Math 3307 Chapter 7 Video Scripts plus Popper

Chapter 7 – Random Variables and Probability Distributions 7.1 – 7.3

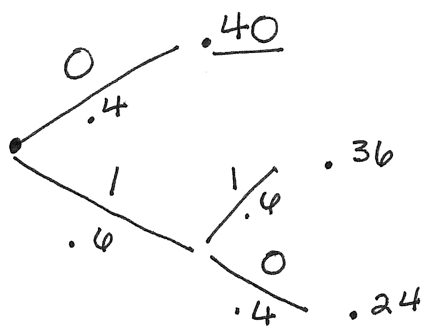
What is a Random Variable?

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:

X	P(X)
0	.40
1	.24
2	.36

$\Sigma P(X) = 1.00$ 100%

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. 189. 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

$$\sigma^2 = \sum [x^2 \cdot P(x)] - \mu^2 \qquad [\text{or } E(X^2)] \text{ same}$$

$$[0^2 \cdot .4 + 1^2 \cdot .24 + 2^2 \cdot .36] - .95^2$$

$$1.68 - .9025 = .7775$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{.7775} \cong .88$$

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
2	3/12
3	2/12
4	4/12
5	1/12
6	1/12

check to see that it adds to 12/12! ie 100% or 1

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:

$$\sqrt{\sigma^2} = \sqrt{\sum (x^2 \cdot P(x)) - E(x)^2}$$

$$\left(\left[1\left(\tfrac{1}{144}\right) + 4\left(\tfrac{9}{144}\right) + \cdots + 36\left(\tfrac{1}{144}\right) - 3.3^2 \right) \right.$$

$\left(\frac{40^2}{144}\right)$ better!

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.

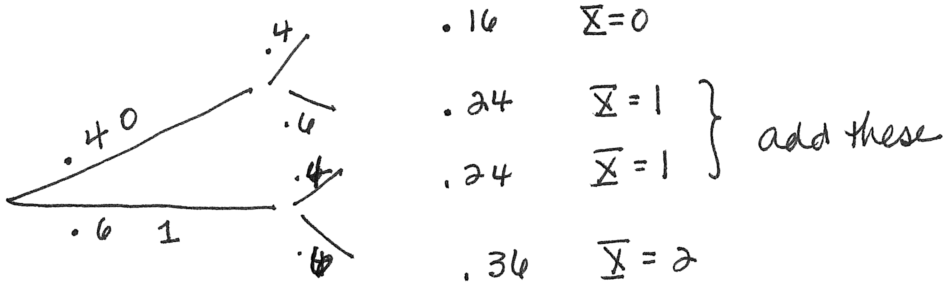
$$Z = \frac{6 - 3.3}{1.4} \sim 1.93$$ nearly 2

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. D

The number of people marching in the Fourth of July parade downtown Houston. D large but still

The measure of voltage for a smoke detector in your kitchen or voltage in general…
on the screen 1 number voltage itself is cts as long as there's power

The temperature in Houston on a thermometer vs temperature itself
D C

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

The number of actors in a randomly selected movie.
D

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

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 .995.

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

$$P(X=5) = \binom{5}{5} .005^5 .995^0$$

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

X	$P(X)$
0	treatment ends
1	$\binom{5}{1}\,.005^1\,.995^4$
2	$\binom{5}{2}\,.005^2\,.995^3$
3	$\binom{5}{3}\,.005^3\,.995^2$
4	$\binom{5}{4}\,.005^4\,.995^1$
5	$\binom{5}{5}\,.005^5\,.995^0$

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?

HHHTT $\quad .5^3 .5^2$
HHTTH $\quad .5^3 .5^2$
HTTHH
TTHHH
HTHTH
⋮

$\binom{5}{3}$ counts up all the ways the event can happen for you in one calculation

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

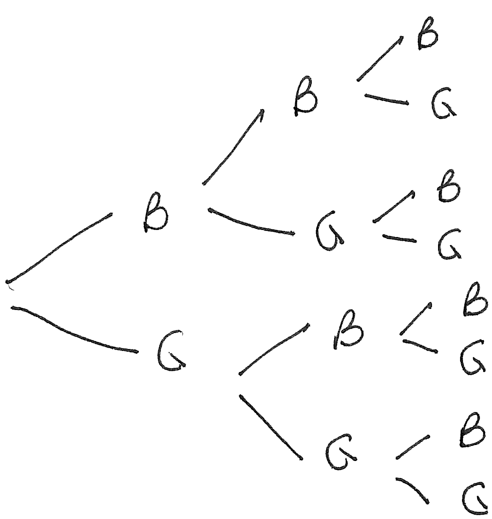
If we only know p, n, and x, we don't have enough info to solve the problem.

A. True B. False

Now what about that combination at the beginning of the formula? It counts up all the ways

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

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



BBB
BBG 2B1G
BGB 2B1G
BGG
GBB 2B1G
GBG
GGB
GGG

And summarize it

BBB 1 way
GGG 1 way
2B1G 3 ways
1B2G 3 ways

The combination of 3 kids taken 2 at a time:

$${}_3C_2 = \binom{3}{2} = \frac{3!}{2!1!} = 3$$

2B1G 3 ways
2G1B 3 ways

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

B/G 50/50

X	
0	.125
1	.375
2	.375
3	.125
	1.00 100%

0	BBB	$1/2^3$	.125
1	GBB	$1/2^3 \times 3$ ways	.125×3
2	GGB	$1/2^3 \times 3$ ways	
3	GGG	$1/2^3$	

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

$\binom{5}{1} .5^4 .5^1$

$\frac{5!}{4!1!} = 5$ ways

$\left.\begin{array}{l} \text{HTTTT} \\ \text{THTTT} \\ \text{TTHTT} \\ \text{TTTHT} \\ \text{TTTTH} \end{array}\right\}$ 5 ways

Are there mean, variance, and standard deviation? Bet your grade on it.
Summary page 205, bottom, box

$\binom{5}{1}$ $n = 5$ $p = 1/2$ $q = 1/2$

For BINOMpdf ONLY

Mean:	np
Variance:	npq
SD:	sqrt npq

$np \quad 5(1/2) = 2\,1/2$

$npq \quad 5(1/4) = 1\,1/4$

$SD \quad \sqrt{5/4} = \sqrt{5}/2$

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

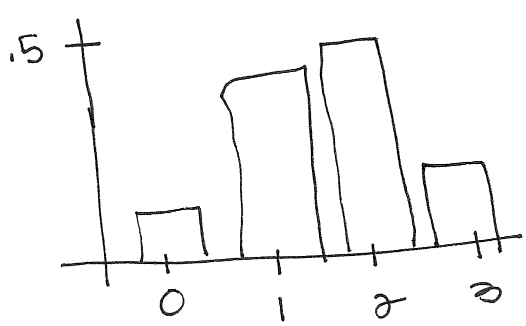
Note the difference from the earlier chart. Suggests the drug might work!

★ 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

nope

try like/dislike approve / disapprove

Rolling a fair die 50 times

6 faces try coin flip

Having kids

yep

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

bad/good yep

Guessing on a T/F test

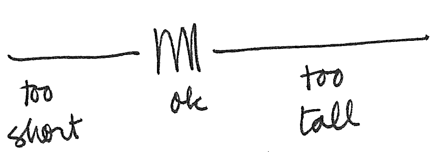
yep

Guessing on a test with 5 answer choices per question

nope

Getting past the height requirement for Navy pilot training

yep



2nd video B