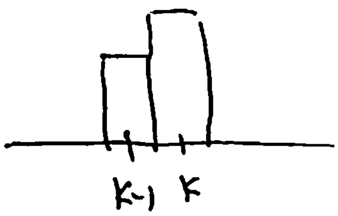
1/26/18

Stat 134 lec 5

I-clicker-tests (see next page),
Mode versus average

Last time using consecutive odds ratio we showed $P(k-1) \le P(k)$ iff $k \le (n+1)p$.

k: integer
$(n+1)p$: not necessarily integer



eg $n=4$
$p = 1/2$
$(n+1)p = 2.5$

$$\text{mode} = \text{int}((n+1)p)$$

$\leftarrow 2$

Average $= \boxed{\mu = np}$

If μ is an integer then it is the mode since if np is an integer then $np+p$ is not an integer $(0<p<1)$ so mode $= \text{int}(np+p) = np$.
μ means "nothing" if not an integer.

The normal dist

The normal curve with center μ and spread σ is given by $y = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty < x < \infty$

2. Is the binomial formula applicable to find the chance that the sum of draws is **3** while drawing **5 times with replacement** from a box with 9 tickets marked **0** and one ticket marked **1**.

a yes

b no

sum of draws is the # successes in 5 Bernoulli trials

$\frac{1}{10}$ $\frac{9}{10}$

1 0

↓ draw 5 w/ replacement

$n = 5$ indep trials

$p = \frac{1}{10}$

$k = 3$

3. Is the binomial formula applicable to find the chance that you have to toss a coin **8** times until you get a total of **2** heads.

a yes

b no

Similar to binomial — called negative binomial

get 1 p

$p \;|\; \underline{q}\;\underline{q}\;\underline{q}\;\underline{q}\;\underline{q}\;\underline{q}\;\underline{p}$ $\binom{7}{1}p^2q^6$ ← this isn't the binomial formula.

$\binom{7}{1} = 7$ ways to get 1p and 6 qs here.

