No activity slide, but you'll want a copy or pen + paper today

Announcements:

- Real World 2 due Monday
- New post on Ed - how grades are made + estimated boundaries

Joint Distributions

CSE 312 Spring 21
Lecture 21

Our Goal

Set a target – I want my margin of error to be 2%. That is, at least 95% of the time, your poll's estimate of the fraction of people in favor will be within 2 percentage points of the true value.

So…how many people are you going to need to interview?

Using the CLT

P / P

What are we looking for? Well we have a margin of error:

$$\mathbb{P}(p - .02 \le \hat{p} \le p + .02) \ge .95$$

That says we're within the 2% margin of error at least 95% of the time.

What is that probability? Well let's setup to use the CLT. Subtract the expectation and divide by the standard devation.

$$\mathbb{P}\left(\frac{p-.02-p}{\sqrt{p(1-p)/n}} \le \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \le \frac{p+.02-p}{\sqrt{p(1-p)/n}}\right) \ge .95$$

Handling $\sqrt{p(1-p)}$

Justification 1: If we make a mistake, we want it to be making n bigger. (since we're trying to say "take n at least this big, and you'll be safe").

The bigger the standard deviation, the bigger n will need to be to control it. So assume the biggest possible standard deviation.

Justification 2:

As $\sqrt{p(1-p)}$ gets bigger, the interval gets smaller (it's in the denominator), so assuming the biggest value of $\sqrt{p(1-p)}$ gives us the most restricted interval. So no matter what the true interval is we have a subset of it. And if our probability is at least .95 then the true probability is at least .95.

What's the maximum of $\sqrt{p(1-p)}$?

Worst value of p

Calculus time!

Set $\frac{d}{dp}\sqrt{p-p^2} = 0$

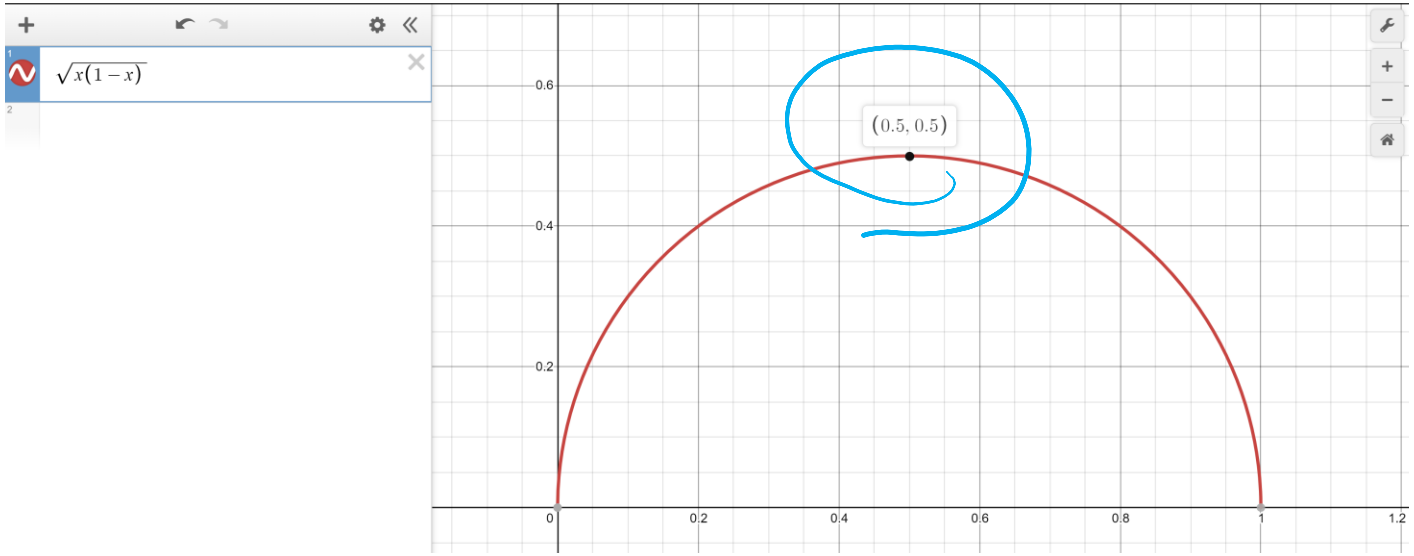
$\frac{1}{\sqrt{p-p^2}}(1-2p) = 0$

$1 - 2p = 0 \rightarrow p = 1/2$

Second derivative test will confirm $p = \frac{1}{2}$ is a maximizer

Or just plot it.

