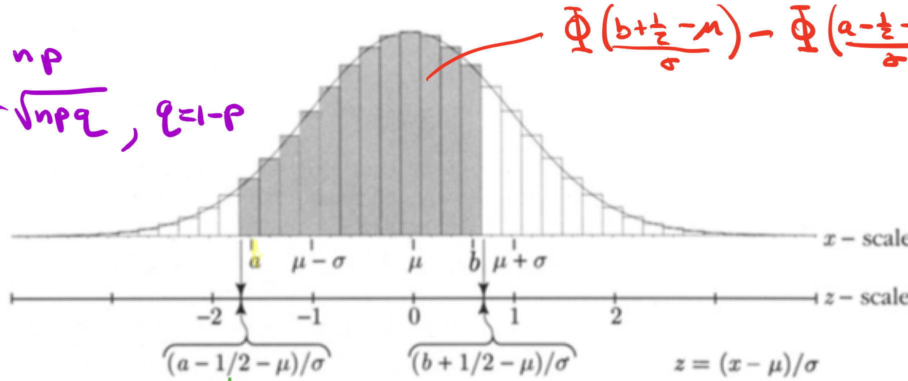
Stat 134 lec 6

Last time sec 2.2 Normal Approx to binomial

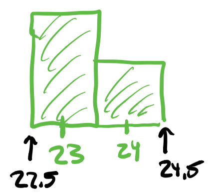
$\mu = np$

$\sigma = \sqrt{npq}$, $q = 1-p$

$\Phi\left(\frac{b+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{a-\frac{1}{2}-\mu}{\sigma}\right)$



why $\frac{1}{2}$?



Continuity correction

We are approximating a discrete distribution (binomial) by a continuous one (normal)

ex

Suppose that each of 300 patients has a probability of 1/3 of being helped by a treatment independent of its effect on the other patients. Find approximately the probability that more than 134 patients are helped by the treatment. (Be sure to use the continuity correction. You will not receive full credit otherwise)

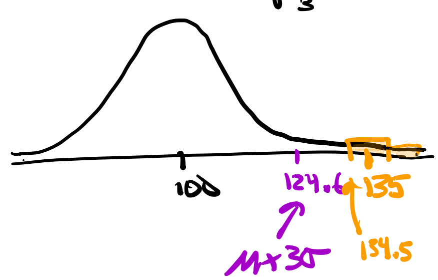
$n = 300$

$p = 1/3$

$\mu = np = 300\left(\frac{1}{3}\right) = 100$

$\sigma^2 = npq = 300\left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{200}{3}$

$\sigma = \sqrt{\frac{200}{3}} = 8.2$



X = # patients helped

$X \sim Bin(300, \frac{1}{3}) \approx N(\underset{\mu}{100}, \underset{\sigma^2}{\frac{200}{3}})$

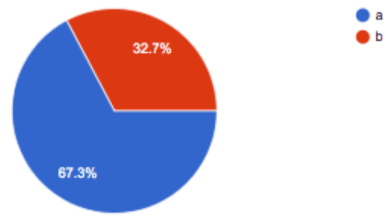
$P(X \geq 135) \approx 1 - \Phi\left(\frac{134.5 - 100}{\sqrt{\frac{200}{3}}}\right) \approx 0$

Today
1. go over student answers to concept test from last time
2. finish sec 2.2
3. sec 2.4 Poisson approximation (skip sec 2.3)

. A fair coin is tossed, and you win a dollar if there are exactly 50% heads. Which is better?

a 10 tosses

b 100 tosses

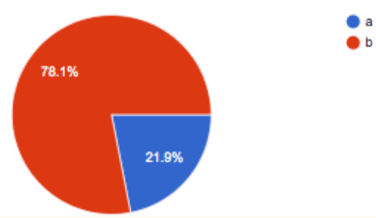


b	100 tosses are better because according to the law of average as the number of tosses increases you are more likely to be closer to 50%.
a	Probability of getting exactly 50 heads out of 100 is lower than 5 out of 10. We can prove this with induction. The base case would be 1 head out of 2. Then we can prove for an arbitrary k flips.
a	If you just think about it logically, there are more possible n's As N grows—with N=100, you could get 49 heads, 48 heads, etc, and all of those are pretty relatively likely. But with N=10, each n is more likely, since there are fewer possibilities.

