>

- <u>The Binomial coefficient:</u>

  It counts all the unique unordered sequences of getting *k* successes in *n* trials.

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

<u>Example:</u> What is the probability of getting exactly the following pattern (2 successes and 3 failures):

<u>F F S F S</u>

- <u>The Binomial coefficient:</u>

  It counts all the unique unordered sequences of getting *k* successes in *n* trials.

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

<u>Example:</u> What is the probability of getting the following pattern: <u>F F S F S</u>

<u>Answer:</u> the hard way (drawing them all out…)

<u>FFFSS FFSFS FFSSF FSFFS FSFSF FSSFF SFFFS SFFSF SFSFF SSFFF</u>

- <u>The Binomial coefficient:</u>

It counts all the unique unordered sequences of getting *k* successes in *n* trials.

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

<u>Example:</u> What is the probability of getting the following pattern: <u>F F S F S</u>

<u>Answer:</u>

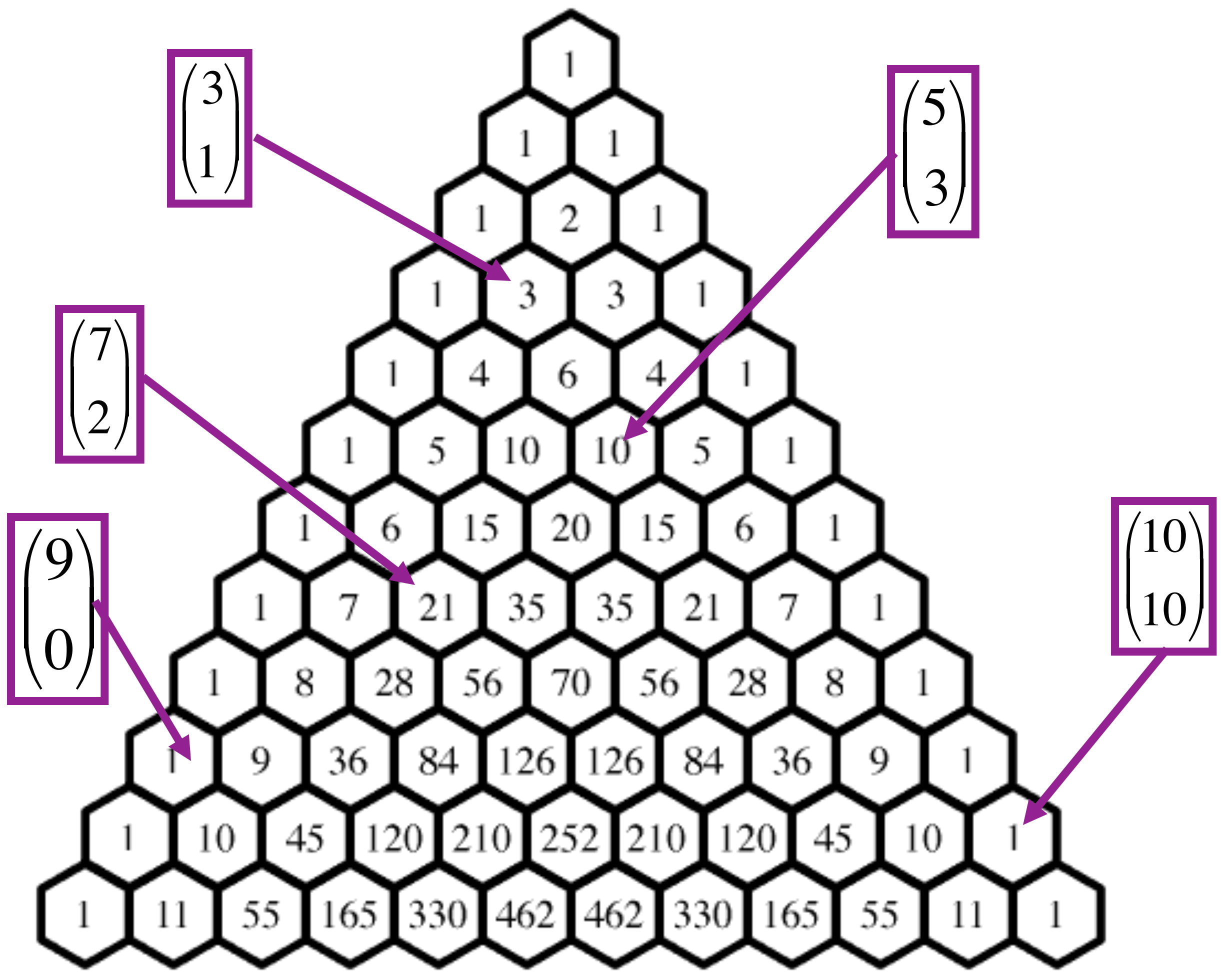
the hard way (drawing them all out…)

