Stat 134 lec 6

Last time: sec 2.2 Normal approx of binomial

- for binomial(n,p), $\mu = np$, $\sigma = \sqrt{npq}$, $q = 1-p$
- when $0 < \mu \pm 3\sigma < n$ and n large we can approx binomial(n,p) by normal.

Today: 1) Finish sec 2.2 law of large numbers

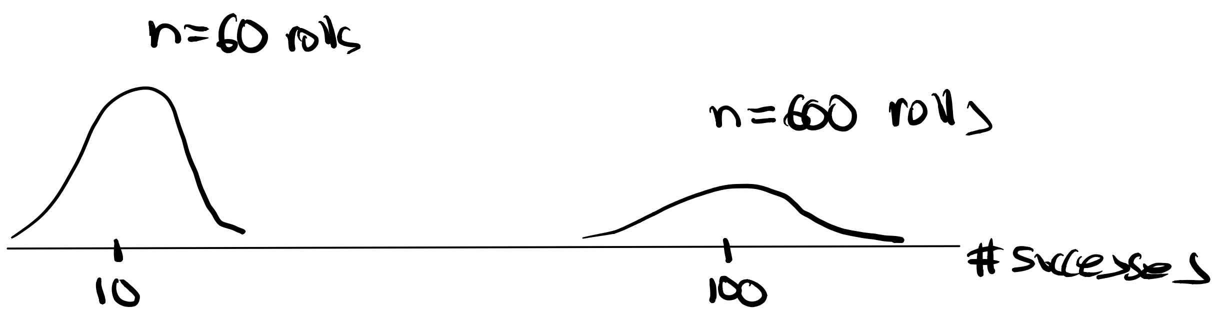
2) Sec 2.4 (skip 2.3) Poisson approx.

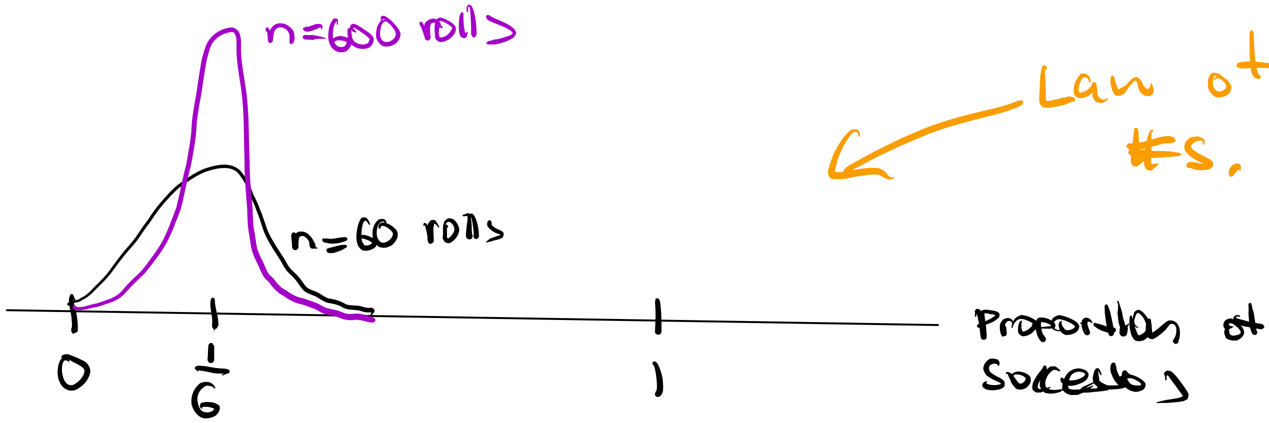
Law of large numbers (sec 2.4)

You roll a die with probability $p = 1/6$ of getting a six.

Success = getting a six.

Pictures: Which is law of large #'s?





Law of large #s.

Law of large #'s says

Proportion of successes $\longrightarrow p$ as $n \to \infty$

You can see this in the bottom picture. As n gets large the distribution of the proportion of successes gets more concentrated around the $1/6$ the true chance of a success.

To understand why, lets call the proportion of successes our estimate of p. Then

$$\text{estimate} = \frac{\#\text{successes}}{n}$$

$$\text{avg}(\text{estimate}) = \frac{\text{avg}(\#\text{successes})}{n} = \frac{np}{n} = p$$