A fair coin is tossed, which is more likely:

a between 4 and 6 heads in 10 tosses

b between 40 and 60 heads in 100 tosses



b	Now the question, although phrased in terms of counts, is asking for the probability that the percentage of heads is between 40% and 60%. The spread around the average for the percentage falls as the number of trials increases.
a	Actually I'm not sure Bc I don't have a calculator to see the actual areas.

For $n=10$ $\sigma = \sqrt{10 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 1.6$

For $n=100$ $\sigma = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$

$n=10$ $[3.5, 6.5] \approx \overset{5}{\mu} \pm 1\overset{1.6}{\sigma} \approx 68\%$

$n=100$ $[39.5, 60.5] \approx \overset{50}{\mu} \pm 2\overset{10}{\sigma} \approx 95\%$

So option b more likely.

(2) Sec 2.2 Normal approximation to the binomial distribution

2 questions

(1) How do we write μ and σ in terms of n, p to match the normal distribution $N(\mu, \sigma^2)$ with the binomial distribution?

$\mu = np$

$\sigma^2 = npq$

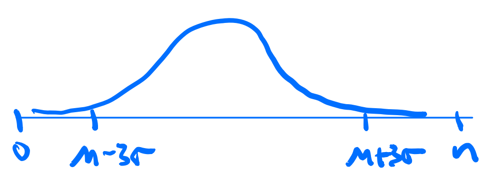
(2) For what n, p is it ok to approximate $Bin(n,p)$ by a normal distribution $N(\mu, \sigma^2)$.

$n \geq 20$ since for fixed p, the binomial is more normal as $n \uparrow$

Outcomes of $Bin(n,p)$ are $0, 1, 2, \ldots, n$

All data is between $\mu \pm 3\sigma$ so we require

$\mu - 3\sigma > 0$ and $\mu + 3\sigma < n$



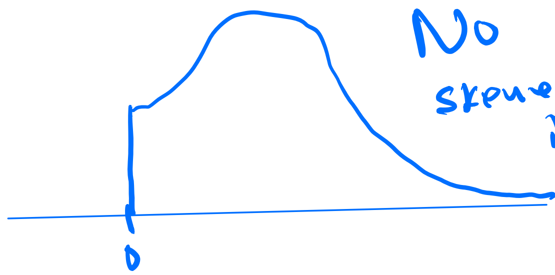
ex Can we approx Bin$(20, \frac{1}{10})$ by the normal?

$n = 20$ ✓

$$\mu - 3\sigma = 20\left(\frac{1}{10}\right) - 3\sqrt{20\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)} = 2 - 4 = -2$$

< 0

✗



No

skewed normal is bad,

$\mu + 3\sigma = 2 + 4 = 6 < 20$ ✓

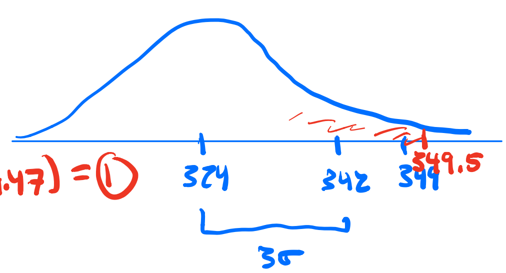
. (3 pts) Southwest overbooks its flights, knowing some people will not show up. Suppose for a plane that seats 350 people, 360 tickets are sold. Based on previous data, the airline claims that each passenger has a 90% chance of showing up. **Approximately**, what is the chance that at least one empty seat remains? (There are no assigned seats.)

