Stat 134 lec 6

Warmup 11:00–11:10

You roll a fair die 600 times.
Using the normal distribution find the chance you get between 74.5 and 75.5 sixes.
Leave your answer in terms of Φ.

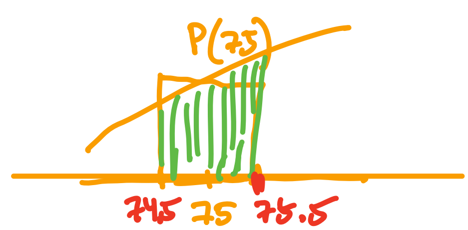
$\mu = np = 600(1/6) = 100$

$\sigma = \sqrt{npq} = \sqrt{600(1/6)(5/6)} = 9.1$

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

Comment exact value $\binom{600}{75}\left(\frac{1}{6}\right)^{75}\left(\frac{5}{6}\right)^{525} = .00087$

Motivation: It is difficult to calculate exact probabilities with the binomial formula. It is easier to calculate the area under the normal curve.

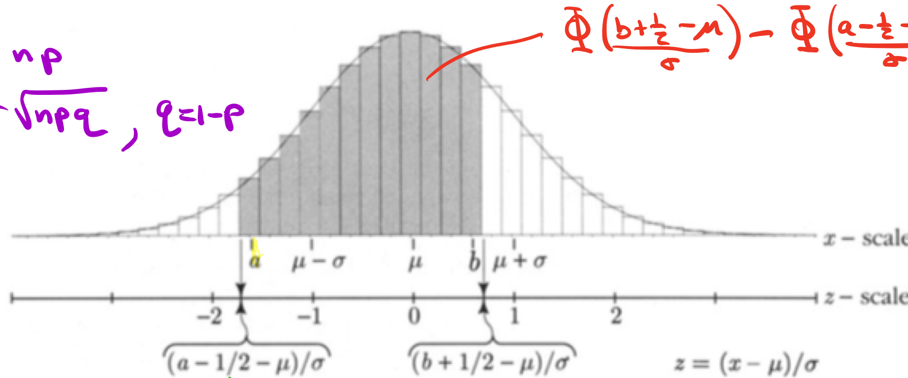


Last time Sec 2.2 Normal Approx to binomial

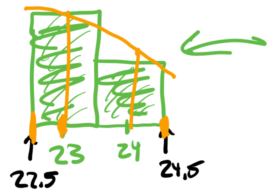
$\mu = np$

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

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



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



Continuity correction (CC)

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)

≥ 135

Hint The mean of $Bin(n,p)$ is $\mu = np$

The standard deviation of $Bin(n,p)$ is $\sigma = \sqrt{npq}$

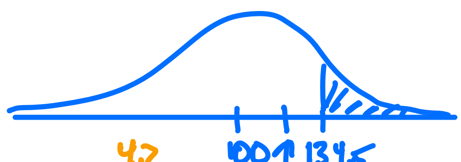
$n = 300$

$p = 1/3$

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

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

X = # patients helped



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

4.2

$\approx 1-1 = \boxed{0}$

$\mu + 3\sigma = 124.6$

so expect area to right of 134.5 to be ≈ 0

Today

1. finish sec 2.2
2. sec 2.4 Poisson approximation (skip sec 2.3)

① **Sec 2.2** Normal approximation to the binomial distribution

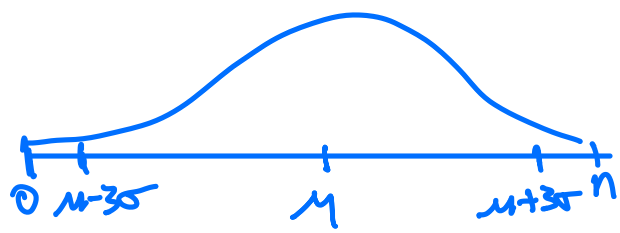
For what n, p is it ok to approximate $Bin(n,p)$ by a normal distribution $N(\mu, \sigma^2)$.

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

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

all of our data between $\mu \pm 3\sigma$ so we require

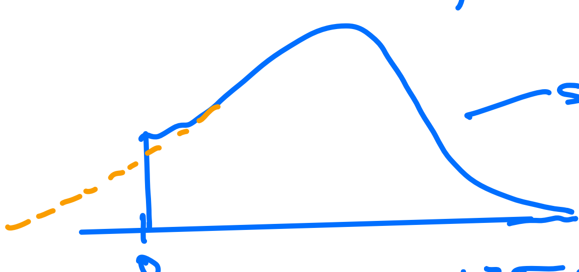
$\mu - 3\sigma > 0$ and $\mu + 3\sigma < n$



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

$n = 20$ ✓

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



$\mu + 3\sigma = 2 + 4 = 6 < 20$ ✓

. (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.)

