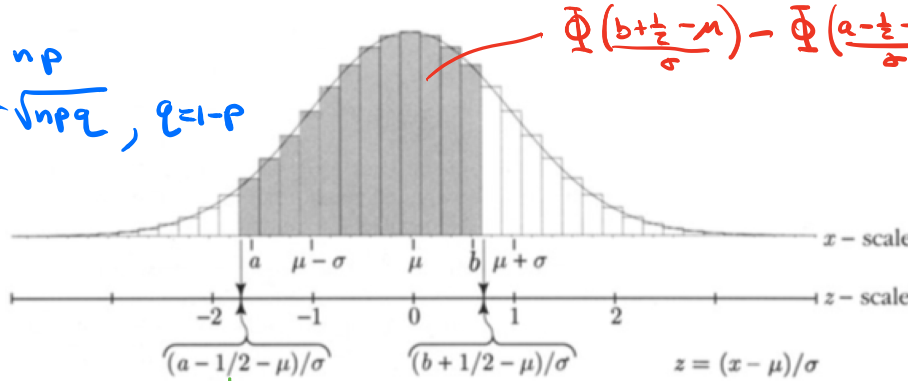
stat 134   lec 6

Last time sec 2.2 Normal Approx to binomial

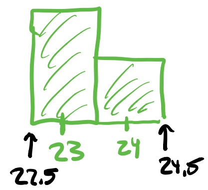
$\mu = np$

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

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



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



Continuity correction

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

ex Find the approximate chance of getting 75 sixes in 600 rolls of a fair die.

soln

$\mu = np = 600\left(\frac{1}{6}\right) = 100$

$\sigma = \sqrt{npq} = \sqrt{600\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} = 9.1$

let $a = b = 75$ then

$$\Phi\left(\frac{75.5-100}{9.1}\right) - \Phi\left(\frac{74.5-100}{9.1}\right) = .00101$$ normal approximation

The exact answer is $\binom{600}{75}\left(\frac{1}{6}\right)^{75}\left(\frac{5}{6}\right)^{525} = .00087$ binomial exact

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%.

P101

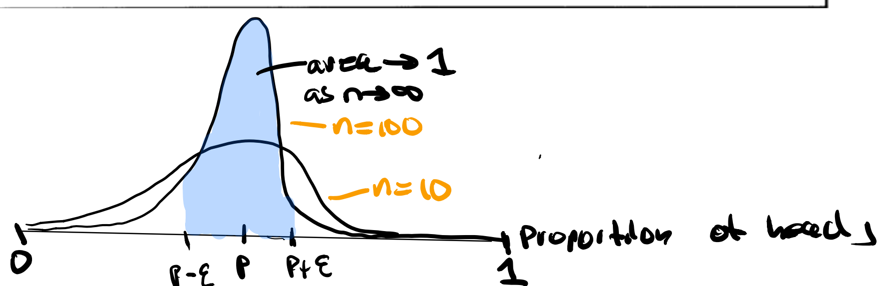
Law of Large Numbers (law of avg)

If n is large, the proportion of successes in n independent trials will, with overwhelming probability, be very close to p, the probability of success on each trial. More formally:

- for independent trials, with probability p of success on each trial, for each $\epsilon > 0$, no matter how small, as $n \to \infty$,

$$P(\text{proportion of successes in } n \text{ trials differs from } p \text{ by less than } \epsilon) \to 1$$

Picture



if n is large
$[p-\varepsilon, p+\varepsilon] \times n$ is a large interval of number of heads so to get exactly p.n heads is small.

	a	Far more spread out for n=100, more possibilities for NOT exactly 50%, imagine the question was n=2 versus n=100 and it's even simpler

	a	look ma i finally got one right <3 UwU

Stat 134

Chapter 2 Friday February 3 2019

1. 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

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

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

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

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

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 curve with the binomial distribution?

Answ let $\mu = np$

$\sigma = \sqrt{npq}$

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

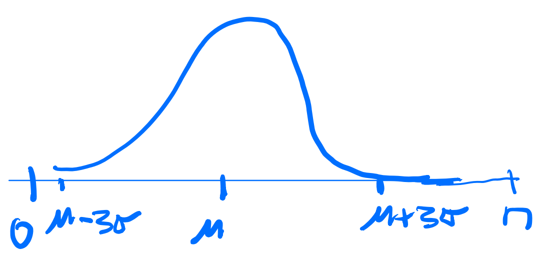
Answ

Lets require $n \geq 20$ 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 $0 \leq \mu - 3\sigma$ and $\mu + 3\sigma \leq n$

Picture



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

$n = 20$ ✓

Need $0 > \mu - 3\sigma$

$$20\left(\frac{1}{10}\right) - 3\sqrt{20 \cdot \frac{1}{10} \cdot \frac{9}{10}} = 2 - 4 = -2 < 0$$ ✗

$$\mu + 3\sigma = 2 + 4 = 6 < 20$$ ✓

NO

skew normal ← bad.

ex

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.)