$\mu = np = 360(.9) = 324$

$\sigma = \sqrt{npq} = \sqrt{360(.9)(.1)} = 5.7$

X = # people who show up

$P(X \leq 349)$



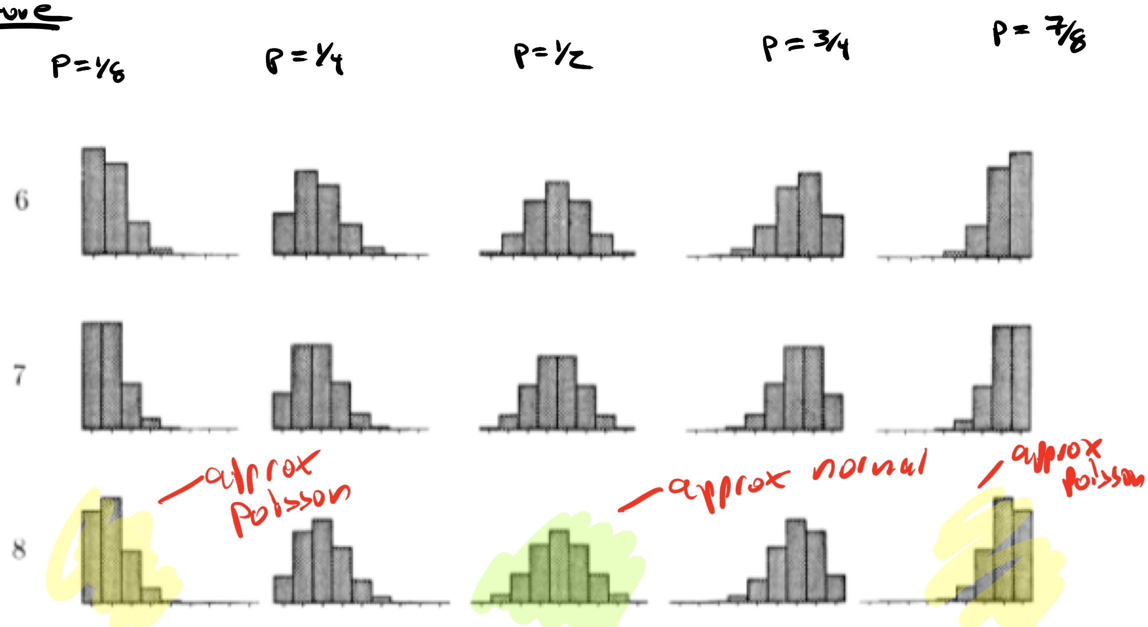
$$\Phi\left(\frac{349.5 - \mu}{\sigma}\right) \approx \Phi(4.47) = 1$$

③ Sec 2.4 (skip 2.3) <u>Poisson approx to Binomial</u>

The normal approximation has almost 100% of data $\pm 3\sigma$ from the mean μ. For this reason we approximated the binomial w/ the normal only when $\mu \pm 3\sigma$ is between 0 and n.

For cases when p is small (or p is close to 1) and n is large, we approximate $Bin(n,p)$ by $Pois(\mu = np)$

<u>Picture</u>



<u>Def^n</u> Poisson(μ) (written Pois(μ))

$$P(k) = \frac{e^{-\mu}\mu^k}{k!} \text{ for } k = 0, 1, 2, \ldots$$ ← infinitely many outcomes.

This approximates the binomial dist. when n is arge and p is small.

Why?

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \longrightarrow \frac{e^{-\mu}\mu^k}{k!}$$ as $n \to \infty$ and $p \to 0$ with $np = \mu$?