$\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)} = \sqrt{1/4}. = \frac{1}{2}$



Doing the algebra

$$\mathbb{P}\left(\frac{p-.02-p}{\sqrt{p(1-p)/n}} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq \frac{p+.02-p}{\sqrt{p(1-p)/n}}\right)$$

$$\approx \mathbb{P}\left(\frac{-\sqrt{n}\cdot .02}{\sqrt{p(1-p)}} \leq Z \leq \frac{\sqrt{n}\cdot .02}{\sqrt{p(1-p)}}\right)$$ by CLT; $Z \sim \mathcal{N}(0,1)$

$$\geq \mathbb{P}\left(\frac{-\sqrt{n}\cdot .02}{\sqrt{1/4}} \leq Z \leq \frac{\sqrt{n}\cdot .02}{\sqrt{1/4}}\right)$$

$$= \mathbb{P}\left(-.04\sqrt{n} \leq Z \leq .04\sqrt{n}\right)$$

$$= \Phi\left(.04\sqrt{n}\right) - \left(1 - \Phi\left(.04\sqrt{n}\right)\right) = 2\Phi\left(.04\sqrt{n}\right) - 1$$

$$2\Phi\left(.04\sqrt{n}\right) - 1 \geq .95 \rightarrow \Phi\left(.04\sqrt{n}\right) \geq \frac{1.95}{2}$$

Using the Φ-table

$\Phi(.04\sqrt{n}) \geq .975$

ppl to sample

Φ-table says:

$.04\sqrt{n} \geq 1.96$

$\sqrt{n} \geq 49$

$n \geq 2401$. gives 95% confidence interval of +/- 2%.

I.e. 95% of the time, our poll gets a value within 2% of the true value.

CLT Wrap-up

It's not ideal that we had an approximation symbol in the middle (that "≥" isn't really a guarantee at this point, it's an approximation)
Observation 1: with our current tools, we wouldn't get an answer in a reasonable amount of time.

But using a binomial would be even harder.

As n changes, the distribution of a binomial changes. Wolfram alpha isn't even enough here (unless you have 2+ hours to spare to guess and check values).
You need a computer program to get the exact value.

You're computer scientists! You can write that program. But it takes time.

Observation 2: if you need an absolute guarantee, you won't get one. The tool you want is a "concentration inequality/tail bound." We'll see those next week.

CLT Wrap-up

Use the CLT when:

1. The random variable you're interested in is the sum of independent random variables.

2. The random variable you're interested in does not have an easily accessible or easy to use pmf/pdf (or the question you're asking doesn't lend it self to easily using the pmf/pdf)

3. You only need an approximate answer, and the sum is of at least a moderate number of random variables.

Joint Distributions

Today

A somewhat out-of-place lecture.

When we introduced multiple random variables, we've always had them be independent.

Because it's hard to deal with non-independent random variables.

Today is a crash-course in the toolkit for when you have multiple random variables and they aren't independent.

Going to focus on discrete RVs.

Joint PMF, support

For two (discrete) random variables X, Y their joint pmf

$$f_{X,Y}(x, y) = \mathbb{P}(X = x \cap Y = y)$$

When X, Y are independent then $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

$$\mathbb{P}(X=x \cap Y=y) = \mathbb{P}(X=x)\,\mathbb{P}(Y=y)$$

Examples

Roll a blue die and a red die. Each die is 4-sided. Let X be the blue die's result and Y be the red die's result.

Each die (individually) is fair. But not all results are equally likely when looking at them both together.

$f_{X,Y}(1,2) = 3/16.$

$f_{X,Y}$	$X=1$	$X=2$	$X=3$	$X=4$
$Y=1$	1/16	1/16	1/16	1/16
$Y=2$	3/16	0	0	1/16
$Y=3$	0	2/16	0	2/16
$Y=4$	0	1/16	3/16	0

Marginals

What if I just want to talk about X?

Well, use the law of total probability:

$\mathbb{P}(X = k) = \sum_{\text{partition } \{E_i\}} \mathbb{P}(X = k | E_i)\mathbb{P}(E_i)$

and use E_i to be possible outcomes for Y For the dice example

$\mathbb{P}(X = k) = \sum_{\ell=1}^{4} \mathbb{P}(X = k \,| Y = \ell)\mathbb{P}(Y = \ell)$

$= \sum_{\ell=1}^{4} \mathbb{P}(X = k \cap Y = \ell)$

$f_X(k) = \sum_{\ell=1}^{4} f_{XY}(k, \ell)$

$f_X(k)$ is called the "marginal" distribution for X (because we "marginalized" Y)

It's the same pmf we've always used; the name emphasizes we have gotten rid of one of the variables.

Marginals

$$f_X(k) = \sum_{\ell=1}^{4} f_{XY}(k,\ell)$$

So

$$f_X(2) = \frac{1}{16} + 0 + \frac{2}{16} + \frac{1}{16} = \frac{4}{16} = \frac{1}{4}$$

$f_{X,Y}$	X=1	X=2	X=3	X=4	
Y=1	1/16	1/16	1/16	1/16	1/4
Y=2	3/16	0	0	1/16	1/4
Y=3	0	2/16	0	2/16	1/4
Y=4	0	1/16	3/16	0	1/4
	1/4	1/4	1/4	1/4	

Different dice

4 sided

Roll two fair dice independently. Let U be the minimum of the two rolls and V be the maximum

Are U and V independent?