Partial soln:

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

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

let X = # people who show up.

$X \sim Bin(360, .9)$

Can we approx as normal?

$\mu \pm 3\sigma = 324 \pm 17.1 = [306.9, 341.1]$ ← between 0 and 360 ✓

Finish the problem.

$P(\geq 1 \text{ empty seat}) = P(X \leq 349)$

continuity correction

$\approx \Phi\left(\frac{349 + .5 - \mu}{\sigma}\right) = \Phi\left(\frac{349.5 - 324}{5.7}\right)$

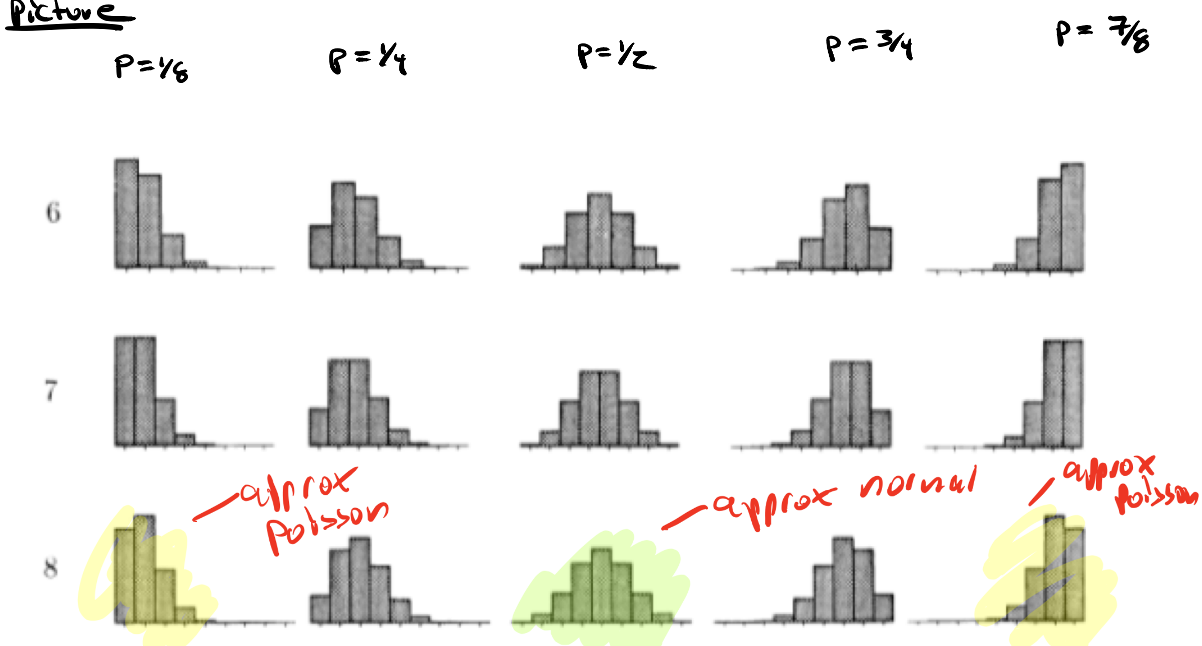
$= \Phi(4.47) = 1$

③ Sec 2.4 (skip 2.3) <u>Poisson approx</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 $Poi(\mu = np)$

<u>Picture</u>



<u>Def</u>

$\underline{\text{Def}^n}$ Poisson(μ) (written Pois(μ))

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

p119

Poisson Approximation to the Binomial Distribution

If n is large and p is small, the distribution of the number of successes in n independent trials is largely determined by the value of the mean $\mu = np$, according to the *Poisson approximation*

$$P(k \text{ successes}) \approx e^{-\mu}\frac{\mu^k}{k!}$$

see end of notes for proof.

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

Chance of winning.

