Stat 250 Final Review

Probability Distributions and Sampling Distributions

Discrete Probability Distributions

We have been told to be able to:

- Identify if a random variable is discrete or continuous
- Express the probability distribution of a discrete random variable in a table
- Verify that a discrete distribution satisfies the properties:
  - All probabilities between 0 and 1
  - Probability of all possible outcomes sums to 1
- Compute the mean (expected value) of a discrete random variable using a probability distribution table

These calculations are easy for the final, so we should be prepared for them!

1. The table below shows X = the number of red lights a commuter may hit on the way to work when crossing through 5 intersections.

Number of red lights (X)	0	1	2	3	4	5
Probability P (X)	0.05	0.25	0.35		0.15	0.05

a.) What is the missing probability?

b.) What is the probability that a commuter will hit 3 or more lights on the way to work?

c.) What is the expected value for the number of lights a commuter will hit on the way to work?

1. The table below shows X = the number of red lights a commuter may hit on the way to work when crossing through 5 intersections.

Number of red lights (X)	0	1	2	3	4	5
Probability P (X)	0.05	0.25	0.35		0.15	0.05

a.) What is the missing probability?

1 – (0.05 + 0.25 + 0.35 + 0.15 + 0.05) = 0.15

b.) What is the probability that a commuter will hit 3 or more lights on the way to work?

0.15 + 0.15 + 0.05 = 0.35

c.) What is the expected value for the number of lights a commuter will hit on the way to work?

0(0.05) + 1(0.25) + 2(0.35) + 3(0.15) + 4(0.15) + 5(0.05) = 2.25

Continuous Probability Distributions

We have learned about five continuous probability distributions:

- Normal – will say the word "NORMAL"
- Binomial – will be given a n and p, with a small sample size
- Uniform – will say the word "UNIFORM"
- t – will only be used in CI for t* and Hypothesis test
- X^2 – will only be used for hypothesis test for categorical data.

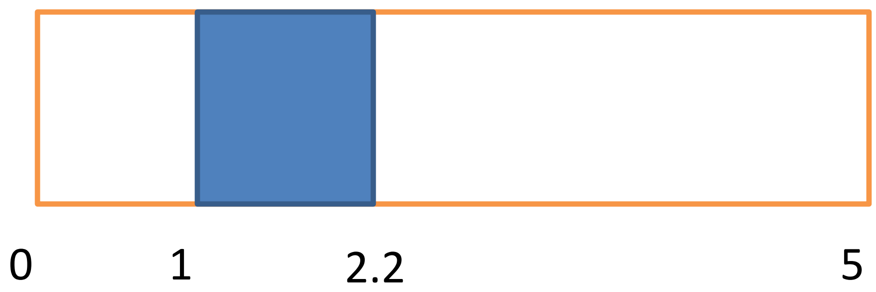
Uniform Distribution

- Know that a probability distribution for a continuous random variable is expressed with a density curve
  - The curve always falls at or above zero
  - The probability a random variable falls in a given interval is equal to the area under the curve on that interval
- Compute probabilities of uniform random variables using a rectangular density “curve” (like the trolley waiting time example)

2. Suppose that time to complete a question has a uniform distribution between 0 and 5 minutes, what is the probability of having a student selected at random takes between 1 and 2.2 minutes to complete the question?

2. Suppose that time to complete a question has a uniform distribution between 0 and 5 minutes, what is the probability of having a student selected at random takes between 1 and 2.2 minutes to complete the question? **0.24**

Reasoning: Uniform means the RECTANGLE!



The entire area of the rectangle is 1, so we can find the height by 1 = 5h, h = 0.2. And we want the shaded region's area from 1 to 2.2. A = bh = (1.2)(0.2)

Normal Distribution

DRAW PICTURES ON SCRAP PAPER!

- Find areas under a given normal distribution (left tail, interval, or right tail) using normalcdf on calculator or NORM.DIST on Excel
- Find quantiles of any normal distribution (a value in the distribution given an area – inverse normal problems) using invNorm on calculator or NORM.INV in Excel
- Know how to turn values from any normal distribution into z-scores from the standard normal. Interpret z score as number of standard deviations above or below mean.