Write the joint distribution in the table

What's $f_U(z)$? (the marginal for U)

pollev.com/cse312

$f_{U,V}(2,2) = \frac{1}{16}$

$f_{U,V}$	U=1	U=2	U=3	U=4
V=1	$\frac{1}{16}$	0	0	0
V=2	$\frac{2}{16}$	$\frac{1}{16}$	0	0
V=3	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	0
V=4	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

5/16

Different dice

Roll two fair dice independently. Let U be the minimum of the two rolls and V be the maximum

$$f_U(z) = \begin{cases} \frac{7}{16} & \text{if } z = 1 \\ \frac{5}{16} & \text{if } z = 2 \\ \frac{3}{16} & \text{if } z = 3 \\ \frac{1}{16} & \text{if } z = 4 \\ 0 & \text{otherwise} \end{cases}$$

$f_{U,V}$	U=1	U=2	U=3	U=4
V=1	1/16	0	0	0
V=2	2/16	1/16	0	0
V=3	2/16	2/16	1/16	0
V=4	2/16	2/16	2/16	1/16

3/16

Joint Expectation

Expectations of joint functions

For a function $g(X,Y)$, the expectation can be written in terms of the joint pmf.

$$\mathbb{E}[g(X,Y)] = \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} g(x,y)\cdot \boxed{f_{XY}(x,y)}$$

$\mathbb{P}(X=x \cap Y=y)$

This definition hopefully isn't surprising at this point (it's the value of g times the probability g takes on that value), but it's good to

Conditional Expectation

Waaaaaay back when, we said conditioning on an event creates a new probability space, with all the laws holding.

So we can define things like "conditional expectations" which is the expectation of a random variable in that new probability space.

$$\mathbb{E}[X|E] = \sum_{x \in \Omega} x \cdot \mathbb{P}(X = x|E)$$

$$\mathbb{E}[X|Y = y] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|Y = y)$$

Conditional Expectations

All your favorite theorems are still true.

For example, linearity of expectation still holds

$$\mathbb{E}[(aX + bY + c)\,|E] = a\mathbb{E}[X|E] + b\mathbb{E}[Y|E] + c$$

Law of Total Expectation

Let $A_1, A_2, \dots, A_k$ be a partition of the sample space, then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i]\mathbb{P}(A_i)$$

Let X, Y be discrete random variables, then

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y = y]\mathbb{P}(Y = y)$$

$$\mathbb{P}(X=x) = \sum_{y \in \Omega} \mathbb{P}(X=x | Y=y)\mathbb{P}(Y=y)$$

Similar in form to law of total probability, and the proof goes that way as well.

LTE

You will flip 2 (independent, fair coins). Call the number of heads X. Then (independently of the coin flips) draw a geometric random variable Y from the distribution $\mathrm{Geo}(X+1)$.

What is $\mathbb{E}[Y]$?

LTE

You will flip 2 (independent, fair coins). Call the number of heads X. Then (independently of the coin flips) draw a geometric random variable Y from the distribution $\mathrm{Geo}(X+1)$.

What is $\mathbb{E}[Y]$?

$\mathbb{E}[Y]$

$= \mathbb{E}[Y|X=0]\mathbb{P}(X=0) + \mathbb{E}[Y|X=1]\mathbb{P}(X=1) + \mathbb{E}[Y|X=2]\mathbb{P}(X=2)$

$= \mathbb{E}[Y|X=0] \cdot \frac{1}{4} + \mathbb{E}[Y|X=1] \cdot \frac{1}{2} + \mathbb{E}[Y|X=2] \cdot \frac{1}{4}$

$= \frac{1}{0+1} \cdot \frac{1}{4} + \frac{1}{1+1} \cdot \frac{1}{2} + \frac{1}{2+1} \cdot \frac{1}{4} = \frac{7}{12}.$

Analogues for continuous

Everything we saw today has a continuous version.

There are "no surprises"– replace pmf with pdf and sums with integrals.

	Discrete	**Continuous**
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
Conditional PMF/PDF	$p_{X\mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_{x} x p_{X\mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X\mid Y}(x \mid y)\,dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Covariance

We sometimes want to measure how "intertwined" X and Y are – how much knowing about one of them will affect the other.

If X turns out "big" how likely is it that Y will be "big" how much do they "vary together"

Covariance

$$\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Covariance

$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$

That's consistent with our previous knowledge for independent variables. (for X, Y independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\mathrm{Var}(X+Y)$?

Covariance

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\mathrm{Var}(X+Y)$?

$\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = 1 - 0^2 = 1$

$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$\mathbb{E}[XY] = \frac{1}{2} \cdot (-1 \cdot 1) + \frac{1}{2}(1 \cdot -1) = -1$

$\mathrm{Cov}(X,Y) = -1 - 0 \cdot 0 = -1.$

$\mathrm{Var}(X+Y) = 1 + 1 + 2 \cdot -1 = 0$