**Math 3307 Chapter 7 Video Scripts plus Popper**

**Chapter 7 – Random Variables and Probability Distributions**

**7.1 – 7.3**

What is a Random Variable?

Technically:

A quantitative variable whose value is determined by the outcome of a chance experiment. We almost always call the variable X with a capital letter. And P(X) is the probability of the random variable X. There are discrete RVs with just a list of values and continuous ones (at the end of the chapter).

The book’s example of a free throw is excellent! It starts on page 183…please check it out on your own as well as below:

Let’s look at the tree diagram:

And check out the geometric presentation on p.186. Again relating AREA to probability. An important idea.

And let’s make a distribution table using the values from the tree diagram:

Note: this is a theoretical distribution. We made it up from facts and not just counting her at a practice. That would be a frequency table.

Now it shouldn’t come as a surprise that we have a measure of center (E(X), the expected value) and a measure of variability (S(X)). The calculation for the expected value is on p. 190. It comes out to .95 in the free throw situation. I’ll do the variability here, the formula is on p. 195. I’ll write it out below. It comes to .88. The expected value is the predicted long term average value of X. So Nikki gets close to 1 point on average in free throw situations, with standard deviation of .88. In a few minutes we’ll review z score for this too!

Variability and SD

**Chapter 7 Popper Question 1**

A theoretical distribution has measures of center and variability just like a distribution table made from actually counting frequency.

A. True

B. False

Let’s look at an example that is not from the book. Suppose we have a loaded die. One we would use to cheat with. Here’s the distribution table:

X P(X)

1 1/12 check to see that it adds to 12/12!

2 3/12

3 2/12

4 4/12

5 1/12

6 1/12

What is the expected value E(X)?

1(1/12) + 2(3/12) + 3(2/12) + 4(4/12) + 5(1/12) + 6(1/12) = 40/12 ~ 3.3…

Formula for SD:

And the Variance and SD

The sum of X sq times P(X) is 156/12

Mu sq is 1600/144

1872 – 1600 all over 144 is 273/144 for Variance. ~1.4 for SD

NOTE TI simulation page 185 This is VERY useful for making up worksheets, quizzes, and tests.

Now just a little bit on Z Score, p.197. It exists! And we use it! It’s the same formula too and means the same thing – distance of the value from the Expected Value, the mean or center.

(Measurement – mean) divided by standard deviation. Suppose we roll a 6.

(6 – 3.3)/1.4 is a z score of about 1.93. Fairly unusual.

Now let’s discuss Fairness. Clearly the above is NOT fair. If a game is fair, any outcome has an equally likely chance of happening. The game in the example on page 188 is pretty interesting. You roll two dice and add the sum of the faces showing upwards. The probability chart for this is on page 199…the two columns on the right. There are 3 players in the game. Player 1 gets a point for {1, 2, 3, 4}; Player 2 gets a point for {5, 6, 7, 8} and Player 3 gets a point for {9, 10, 11, 12}

There’s a 6 out of 36 chance on any one number for Player 1, a 20/36 chance on numbers for Player 2, and 10/36 for Player number 3 for just rolling the dice and recording a number. Clearly not fair…Player 2 has an advantage.

Now here’s a riff on Problem 3 page 229.

60% of the customers at the Dollar Store pay in cash. Two customers are in line. Use a tree diagram to find the Probability Distribution for X, the number of customers who pay cash checking out right now this time. The values for X are

{0, 1, 2}

Note: a very specific table. General tables and Distributions next

The table looks like:

X P(X)

0 .36

1 .48

2 .16

E(X) = 2(.36) + 1(.48) + 0(.36) = 1.2

S(X) = [4(.36) + 1(.48) + 0(.36)} – (1.2)(1.2) = .48

SD is sqrt .48 ~ .693

Now let’s turn the information into the most abstract representation of all:

**Discrete Random Variable:** a finite number or a countable number of outcomes.

 Like the ones we’ve just done.

**Continuous Random Variable:** inifinitely many variables, situated on a numberline with no gaps or interruptions.

 Like the normal distribution, or Bell Curve.

Which of the following are discrete? Continuous?