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

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

$\mu + 3\sigma < n$ and $\mu - 3\sigma > 0$ and $n = 360 \geq 20$ ✓

so we can use the normal approx.

X = # people who show up.

cont. correction

$P(X \leq 349) \approx P(X \leq 349.5)$

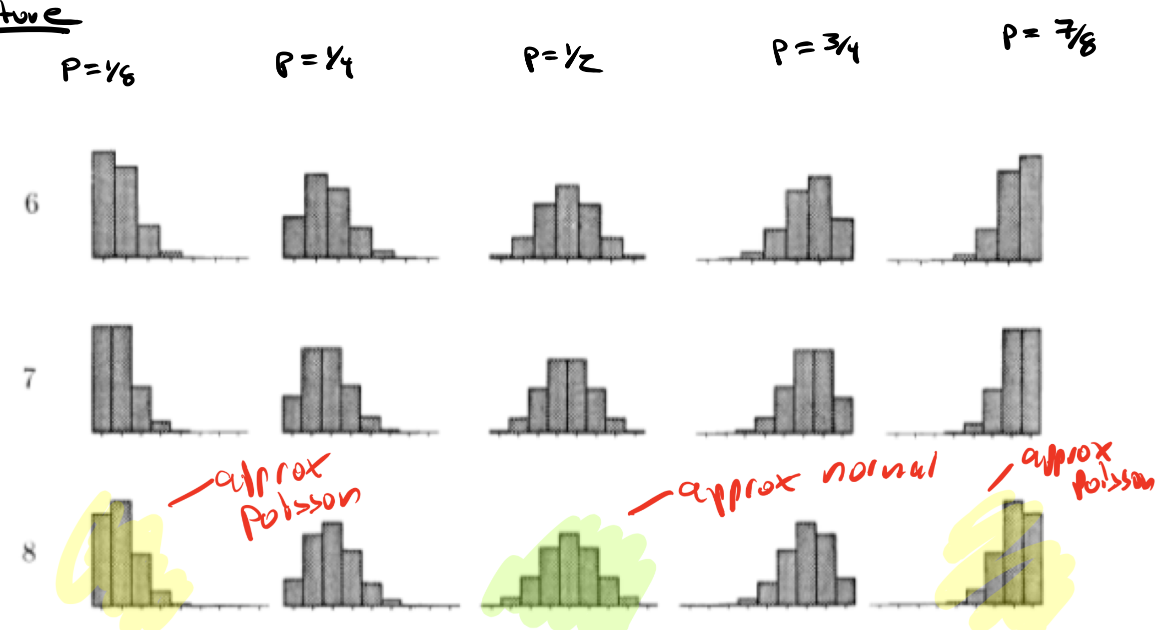
$= \Phi\left(\frac{349.5 - \mu}{\sigma}\right) = \Phi(4.47) = 1$

(2) Sec 2.4 (skip 2.3) Poisson approx to Binomial

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$(\mu = np)$

Picture



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.

You can just define the Poisson(μ) distribution this way or think of it as a limit of the Binomial formula when n is large and p is small and $np \to \mu$.

Proven in appendix at end of lecture notes,

Thm

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

with $np \to \mu$

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

Approximate the chance of winning at least once.

Don't use CC since Poisson is discrete.

Def^n Poisson (μ)

$$P(k) = \frac{e^{-\mu}\mu^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

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

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

$$= 1 - \frac{e^{-500(1/1000)} \cdot \left(\frac{500}{1000}\right)^0}{0!} = 1 - e^{-1/2} = .3943$$

exactly (binomial)

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

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

Calculating $P(k)$ using the Poisson formula versus the Binomial formula is a little easier. The main point I want to make is that Pois (μ) is related to Bin (n, p).

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.

p large

large

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?

<u>Def</u>

Defn Poisson(μ)

$$P(k) = \frac{e^{-\mu}\mu^k}{k!} \quad \text{for } k=0,1,2,\dots$$

p = Prob have fridge $= .978$

$n = 100$

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

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

$$\approx P(0) + P(1) + P(2)$$

use $\text{Pois}(\mu = \overset{100}{n}\overset{.022}{q}) = \text{Pois}(2.2)$

$$\approx \boxed{e^{-2.2} + \frac{e^{-2.2}(2.2)^1}{1!} + \frac{e^{-2.2}(2.2)^2}{2!}}$$

Appendix

<u>Thm</u> Let $P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$ (binomial formula)

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

Pf/ The claim follows if we show these 2 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

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

p small, np = μ

$\Rightarrow \log P_n(0) = n\log(1-p) \approx n(-p) = -\mu$

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

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

Proof of fact ②:

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

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

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