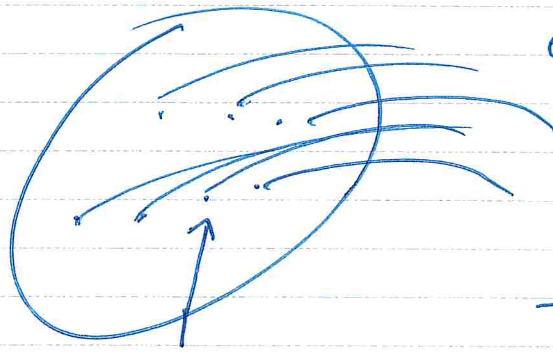
STAT 202 Means / Std of RV's

Review - Random Variables

Needed Concepts include Sample Space, Outcomes
Value of random variable for each outcome
Value must be numeric



outcomes

each outcome is mapped to a value
Value must be numeric

Each outcome must map to one and only one number

Two outcomes can map to same number
But same outcome can't map to two numbers (vertical line test)

Binomial Random Variable

Example - Flip a coin 10 times (fair coin)
Example question: What is the probability of 5 or fewer heads,

This question set up by default in StatCrunch

Show

Using Binomial Calculator

Repeat trials $\boxed{n}$ times, each probability $\boxed{p}$ of "success". Question answered "What is the probability of $\boxed{\text{expression for } X}$"

Where X is the binomial random variable (count of successes in the n trials)

$\boxed{n}$ and $\boxed{p}$ are called parameters

Comparing Calculators eg Normal or Binomial

Input $\boxed{\text{Parameters}}$ and $\boxed{\text{Expression for X}}$

Get: Probability of expression for X

Sometimes go backwards (e.g. Normal calculator you can plug in $\boxed{\text{Parameters}}$ and ~~Express~~ $\boxed{\text{Probability of expression for X}}$

And it fills in some of the blanks for $\boxed{\text{expression for X.}}$

The binomial RV

maps outcomes (eg SFSSFFFSF) to numbers of successes eg (4),

outcome $\mapsto$ number for each outcome

Conditions for applying Binomial distribution/RV/calculator

1) must be a fixed number $\boxed{n}$ of trials
2) Outcome of each trial binary (success/failure)
3) Probability of success in each trial i $\boxed{p}$ for all trials
4) Trials are independent

Let's revisit Binomial Vs Normal

Binomial is Discrete
Normal is Continous

Discrete was defined as having a finite number of values (we will update that def'n soon).

Continous was defined as ~~having~~ having a range of possible values that is an interval in the real numbers

Here's an example of one random variable, similar to the Binomial Random variable that does not fit in either category.

The Geometric Distribution: Run trials exactly like the ones in the Binomial distribution (these trials are called Bernoulli Trials)

Binomial: Run exactly n trial and ~~count~~ count the number of successes

Geometric: Run trials, stop when there is a success count either the number of trials OR the number of failures before a success

Statcrunch will do both. Obviously they are related:

$$(\#\text{ of failures}) + 1 = (\#\text{ of trials})$$

last success

Show geometric calculator

Problem with Geometric Calculator

What is the probability that you will have to toss a fair coin 5 or more times to get your first head? (instead of all tails)

Question Does the Geometric Distribution look more discrete or more continuous

But think. Can you throw a coin 100 times a get all tails? How about 1000 times, Probability is small but not zero.

All ~~positive~~ ~~nonnegative~~ integers starting with 1 and up (for counting trials) and starting with 0 and up (for counting failures) are possible.

This random variable has an infinite number of values

~~It can ~~even~~ conceivably take on a value of ∞ — possible but that outcome has probability 0~~
~~must be ignored This value is not a number which we have to ignore it to consider the function a random variable.~~

Many advanced books define discrete as one whose values can be put in a list

All ~~finite~~ Random variables with finite number of values can but so can the geometric ones. Here is the list 1, 2, 3, ...

the list can be infinite.

You can't do this with an interval of real numbers Ask me!

Mean of Discrete Random variable

Value	x_1	x_2 ...
Prob	p_1	p_2 ...

mean:

Binomial Toss 3 coins

Value (# of heads)	0	1	2	3
Probability	1/8	3/8	3/8	1/8

3/8th of the time it is 1
3/8th of the time it is 2
1/8th of the time is 0
1/8th of the time it's 3

Instead of averaging the data
We want to weight each value with the fraction of times it occurs

$$0\cdot\tfrac{1}{8} + 1\cdot\tfrac{3}{8} + 2\cdot\tfrac{3}{8} + 3\cdot\tfrac{1}{8} = \sum x_i p_i$$

$$= \tfrac{3}{8} + \tfrac{6}{8} + \tfrac{3}{8} = \tfrac{12}{8} = \tfrac{3}{2} = 1.5$$

Problem: a 4 sided die is thrown

red $\mapsto$ \$10
green $\mapsto$ \$1
blue $\mapsto$ −1
~~green~~ white $\mapsto$ −\$30

thats not quite enough info. Let's say all ~~eq~~ outcomes are equally likely (We need probabilities)

Write a probability table
What is the mean of this RV?