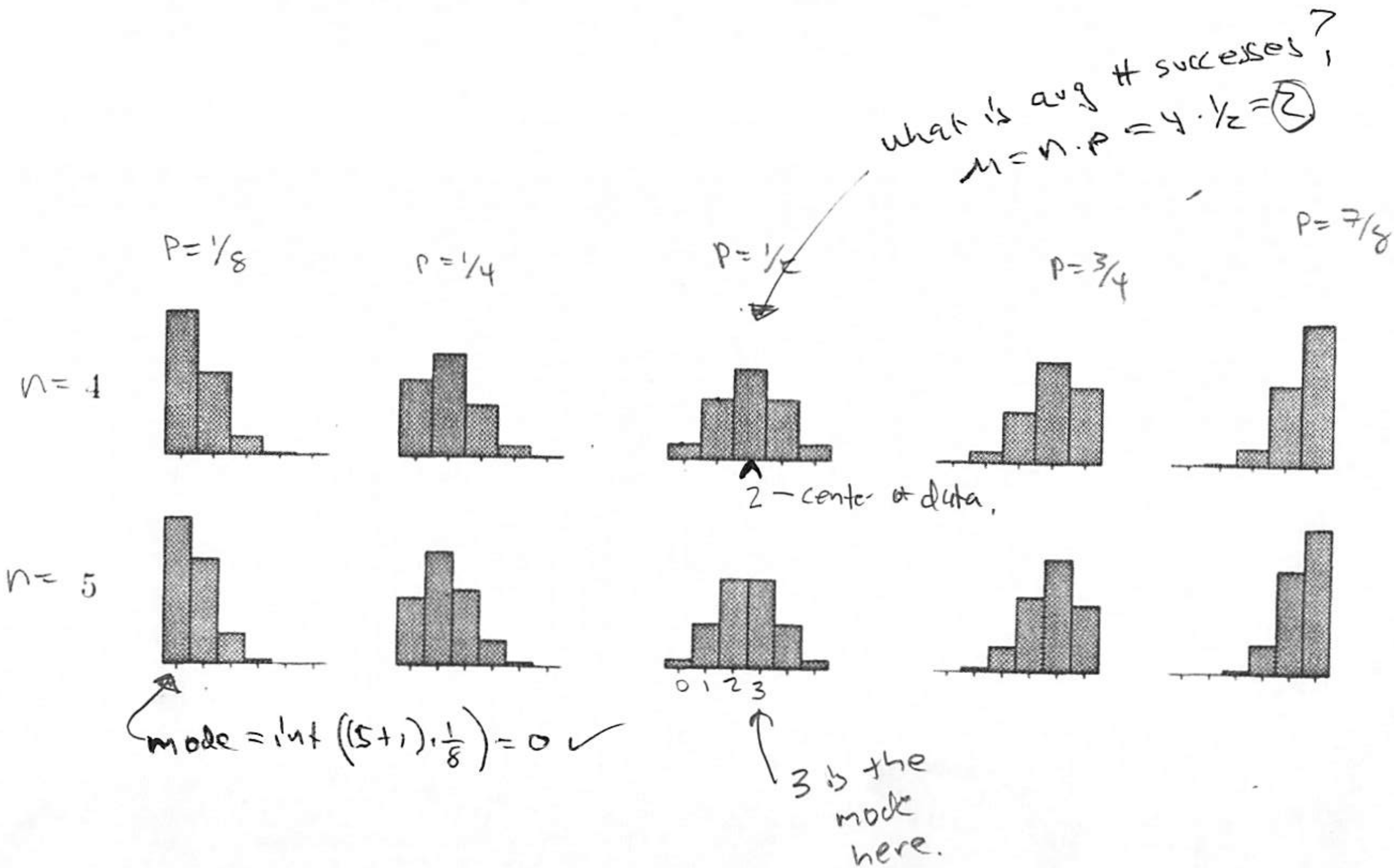
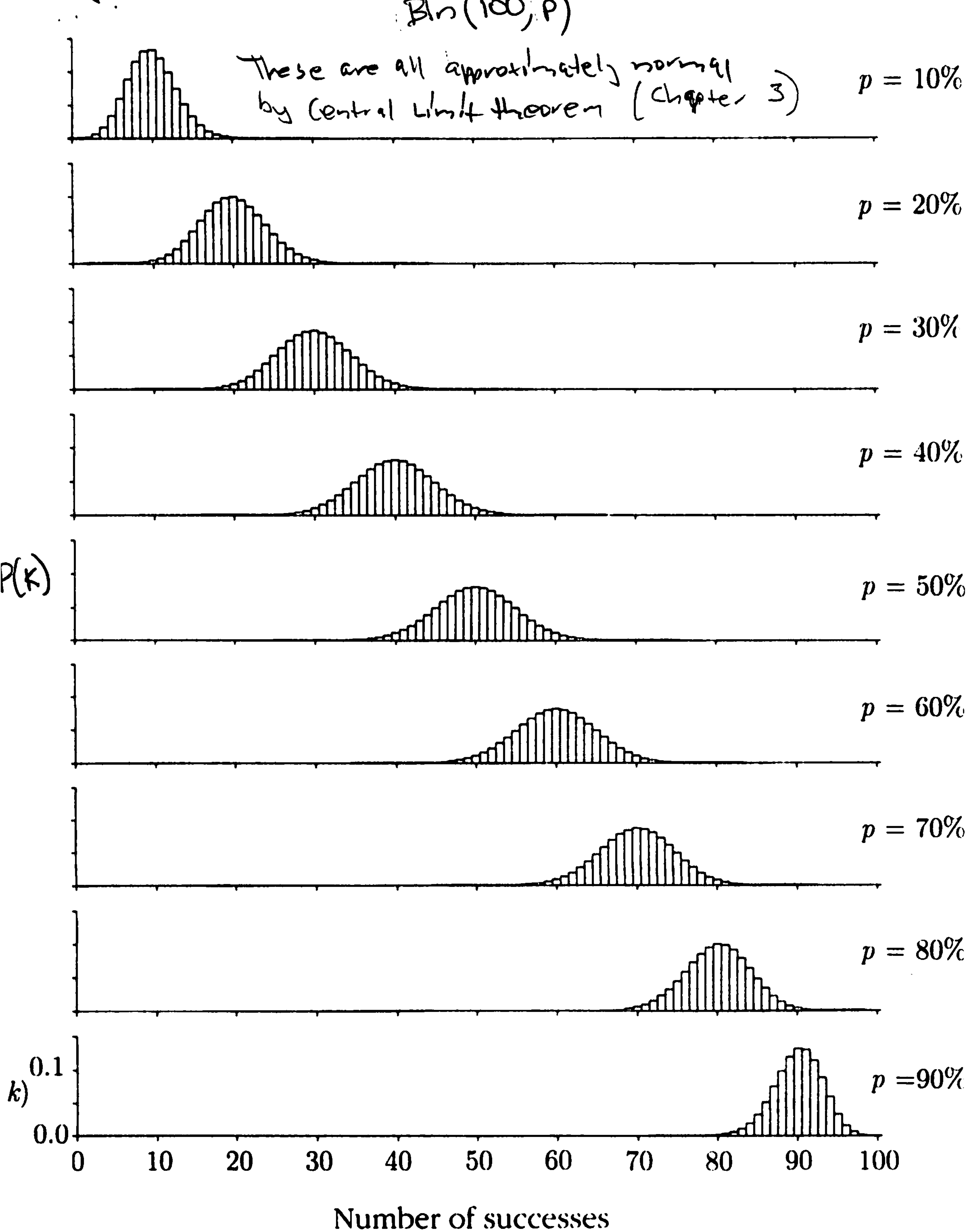
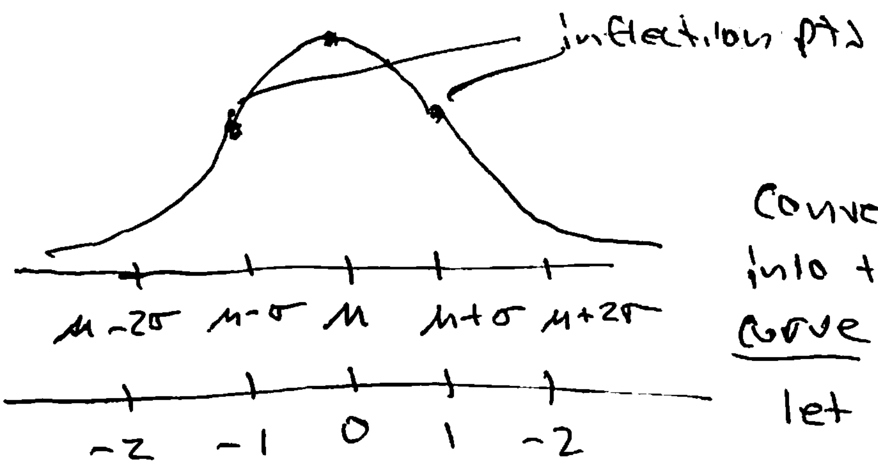