The number of eggs received by the shipping department at the local Krogers on a given day.

The number of people marching in the Fourth of July parade downtown Houston.

The measure of voltage for a smoke detector in your kitchen or voltage in general…

The temperature in Houston on a thermometer vs temperature itself

The exact playing time for a given baseball game vs. time itself

The number of actors in a randomly selected movie.

The weight of a randomly selected human vs weight as a physical quantity.

We will continue this discussion in the upcoming videos.

**End Video A**

**7.4 Binomial Random Variables**

We’ll look at the Binomial Distribution here. It’s a big generalized system that covers many, many situations. It is discrete. It is necessary that each situation that we label as Binomial meets the following, know by heart, conditions. (p. 203)

1. There are repeated identical trials that are

2. independent of each other. The preceding doesn’t influence the next.

3. There are only two possible outcomes. (success/failure)

4. P(success) = p; P(failure) = q. p + q = 1 (100%); q = 1 – p

5. x, the rv, is the number of successes in n trials. n – x = # failures.

Here’s the formula for finding the probability of any given number of successes:

This is called the “binomial probability density function” (binompdf)

Let’s look at an example. I have a rescue dog named Toasty. He was a badly treated yard dog that Kim Ogg’s team rescued and passed on to me. He was diagnosed with hemangiosarcoma: a fairly vicious cancer caused by over exposure to the sun. We are 15 months post chemo. He’s had negative tests every three months for a total of 4 trials. He’s got one more in March 2021.

This is binomial. There are repeated identical tests. He tests positive or negative. Last quarter’s test doesn’t affect this quarter’s test. From a testing stand point, testing positive is a success for the test; it found a recurrence of the cancer. So we’ve got 4 failures in a row and are hoping for a fifth. The probability of finding the cancer is .005; and not is .005.

Don’t get into calling a positive a good thing. It’s a math term and medical term not a conversational convention here.

Let’s do an outline of the full table and fill in a couple of probabilities starting with test 1 more than a year ago. Next page

Often we have a situation with **repeated identical trials**. Tossing a free throw (it goes in or it doesn’t), tossing a coin (heads/tails), landing a plane (ok/crash), having a baby (boy/girl), taking a T/F test; hitting the height to be a Navy pilot.

Let’s look at some other ways to calculate binomial probabilities.

What about a coin toss with a fair coin. What’s the probability of getting exactly 3 heads in 5 tosses? What’s p? What’s q? What’s n – x? What’s the combination of 5 things taken 3 at a time doing in the formula?

What about at most 3 heads in 5 tosses?

P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)

**Chapter 7 Popper Question 2**

If we only know p, n, and x, we don’t have enough info to solve the problem.

A. True B. False

Now what about that combination at the beginning of the formula?

Note that we will be using COMBINATIONS when we count outcomes:

Let’s look at the 3 child family with a tree diagram:

And summarize it

BBB

GGG

2B1G

1B2G

The combination of 3 kids taken 2 at a time:



P(3girls) = 3/8, multiplying out the branches. Let’s make a table below and check it out. And check it with the formula, too. Suppose x = # girls

**Chapter 7 Popper Question 3**

The combination factor in the binompdf formula counts up all the ways an event can occur.

A. True B. False

Let’s check that combination factor again with coin tosses n = 5 x = 1 p = ½

Are there mean, variance, and standard deviation? Bet your grade on it.

Summary page 205, bottom, box

For BINOMpdf ONLY

Mean: np

Variance: npq

SD: sqrt npq

Easy and short but of zero value if it’s not a binomial scenario.

And while we are on housekeeping, here are some computational tips.

If p is an unknown, use ½

Read the language very closely. If it says “at most” (n-1) out of n outcomes use the complement rule: 1 – P(n)

TI – let’s learn how to do this efficiently: page 206

If it says at most 5 out of 100…add up 0, 1, 2, 3, 4, 5

“At least 5” out of 10…start your adding probabilities with 5, 6, 7, 8, 9, 10

And learn how to do it in a program or on a calculator. By hand is not fun at all.

EXAMPLE

In a drug study, there is a control group and a group of people not taking the drug. The drug is to help you have girls for children in a 3 child family. These are VERY large groups.