3. The weights of adorable, fluffy kittens are normally distributed with a mean of 3.6 pounds and a standard deviation of 0.4 pounds.

What percent of adorable, fluffy kittens weigh between 2.8 and 4.8 pounds?

What percent of adorable, fluffy kittens weigh less than 2.4 pounds?

What value corresponds to a 97.5th percentile of kitten weights?

3. The weights of adorable, fluffy kittens are normally distributed with a mean of 3.6 pounds and a standard deviation of 0.4 pounds.

What percent of adorable, fluffy kittens weigh between 2.8 and 4.8 pounds? **97.35%**

What percent of adorable, fluffy kittens weigh less than 2.4 pounds? **0.15%**

What value corresponds to a 97.5th percentile of kitten weights? **4.4 pounds**

4. Assume college women's heights follow a normal distribution with a mean height of 65 inches and a standard deviation of 2.7 inches. Which of the following correctly describes the probability that a college woman, selected at random, will be shorter than 62 inches?

A. The area to the left of 1.11 on the standard normal curve.
B. The area to the right of -1.11 on the standard normal curve.
C. The area to the left of -1.11 on the standard normal curve.
D. The area between -1.11 and 1.11 on the standard normal curve.

4. Assume college women's heights follow a normal distribution with a mean height of 65 inches and a standard deviation of 2.7 inches. Which of the following correctly describes the probability that a college woman, selected at random, will be shorter than 62 inches?

A. The area to the left of 1.11 on the standard normal curve.

B. The area to the right of -1.11 on the standard normal curve.

C. The area to the left of -1.11 on the standard normal curve.

D. The area between -1.11 and 1.11 on the standard normal curve.

Reasoning: $z = \frac{x-\mu}{\sigma} = \frac{62-65}{2.7} = -1.11$ **and the shorter keyword lets us know to shade to the left.**

Binomial Distribution

- Know the properties of a binomial experiment and identify whether a series of trials satisfies the criteria for a binomial experiment.
- Know the formula for the P(X=x) in a binomial distribution and be able to describe the components of it
- Compute the mean and standard deviation of a binomial distribution
- Compute binomial probabilities using binomcdf, binompdf, or binom.dist.
- Know under which conditions the normal distribution can be used to approximate the binomial distribution
- Use the normal distribution to approximate the binomial distribution when appropriate (continuity correction not required) - find approximate probabilities using the normal curve

if we’re looking at a binomial experiment with n trials and p probability of success, the probability of getting k successes is (this is what binompdf does, or binom.dist with cumulative=False):

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

To sum up multiple values if asked for a range, we can use the sigma notation, so the probability of less than or equal to k successes (this is what binomcdf does, or binom.dist with cumulative=True):

$$P(X \leq k) = \sum_{x=0}^{k} \binom{n}{x} p^x (1 - p)^{n-x}$$

5. Sixty-five percent of men consider themselves knowledgeable soccer fans. If 10 men are randomly selected, find the following probability that exactly 5 of the men chosen consider themselves knowledgeable soccer fans.

A. 0.65

B. (0.65)^5

C. $\binom{10}{5}.65^{5}\,(.35)^{10-5}$

D. $\sum_{x=0}^{5}\binom{10}{x}.65^{x}\,(.35)^{n-x}$

5. Sixty-five percent of men consider themselves knowledgeable soccer fans. If 10 men are randomly selected, find the following probability that exactly 5 of the men chosen consider themselves knowledgeable soccer fans.

A. 0.65

B. (0.65)^5

C. $\boldsymbol{\binom{10}{5} .65^{5} (.35)^{10-5}}$ **- also be able to compute this value.**

D. $\sum_{x=0}^{5} \binom{10}{x} .65^{x} (.35)^{n-x}$

6. Sixty-five percent of men consider themselves knowledgeable soccer fans. If 10 men are randomly selected, find the following probability that at least 2 of the men chosen consider themselves knowledgeable soccer fans.

A. 0.65

B. $\binom{10}{2}.65^2\,(.35)^{10-2}$

C. $\sum_{x=0}^{2}\binom{10}{x}.65^x\,(.35)^{n-x}$

D. $\sum_{x=2}^{10}\binom{10}{x}.65^x\,(.35)^{n-x}$

6. Sixty-five percent of men consider themselves knowledgeable soccer fans. If 10 men are randomly selected, find the following probability that at least 2 of the men chosen consider themselves knowledgeable soccer fans.