<u>FFFSS FFSFS FFSSF FSFFS FSFSF FSSFF SFFFS SFFSF SFSFF SSFFF</u>

The easy way: 5 choose 3 = 5x4x3x2x1/ {(2x1)(3x2x1)} = 10

Binomial Distribution

We can use “Pascal’s triangle” to determine the binomial coefficient

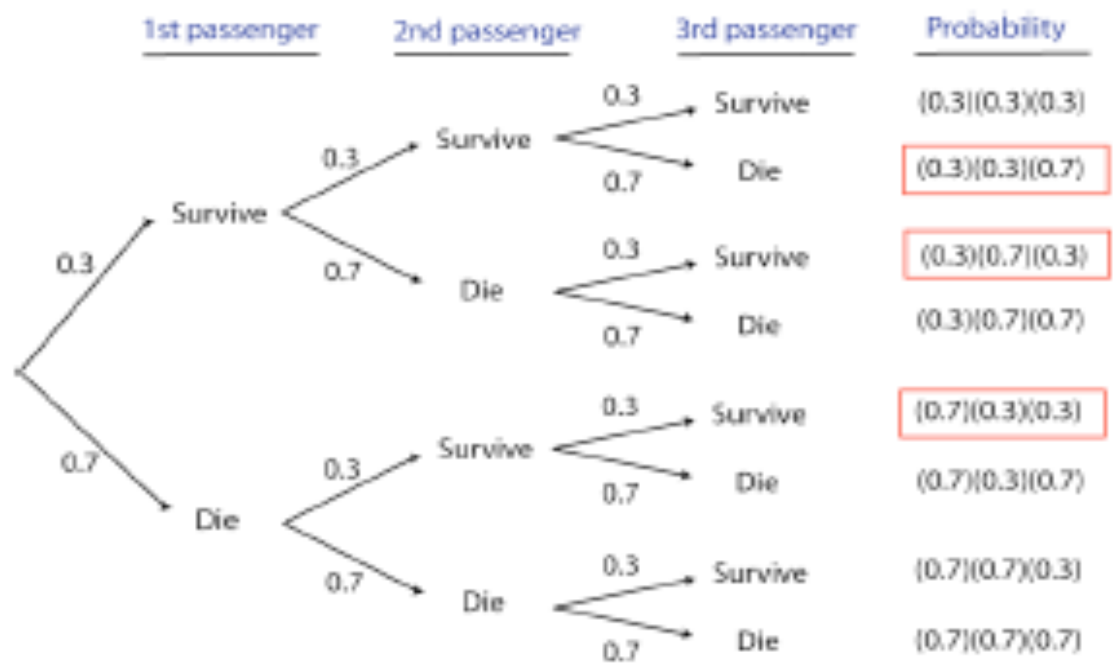


Example: 2092 passengers on the titanic; 654 survived

P(surviving) = 654/2092 = 0.3

What is the probability that 2 out of 3 randomly chosen passengers survived?

Answer: The hard way…



<u>Example:</u> 2092 passengers on the titanic; 654 survived

P(surviving) = 654/2092 = 0.3

What is the probability that 2 out of 3 randomly chosen passengers survived?

<u>Answer:</u> The easy way….

$$P(X = 2) = \binom{3}{2}(0.3)^2(0.7)^1$$

Binomial Distribution

- In general, questions that involves phrases such as *at least, at most, less than, more than* etc

<u>How would you answer this question?</u>

- 70% of colon cancers are cured if detected early using a new screening test; what is the probability that 9 out of 10 patients are cured?

<u>How about this one?</u>

- A woman has three sons. What is the probability that her next child is a girl?

12 people have been exposed to a particular disease. Each one independently has a 40% chance of contracting the disease. A small, county hospital has the capacity to handle 9 cases of the disease. What is the probability that the hospital's capacity will be exceeded?

a. 0.9972

b. 0.0028

c. 0.00562

d. 0.9944

Asymmetry is present in animals in many of their preferences including hand usage and dominant eye and ear. To test ear usage in human beings, an observer scored whether the person approached turned either the left or right ear toward the questioner. Of 12 participants, 10 turned the right ear toward the questioner and 2 offered the left ear. Use an alpha=0.05.

----

a. P=0.0193, reject

b. P=0.0386, fail to reject

c. P=0.0193, fail to reject

d. P=0.0386, reject

Properties of the Binomial Distribution:

A binomial random variable has values that are the number of successes

The long way of demonstrating the mean and standard deviation:

If 40% of brand A widgets have a particular defect, and I buy 5 of these widgets, what is the expected number of defective widgets that I now own?