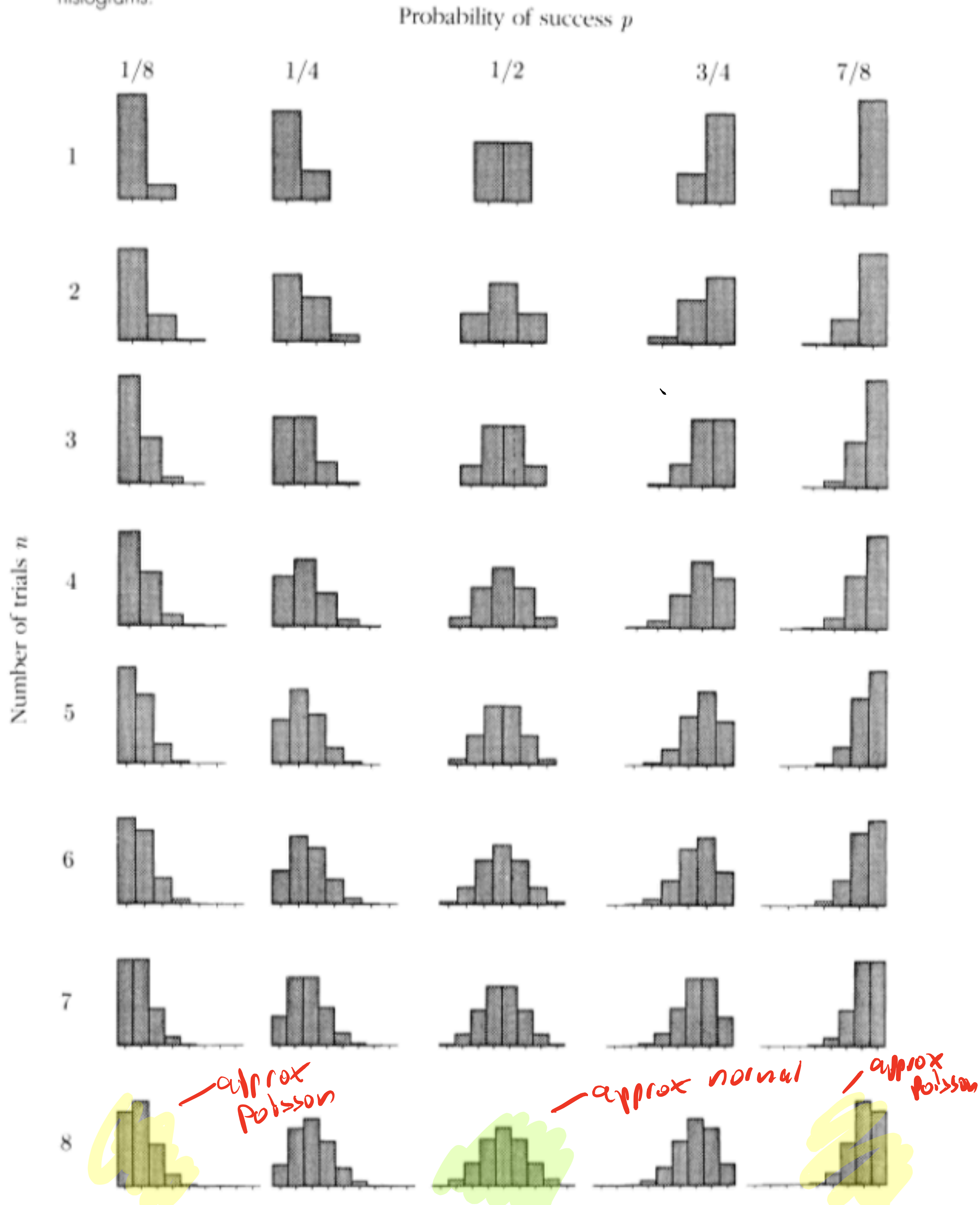
$$\sigma(\text{estimate}) = \frac{\sigma(\#\text{successes})}{n} = \frac{\sqrt{npq}}{n} = \frac{\sqrt{pq}}{\sqrt{n}}$$

avg(estimate) $= p$ says that the center of our distribution of our estimate is p independent of what n is.

$\sigma(\text{estimate}) = \frac{\sqrt{pq}}{\sqrt{n}}$ is called the <u>square root law</u> and says that the distribution of our estimate gets steeper as n gets bigger since $\sqrt{n}$ is in the denominator.

P87

FIGURE 3. Histograms of some binomial distributions. The histogram in row n, column p shows the binomial (n, p) distribution for the number of successes in n independent trials, each with success probability p. In row n, the range of values shown is 0 to n. The horizontal scale changes from one row to the next, but equal probabilities are represented by equal areas, even in different histograms.



Sec 2.4 (skip 2.5) Poisson approx.

The normal approx has almost 100% of data $\pm 3\sigma$ from mean μ.

For this reason we approximated binomial w/ normal only when $\mu \pm 3\sigma$ is between 0 and n.