Facts

① $P_n(0) \approx e^{-\mu}$

② $P_n(k) = P_n(k-1)\frac{\mu}{k}$

so $P_n(1) = e^{-\mu}\frac{\mu}{1}$

$$P_n(2) = P_n(1)\frac{\mu}{2} = e^{-\mu}\frac{\mu}{1}\cdot\frac{\mu}{2} = e^{-\mu}\frac{\mu^2}{2!}$$

etc

Proof of fact ①: $P_n(0) \approx e^{-\mu}$

Remember from Calculus $\log(1+x) \approx x$ for x small

let $P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$ binomial formula

$$P_n(0) = (1-p)^n$$

$$\Rightarrow \log P_n(0) = n\log(1-p) \overset{p \text{ small}}{\approx} n(-p) \overset{np=\mu}{=} -\mu$$

$$\Rightarrow P_n(0) = e^{-\mu}$$

$\square$

$$P_n(k) = P_n(k-1)\frac{\mu}{k}$$

Proof of fact ②:

Remember from sec 2.1 p85, $\frac{P_n(k)}{P_n(k-1)} = \left[\frac{n-k+1}{k}\right]\frac{p}{q}$

$$\Rightarrow P_n(k) = P_n(k-1)\left[\frac{n-(k-1)}{k}\right]\frac{p}{q}$$

$$= P_n(k-1)\left[\frac{\overset{\to \mu}{np} - \overset{\to 0}{(k-1)p}}{k}\right]\overset{\to 1}{\frac{1}{q}} \approx P_n(k-1)\frac{\mu}{k} \quad \square$$

ex Bet 500 times (large), independently, on a bet with 1/1000 (small)

Find the chance of winning at least once.

$P(\text{win} \geq 1 \text{ bet})$

$P(k \geq 1) = 1 - P(0)$

exactly (binomial)

$$1 - P(0) = 1 - \binom{500}{0}\left(\frac{1}{1000}\right)^0\left(\frac{999}{1000}\right)^{500}$$

$$= 1 - \left(\frac{999}{1000}\right)^{500} = .3936$$

approx (Poisson)

$$1 - P(0) = 1 - \frac{e^{-1/2}\left(\frac{1}{2}\right)^0}{0!} = 1 - e^{-1/2} = .3943$$

What about those binomials with p close to 1?

p = chance of success

q = chance of failure

if $p \approx 1$ then $q = 1 - p \approx 0$

$\text{Bin}(n, q) \approx \text{Pois}(\mu = nq)$ for large n, small q.

ex: 97.8% of approx 30 million poor families in the US. have a fridge. If you randomly sample 100 of these families roughly what is the chance 98 or more have a fridge?

p = prob have a fridge $= .978$ ← close to 1

$n = 100$ → large ✓

find $P(98 \text{ or more } \overbrace{\text{have fridge}}^{= \text{success}})$

$q = .022$

$n = 100$

$\mu = nq = 2.2$

$= P(2 \text{ or less don't have fridge})$

$= P(0) + P(1) + P(2)$

$\approx e^{-2.2} + e^{-2.2}(2.2) + \dfrac{e^{-2.2}(2.2)^2}{2!}$

ex

Suppose you and I each have a box of 600 marbles. In my box, 4 of the marbles are black, while 3 of your marbles are black. We each draw 300 marbles **with replacement** from our own boxes. **Approximately**, what is the chance you and I draw the same number of black marbles?

Soln

Use Poisson approx since n large p small.

X = # black I draw

Y = # black you draw

$$\mu_X = 300 \cdot \frac{4}{600} = 2$$

$$\mu_Y = 300 \cdot \frac{3}{600} = 1.5$$

we draw with replacement so k can be 0, 1, 2, …, 300.

Find $P(X=Y) = \sum_{k=0}^{300} P(X=k, Y=k)$ by addition rule,

$= \sum_{k=0}^{300} P(X=k)P(Y=k)$ by independence since draw with replacement

$$\approx \sum_{k=0}^{300} \frac{e^{-2}2^k}{k!} \frac{e^{-1.5}1.5^k}{k!}$$