Here is a table for the drug use group. Is it a Probability Distribution Table?

|  |  |
| --- | --- |
| X | P(X) |
|  0 | .120 |
| 1 | .370 |
| 2 | .380 |
| 3 | .130 |

Note the difference from the earlier chart. Suggests the drug might work!

Chapter 7 Popper Question 4

Is there an Expected Value and a Standard Deviation for this distribution?

A Yes B. No

What does the histogram look like?

**Which of the following are binomial experiments?**

Surveying 1000 people and asking them to rate the president on a scale of 1 – 5

Rolling a fair die 50 times

Having kids

Determining whether 12,000 pacemakers are defective or not, one by one

Guessing on a T/F test

Guessing on a test with 5 answer choices per question

Getting past the height requirement for Navy pilot training

BP Problem 1

Bob is a self-proclaimed mentalist who claims he can read minds. To test this, he is given 14 T/F questions.

A. He gets 8 of them right. What is the expected value and is this unusual?

B. He gets 11 of them right. What is the expected value and is this unusual?

C. He gets 2 of them right? EV is? Is this unusual?

BP Problem 2

There is a 0.723 probability that an airplane will land on time at Hobby in a specific set of hours, 10am to 3pm weekdays

A Find the probability that at least 5 out of 6 airplanes arrive on time in the given period of time.

B Find the probability that at most 2 airplanes arrive on time in the given period of time.

C Find that probability that exactly 3 airplanes land on time in the given period of time.

BP Problem 3

Internal surveys show that directory assistance providers give the wrong number 15% of the time. Assume you are testing a provider by making 10 requests. Assume further that this is a very average company and gives wrong answers 15% of the time.

Find the probability of getting one wrong answer. Is this unusual?

Find the probability of getting at most one wrong answer. Is this unusual?

Is the probability really 15% for this company?

BP Problem 4

A study was conducted to determine whether there were significant differences between medical students admitted through special programs and medical students admitted through the regular admissions criteria. It is claimed that the graduation rate for the students admitted through the special programs is 94%.

If 10 students from the special programs are randomly selected, find the probability that at least 9 of them graduated.

Would it be unusual to randomly select 10 and find that 7 graduated? Why or why not?

**End Video B**

**Next up a continuous generalized distribution.**

**7.5 The Normal Curve and just a bit of 7.6**

The standard normal curve the bell curve

symmetric, mound shaped, continuous

Let’s discuss continuous versus discrete again

**For the standard normal curve the mean is zero and**

**the standard deviation is 1.**

It is symmetric about **z** = 0…not x? why?

Probabilities correspond to area under the curve.

Let’s review the Empirical Rule (p. 71) right now with a picture:

Now let’s look at the standard normal probability table.

Given a z-score of 1.28, what is the probability that a measurement is at or below this value?

page 210 rows: 1.2 across to under .08 Area is .8997 at or below this z score.

Now for using the chart with “greater than or equal to”…a version of the complement rule! P(Z >1.32) = 1 - .9066

Or between two measurements! P( .83< Z < 1.03) P(.83 < Z) = .7967 and

P(Z < 1.03) = .8485

Using the table in reverse: from a probability to a z-score:

Page 212 Suppose the P is 61%...look in the table…find .6064 and .6103 look up and across:

.26 and.27 for z scores. Since between can extrapolate to .267 or .268 (closer to .27). This is the sixty-first percentile z score.

In reality, MOST normal curves are NOT standard! How do we rescale to make use of our standard normal chart? With z-scores! All normal curves are proportional and we use the z-score calculation to make them “fit” the table.

Page 213

Let’s look at a normal curve with mean 5 and sd .75. What’s the area between 1 sd above and below the mean? Empirical Rule. What are those measurements for THIS curve. How will we use z-score to discover this on the chart for the standard normal distribution?

From another source – TI83 instructions for

Areas between two bounds:

2nd VARS [2: normal cdf(left z score, right z score)]

Normal Distributions: together!

The Precision Scientific Instrument Company manufactures thermometers. To check the accuracy, they test the thermometers in freezing water and make sure it registers 0 degrees F. Of course some are high and some are low. Assume there is a standard deviation of 1 degree F. Find the area and show it on a standard normal curve!