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

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

<u>exactly</u> (binomial)

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

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

<u>approx</u> (Poisson)

$$1 - P(0) = 1 - \frac{e^{-1/2}\left(\frac{1}{2}\right)^0}{0!} = 1 - e^{-1/2} = .3935$$

What about those binomials with p close to 1?

p = chance of success

q = chance of failure

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

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

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

Soln

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

$n = 100$ — large

find $P($98 or more have fridge$)$ success $p = .978$

$= P($2 or less dont have a fridge$)$ failure $q = .022$

$q = .022$

$n = 100$

$\mu = nq = 2.2$

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

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

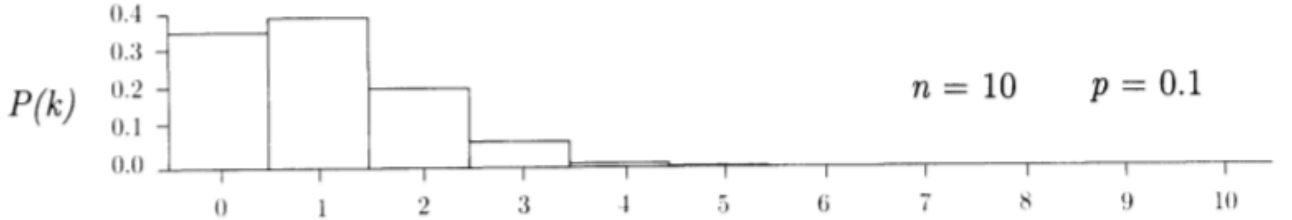
$$= e^{-2.2}\left(1 + 2.2 + \frac{2.2^2}{2}\right) = .6227$$

Here is a picture of how Pois(1) is a limit of binomials as $p \to 0$ and $n \to \infty$ and $np \to 1$

Example 1. **The binomial (10, 1/10) distribution.**

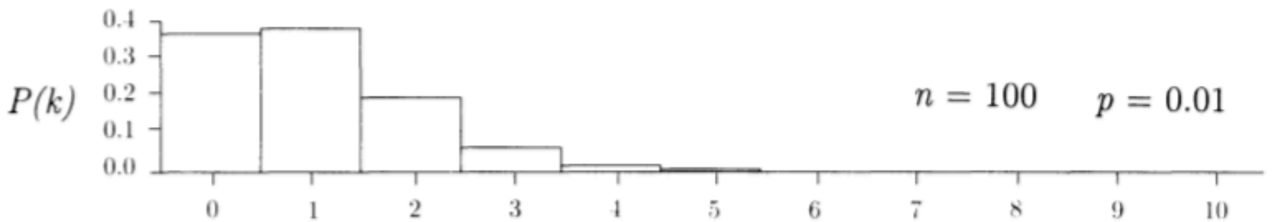
P117

This is the distribution of the number of black balls obtained in 10 random draws with replacement from a box containing 1 black ball and 9 white ones.



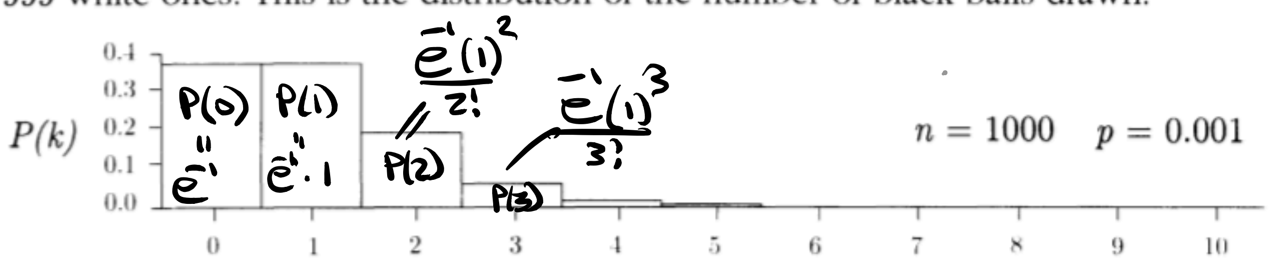
Example 2. **The binomial (100, 1/100) distribution.**

This is the distribution of the number of black balls obtained in 100 random draws with replacement from a box containing 1 black ball and 99 white ones.



Example 3. **The binomial (1000, 1/1000) distribution.**

Now take 1000 random draws with replacement from a box with 1 black ball and 999 white ones. This is the distribution of the number of black balls drawn:



As these examples show, binomial distributions with parameters n and $1/n$ are always concentrated on a small number of values near the mean value $\mu = 1$, with a 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 $0, 1, 2, \ldots, 300$.

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

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

Appendix

Why

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k} \longrightarrow \frac{e^{-\mu} \mu^k}{k!} \quad \text{as } n \to \infty \text{ and } p \to 0$$

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:

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

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

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

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

□

Proof of fact ②:

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

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

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

□