Properties of the Binomial Distribution:

A binomial random variable has values that are the number of successes

ANSWER: (Make sure you can replicate these calculations…)

Outcome:	0 widgets	1 widget	2 widget	3 widget	4 widget	5 widget
Probability:	0.07776	0.25920	0.34560	0.23040	0.07680	0.01024
Random Variable:	0	1	2	3	4	5

$$\boxed{\bar{X} = 0(0.07776) + 1(0.25920) + 2(0.34560) + 3(0.23040) + 4(0.07680) + 5(0.01024) = 2}$$

This is the same answer as would be obtained by simply multiplying the probability of "success" times the number of cases….

Properties of the Binomial Distribution:

A binomial random variable has values that are the number of successes

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

As $n \rightarrow \infty$, the Binomial dist'n is approached by the Normal dist'n (deMoivre originally showed this by using "the calculus" with $p = 0.5$ to obtain the standard normal distribution but it works When $p \neq 0.5$ if n is big enough thanks to the CLT) https://homepage.divms.uiowa.edu/~mbognar/applets/binnormal.html

Binomial distribution describes the number of successes but *since the number of trials is **fixed*** it also gives the proportion of successes. Here is how we get the mean:

$$\hat{p} = \frac{X}{n}$$

$$E(\hat{p}) = E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}E(n\rho) = \rho$$

Mean: $\hat{p} = \rho$

Proportions

Properties of sample proportions:

Mean:

$$\hat{p} = \rho$$

Variance:

$$\sigma_{\hat{p}}^2 = \frac{\rho(1-\rho)}{n}$$

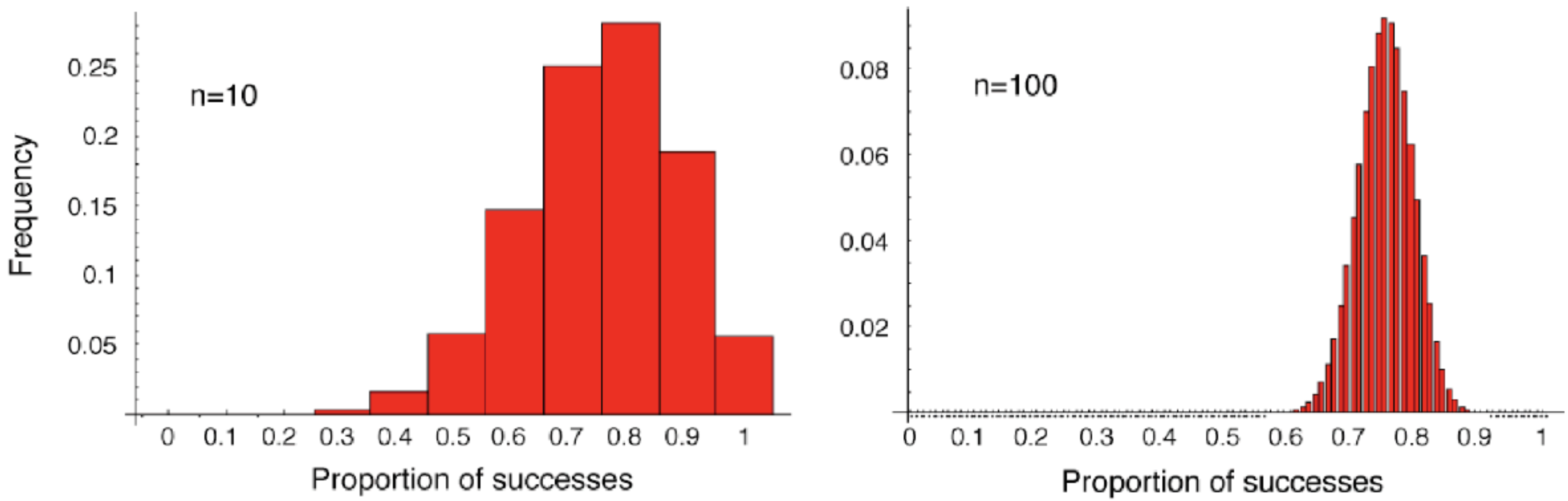
(Variance can be simply derived from Bernoulli like so:)

$$Var(\frac{X}{n}) = \frac{Var(X)}{n^2} = \frac{n\rho(1-\rho)}{n^2} = \frac{\rho(1-\rho)}{n}$$

Standard Error of the estimate of a proportion:

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The law of large numbers:



$\bar{X}_n \rightarrow \mu$ as n $\rightarrow \infty$ with probability of 1. In words this means that the sample mean converges to the true mean …eventually (with a large enough sample)

The greater the sample size, the greater the precision of the estimate of a proportion

A good explanation of the Law of Large Numbers and the closely related CTL:

https://www.youtube.com/watch?v=9yQpg3z9_DM