what is avg # successes?
$\mu = n \cdot p = 4 \cdot \frac{1}{2} = 2$
$p = 1/8$
$p = 1/4$
$p = 1/2$
$p = 3/4$
$p = 7/8$
$n = 4$
$n = 5$
2 – center of data.
0 1 2 3
mode = int$\left((5+1)\cdot\frac{1}{8}\right) = 0$ ✓
3 is the mode here.

Fig 5 p69


Bin(100, p)
These are all approximately normal
by Central Limit theorem (Chapter 3)
p = 10%
p = 20%
p = 30%
p = 40%
p = 50%
p = 60%
p = 70%
p = 80%
p =90%
P(k)
k)
0.1
0.0
0
10
20
30
40
50
60
70
80
90
100
Number of successes



Convenient to transform into the <u>standard normal curve</u>

let $z = \dfrac{x - \mu}{\sigma}$

$\mu \to 0$

$\mu + \sigma \to 1$

$\mu + 2\sigma \to 2$

$\mu - \sigma \to -1$

The std normal curve has mean 0 and std dev 1 and has formula $\phi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ $\quad -\infty < z < \infty$

(will show using CVT chap 4).

We find the area under the curve to the left of z using <u>cummulative distribution function</u>

CDF $\quad \Phi(z) = \displaystyle\int_{-\infty}^{z} \phi(s)\,ds$.

Find area between -1 and 1: $\Phi(1) - \Phi(-1)$

$= .8413 - (1 - .8413)$

$= 2(.8413) - 1$

$= .68$

Find area between -2 and 2: $2\Phi(2) - 1 = .95$

-3 and 3: $.997$

<u>Empirical rule</u>

Normal approx to binomial (n, p) dist when n is large,

Use normal curve with $\mu = np$ and $\sigma = \sqrt{npq}$.

eg

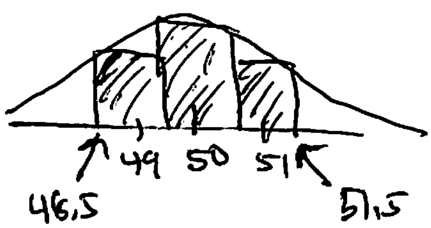
If $n = 100$, $p = .5$, then use $\mu = 50$, $\sigma = 5$

P(between 49 and 51 heads inclusive)

$$= \sum_{k=49}^{51} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k}$$

$$= \left(\binom{100}{49} + \binom{100}{50} + \binom{100}{51} \right) \cdot \left(\frac{1}{2}\right)^{100} = .2359$$

in R.



$$\Phi\left(\frac{51.5 - 50}{5}\right) \quad \Phi\left(\frac{48.5 - 50}{5}\right)$$

$= .3$

$$P(a \text{ to } b \text{ successes}) \approx \Phi\left(\frac{b + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{a - .5 - \mu}{\sigma}\right)$$

called continuity correction

very important when σ is small

approx: $2\Phi(.3) - 1 = 2(.6179) - 1 = .2358$ ← w/ continuity correction

w/o cont correction

$2\Phi(.2) - 1 = .1585$ ← bad.