It's easy when all outcomes are equally likely, just average (mean) the values.

Two Random variables on the same sample space can be added, subtracted, etc

X	Y
red $\mapsto 5$	red $\mapsto 3$
green $\mapsto 4$	green $\mapsto -1$
blue $\mapsto 3$	blue $\to 0$
white $\mapsto 0$	white $\to 1$

$X+Y$

red $\mapsto 8$
green $\mapsto -3$
blue $\mapsto 3$
white $\mapsto 1$

μ stands for mean

$$\mu_{X+Y} = \mu_X + \mu_Y$$

One die

value	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

mean $\frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$

two die (monopoly) $\mu_{X_1+X_2} = \frac{7}{2} + \frac{7}{2} = 7$

If a and b are numbers and X is a random variable

$$\mu_{a+bX} = a + b\mu_X$$

~~Random Variable represents temperature in °F~~

X = temperature of a random person's refigerator chosen at ~~random~~ random (equally likely)

Bob's frig ⟼ 34°
Sally's frig ⟼ 37°
Jo's frig ⟼ 35°
Jane's frig ⟼ 38°

Conversion: $X_{new} = \frac{5}{9}(X_{old} - 32)$

X_{old} °F
X_{new} °C

Last formula says we can "average temperatures" (take mean of RV) in °F then convert

OR ~~average~~ convert temperatures then average converted temperatures

Both give same result!

Don't have to have equally likely outcomes though

$$\underbrace{\mu_{a+bX}}_{\text{new random variable that's converted temperature}} = a + b\mu_X$$

mean of new RV

mean random variable that ~~...~~ old temperat then convert,

$$\mu_{x-y} = \mu_x - \mu_y$$

$$\mu_{ax} = a\,\mu_x$$

$$\mu_{b+x} = b + \mu_x$$

~~Example~~ .

Standard deviation: Toss 3 coins

X = number of heads, standard deviation

Values	0	1	2	3
Prob	1/8	3/8	3/8	1/8

Mean = 3/2

$$\sqrt{\sum (X_i - \mu_x)^2 p_x}$$

Compare with data

$$\sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

↑ (or $\frac{1}{n}$) equally likely all have same weight

↑ mean needs adjustment because we are using sample mean $\bar{x}$ instead of μ_x pop. mean

Do this for 3 coin tosses

$$\sqrt{(0-3/2)^2\cdot 1/8+(1-3/2)^2\cdot 3/8+(2-3/2)^2\cdot 3/8+(3-3/2)^2\cdot 1/8}$$

$$=\sqrt{9/4\cdot 1/8+\frac{1}{4}\cdot 3/8+\frac{1}{4}\cdot 3/8+9/4\cdot 3/8}$$

$$=\sqrt{\frac{9+3+3+9}{32}}=\sqrt{\frac{24}{32}}=\sqrt{\frac{12}{16}}=\sqrt{\frac{3}{4}}$$

Here's a hint: For binomial problems

mean $= np$
std dev. $=\sqrt{np(1-p)}$] just Binomial.

$1-p$ is sometimes called q

So $1-p=q$ $\quad p+q=1$

std dev $=\sqrt{npq}$

$$=\sqrt{3\cdot 1/2\cdot 1/2}=\sqrt{3/4}$$

To take mean and standard dev of geometric series you'd need

$\sum_{n=1}^{\infty} x_i p_i$ ~~an~~ infinite series

For continuous Random variables, you'd need integrals $\int_{-\infty}^{\infty} x\,p(x)\,dx$ Both need calculus

Formulas for standard deviation
(actually variance, square of std dev.
find variance, then take sqare root)

$$\sigma^2_{X+Y} = \sigma_X^2 + \sigma_Y^2$$ (only when X and Y are independent)

Difference

$$\sigma^2_{X-Y} = \sigma_X^2 + \sigma_Y^2$$ (independent)

$\hookrightarrow$ not a typo.

$$\sigma^2_{b_1X_1 + b_2X_1} = b_1^2 \sigma^2_{X_1} + b_2^2 \sigma^2_{X_2}$$

$$\sigma^2_{a+bX} = b^2 \sigma^2_X \qquad (1)^2 = 1 \qquad (-1)^2 = 1$$

General addition rule
depends on something called correlation ρ
(when independent $\rho = 0$)

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$$

HW #17