For the cases when p is small or p is close to 1 and n is large, we approximate

(See picture of binomial histograms above)

Binomial (n, p) by Poisson ($\mu = np$)

Defn Poisson (μ)

Outcomes 0, 1, 2, ... ← infinitely many possible outcomes

$$\boxed{P_\mu(k) = \frac{e^{-\mu}\mu^k}{k!}}$$

Ex Bet 500 times, independently, on a bet with $\frac{1}{500}$ chance of winning.

large

small

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

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

exactly (binomial)

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

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

approximately (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 $q = 1 - p \approx 0$

Binomial$(n, q) \approx$ Poisson$(\mu = nq)$ for large n, small q

ex

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

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

$n = 100$ — large.

success

Find P(98 or more have fridge)

failure

= P(2 or less $\underbrace{\text{don't have a fridge}}_{q = .022}$)

$q = .022$
$n = 100$
$\mu = 2.2$

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

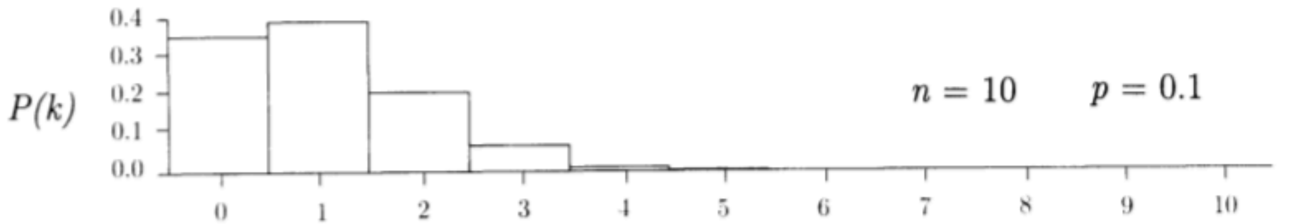
$\approx e^{-2.2} + e^{-2.2}(2.2) + \dfrac{e^{-2.2}(2.2)^2}{2!}$

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

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

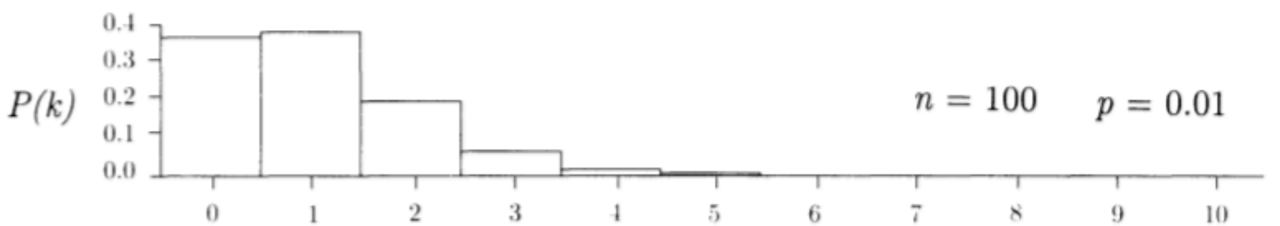
Example 1. **The binomial (10, 1/10) distribution.**

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.



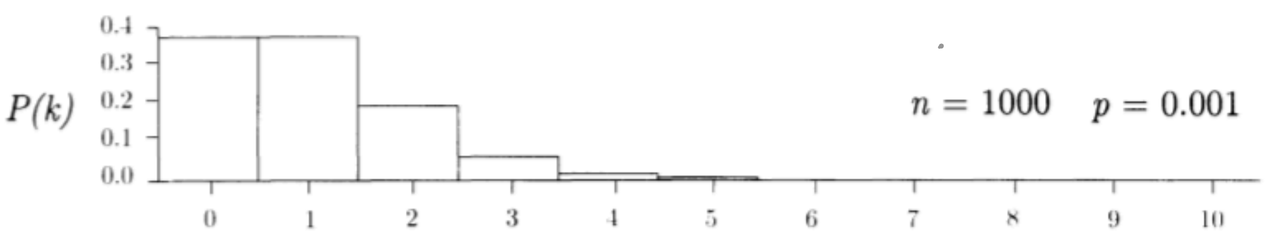
Example 2. **The binomial (100, 1/100) distribution.**

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

Why

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

<u>Facts</u>

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

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

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

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

$\square$

Proof of fact ② :

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

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

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

$np \to \mu$, $(k-1)p \to 0$, $q \to 1$

Stat 134

Chapter 2 Wednesday September 5 2018

1. Recall that $np + p$ is used to test for the mode of a binomial. The mode of Poisson(5) is:

   4 and 5

   a 5

   b 4,5

   c 0,1

   d none of the above

For binomial,

$m = \lfloor np+p \rfloor$ ← floor of $np+p$

$$\text{mode} = \begin{cases} m & \text{if } k \neq np+p \\ m-1, m & \text{if } k = np+p \end{cases}$$

as $p \to 0$ and $np \to 5$

only $\mu = 5$ is neccessary to find the mode.

For Poisson(μ), let

$m = \lfloor \mu \rfloor$

$$\text{mode} = \begin{cases} m & \text{if } k \neq \mu \\ m-1, m & \text{if } k = \mu \end{cases}$$

we have $\lfloor 5 \rfloor = 5$ so $\boxed{\text{mode} = 4, 5}$