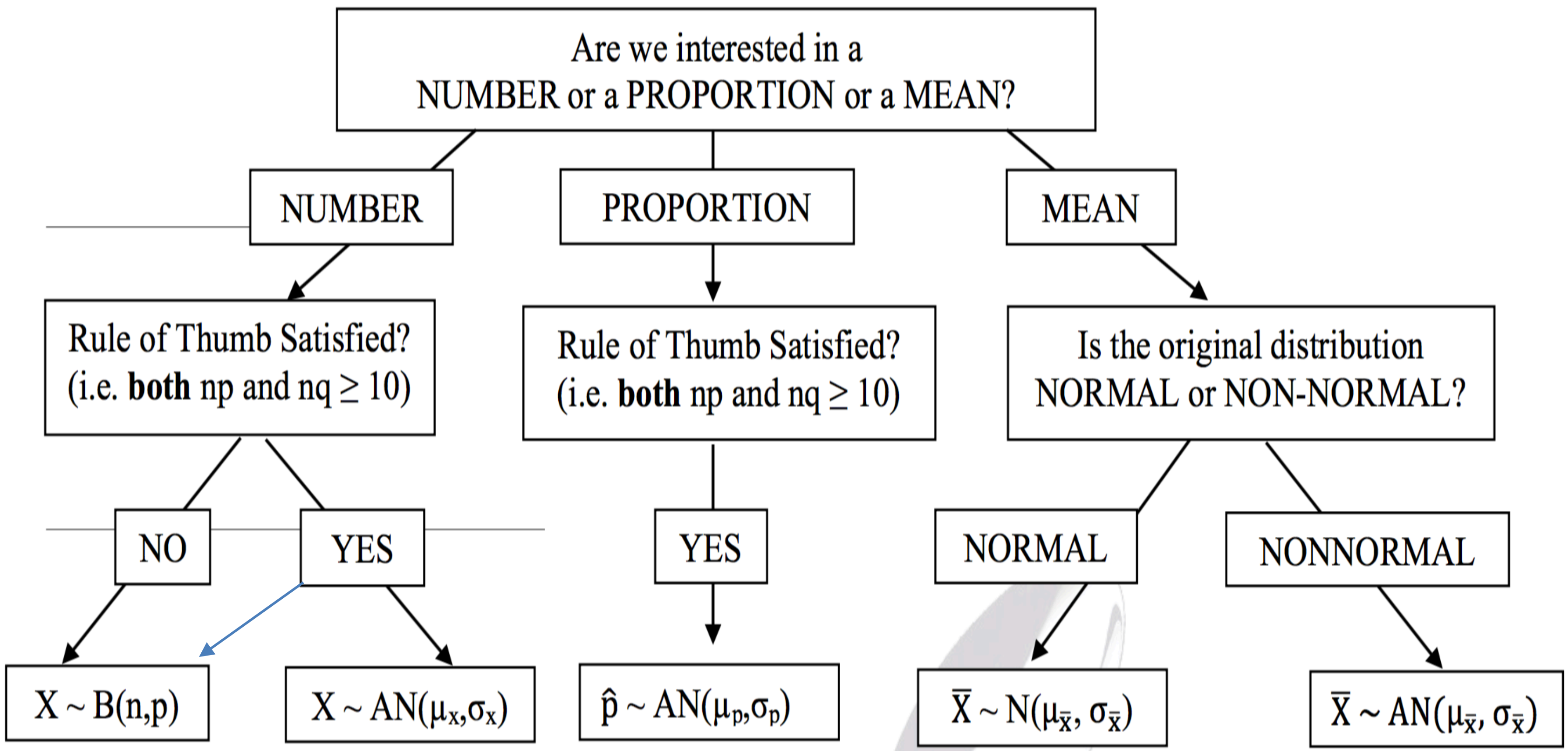
A. 0.65

B. $\binom{10}{2}.65^2\,(.35)^{10-2}$

C. $\sum_{x=0}^{2}\binom{10}{x}.65^x\,(.35)^{n-x}$

D. $\boldsymbol{\sum_{x=2}^{10}\binom{10}{x}.65^x\,(.35)^{n-x}}$ **- you'll have to subtract to get this ; 1-binomcdf(10,.65,1)**