What is the probability that the reading is less than 1.58°?

You should get 94.29%

What is the probability that the reading is above −1.23°?

You should get .8907

What is the probability that the reading is between −2° and 1.5°?

You should get 91.04%

Working backwards in the chart:

Find the temperature associated with the 95th percentile. z = 1.645

How does this work?

Find the temperatures separating the bottom 2.5% and the top 2.5%

These are called tolerances. (−1.96 and 1.96 for z’s).

How does this work?

**Fill in the blanks:**

About \_\_\_\_\_\_\_\_\_% of the area is within 1 standard deviation of the mean

About \_\_\_\_\_\_\_\_\_% of the area is within 2 standard deviations of the mean

About \_\_\_\_\_\_\_\_\_% of the area is within 3 standard deviations of the mean

Enrichment: The Triangle Distribution. Manufacturing fill problems



Here is a probability distribution. Find the value of c.

Show the probability that *x* is between 0 and 3.

Show the probability that *x* is between 2 and 9.

UNIFORM Distribution

Sketch one and find c for a horizontal axis from 0 to 8

Homework problem 1

Air Force ejection seats are designed for people weighing between 140 lb and 211 lb. Women’s weights are normally distributed with a mean of 143 lb and a standard deviation of 29 lb. What percentage of women have weights in those limits?

I’ll get you started. What’s X?

Homework problem 2

The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days.

A woman wrote to Dear Abby claiming that she gave birth 290 days after a brief visit with her husband who was fleet Navy and ship bound else. Is this credible?

Again, we’ll start together

Premature is being born in the 4th percentile of length of pregnancy…what length of time is this?

Hint: use the z score formula to back solve for the days because you’ll know the z score but not the X from the table

Finishing up the poppers

**Chapter 7 Popper Question 5**

The z score associated with the 55 percentile is approximately:

A. 0.15 B. 0.21

**Chapter 7 Popper Question 6**

To work a probability problem with a Normal Distribution that is NOT the Standard Distribution, you set the problem up with X and then convert it to Z using the z score formula and the non-standard mean and standard deviation.

A. True B. False

**Chapter 7 Popper Question 7**

To find the P(Z > .91), use the Complements Rule: 1 – P(Z < .91)

A. True B. False

**Chapter 7 Popper Question 8**

All normal distributions are proportional and the probability that X is between 1 and – 1 is 68% for them all.

A. True B. False

Chapter 7 Popper Question 9

Given N(7, 2), the Probability that X is between 3 and 11 is 95%.

A. True B. False

Now for a brief summing up:

We looked first at Discrete Distributions – both individual ones that pertained to just one problem and the Binomial Distribution. In all cases, there was a mean and a standard deviation, and the distribution probabilities added up to one.

We used tree diagrams and area charts and frequency data to find the probabilities and then summarized them in tables with a finite number of rows.

We can have negative values for X even though we didn’t work with any of these. These are in the first column of the distribution. The second column is made up of probabilities and these are all positive numbers between and including 0 and 1.

The Binomial Distribution is a general case distribution. You need

 Repeated identical trials, n of them

 Independent trials

 Exactly two outcomes (P and Q)

 p + q = 1

 x is the number of successes and n – x is the number of failures

The formulas for an individual discrete distribution m and sd are DIFFERENT than those for the Binomial Distribution.

We looked at 3 Continuous Distributions: Normal, Triangle, and Uniform. These are all in Quadrants 1 and 2 and over an interval of the horizontal axis. Finding probabilities is a matter of finding the area between, to the left, or the right of a given value. Less than or equal to has the SAME probability as a straight less than because adding in a line does NOT add area. This is not usually true with Discrete

Distributions. Continuous Distributions have the area under the curve = 1 and they each have a mean and a standard deviation too. N(0,1) is the Standard Normal Distribution and is generalized to cover all scenarios that are normally distributed.

Summary of Chapter 7: A 9 question popper, no essays, 2 homework problems to complete in 7 C plus from the book: 2, 6, 11, 18 (only a and f).