Sampling Distributions



Sampling Distributions

For means and proportions:

- As the sample size increases, the mean of the sampling distribution of phat or xbar will stay the same and the standard deviation will decrease.
- If X is a count of successes from an experiment with n trials and p probability of success or population proportion in each trial, then the SAMPLE PROPORTION ($\hat{p}$) has mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$
- If X follows a distribution with mean μ and standard deviation σ, then the SAMPLE MEAN ($\bar{x}$) has mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

7. An agriculturalist is trying to breed a plant with with a particular genetic variant. There are 320 plants on a table. In the current breeding process, 12.5% of plants have this variant. Let Y be the number of plants on a table with the variant. Which of the following correctly describes Y?

I. Approximate normal(40, 5.916)

II. Binomial(320, 0.125)

III. Uniform(1,320)

A. I only
B. II only
C. I and II
D. I, II, and III

7. An agriculturalist is trying to breed a plant with with a particular genetic variant. There are 320 plants on a table. In the current breeding process, 12.5% of plants have this variant. Let Y be the number of plants on a table with the variant. Which of the following correctly describes Y?

I. Approximate normal(40, 5.916)

II. Binomial(320, 0.125)

III. Uniform(1,320)

A. I only
B. II only
C. **I and II**
D. I, II, and III

Reasoning: NUMBER! Always binomial and n is large enough (np and nq are greater than 10, so we can use the approximate normal. Does not say uniform.

8. A group of scientists wanted to estimate the proportion of geese returning to the same site for the next breeding season. Suppose they decided to increase the sample size from 200 to 500. How will this affect the distribution of the sample proportion?

A. The distribution of the sample proportion will be less spread out.
B. The distribution of the sample proportion will be more spread out.
C. The spread of the distribution of the sample proportion will remain unaffected.
D. The distribution of the sample proportion will more closely resemble the binomial distribution.
E. The distribution of the sample proportion will be more strongly skewed.

8. A group of scientists wanted to estimate the proportion of geese returning to the same site for the next breeding season. Suppose they decided to increase the sample size from 200 to 500. How will this affect the **distribution of the sample proportion**?

A. The distribution of the sample proportion will be less spread out.

B. The distribution of the sample proportion will be more spread out.

C. The spread of the distribution of the sample proportion will remain unaffected.

D. The distribution of the sample proportion will more closely resemble the binomial distribution.

E. The distribution of the sample proportion will be more strongly skewed.

Reasoning:

We know: $\boldsymbol{\hat{p} \sim AN\left(p, \sqrt{\frac{pq}{n}}\right)}$

If we increase n, the standard deviation goes down, but nothing else changes.

9. The Harvard College Alcohol Study finds that 67% of college students support efforts to “crack down on underage drinking.” The administration of a college surveys 100 students and finds that 62% support a crackdown on underage drinking. Which is the mean for the sampling distribution of the sample proportion of students who support a crackdown on underage drinking.

A. 0.67
B. 0.62
C. 67
D. 62

9. The Harvard College Alcohol Study finds that 67% of college students support efforts to “crack down on underage drinking.” Assume the value is correct. The administration of a college surveys 100 students and finds that 62% support a crackdown on underage drinking. Which is the mean for the sampling distribution of the **sample proportion** of students who support a crackdown on underage drinking.

A. 0.67

B. 0.62

C. 67

D. 62

Reasoning:

We know: $\hat{p} \sim AN\left(p, \sqrt{\frac{pq}{n}}\right)$

The mean is p, REAL p, which is the 0.67! 0.62 is the p-hat. C would be the correct answer for the mean of the NUMBER of students…

10. The GPA for incoming freshman at a large college follows a normal distribution with mean GPA of 3.62 and standard deviation of 0.29. If a random sample of 19 incoming freshman are selected, what is the probability the mean GPA in the sample is at least 3.5?

10. The GPA for incoming freshman at a large college follows a normal distribution with mean GPA of 3.62 and standard deviation of 0.29. If a random sample of 19 incoming freshman are selected, what is the probability the mean GPA in the sample is at least 3.5?

0.964