Stat 134 lec 5

last time

sec 2.1

binomial distribution — n independent Bernoulli (p) trials

binomial formula $P(k) = \binom{n}{k} p^k q^{n-k}$ where $q = 1-p$

The consecutive odds ratio is

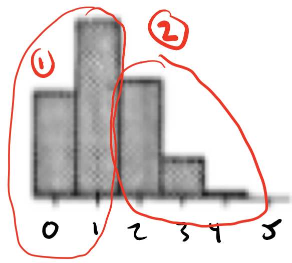
$$\frac{P(k)}{P(k-1)} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \boxed{\frac{n-k+1}{k} \cdot \frac{p}{1-p}}$$

helps show* — see end of lecture notes.

ex $n=5$, $p=\frac{1}{4}$

$np+p = 1\frac{1}{2}$

$\lfloor np+p \rfloor = 1$



① $k < np+p$ iff $P(k-1) < P(k)$

② $k > np+p$ iff $P(k-1) > P(k)$

③ $k = np+p$ iff $P(k-1) = P(k)$

This helps us find the mode (most likely outcome)

$$\text{mode} = \begin{cases} m & \text{if } np+p \notin \mathbb{Z} \quad \text{where } m = \lfloor np+p \rfloor \\ m, m-1 & \text{if } np+p \in \mathbb{Z} \end{cases}$$

Today ① review student responses to concept test

② Finish sec 2.1 Binomial distribution

③ Start sec 2.2 Normal approximation to the binomial.

1

1. Ten cards are dealt off the top of a well shuffled deck. The binominal formula doesn't apply to find the chance of getting exactly three diamonds because:

   a The probability of a trial being successful changes

   b The trials aren't independent

   c There isn't a fixed number of trials

   d more than one of the above

d	b	what's the difference between unconditional and conditional probability

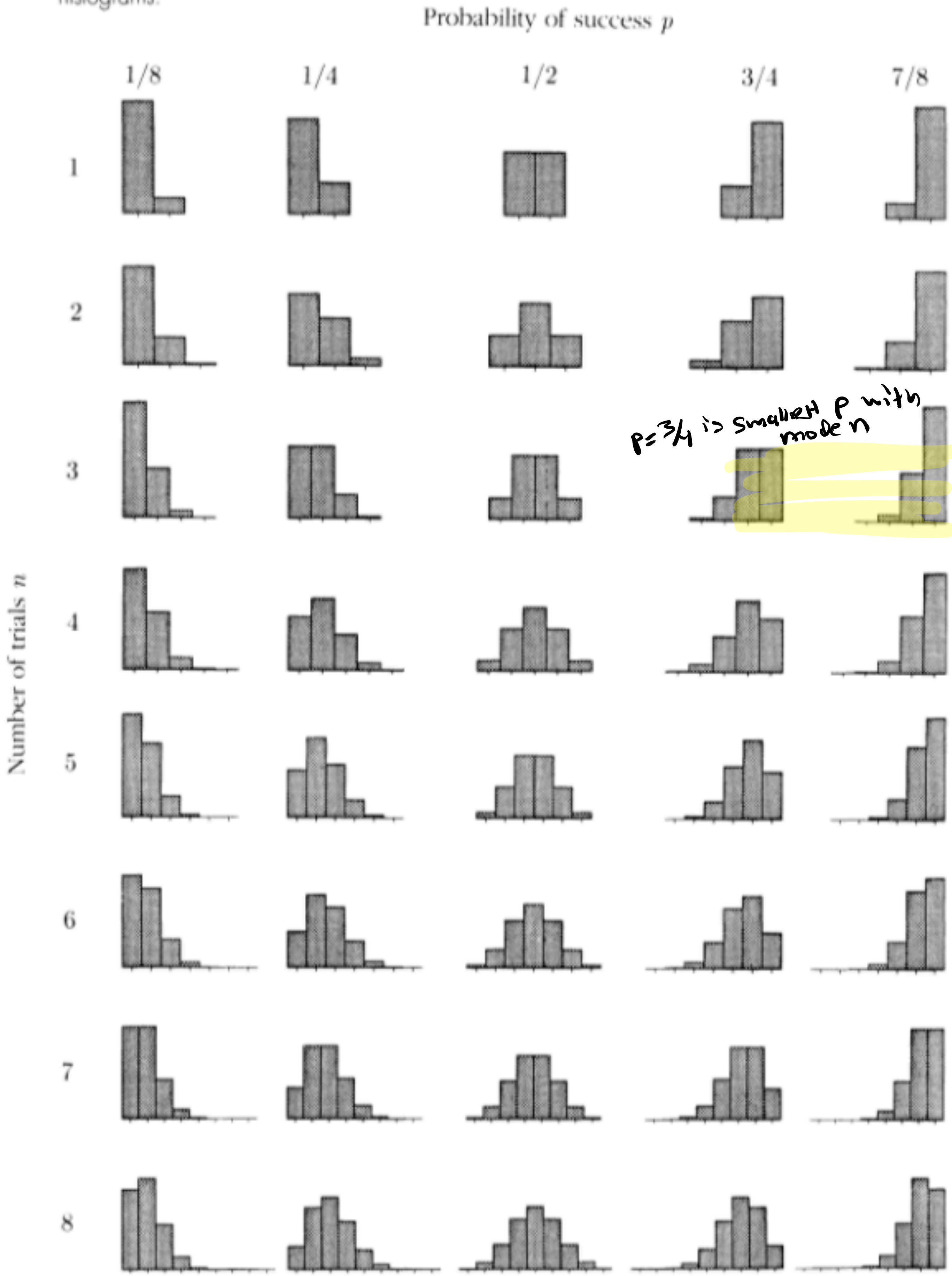
d	d	There are 10 trials, but the probability the card is a diamond depends on the previous cards. So, both A & B are true.	

d	d	A and B are literally talking about the same thing. The chance of a single trail (if not the first time) depends on the results from previous ones.	

d	b	The unconditional probability of getting an ~~ace~~ diamond is always 1/13/ If we were drawing from two decks of cards having different number of ~~aces~~ diamonds than a would be false.	

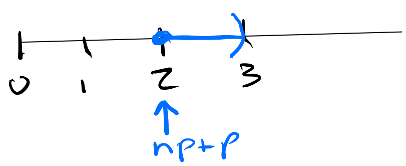
(2) Sec 2.1
p87

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.

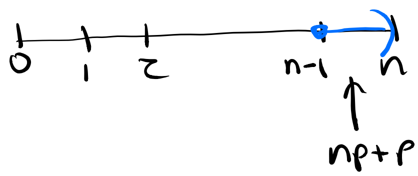


What is the mode in each of the following scenarios?

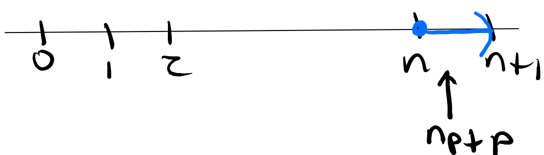
Picture



$m = \lfloor np+p \rfloor = ?$ — 2



$m = \lfloor np+p \rfloor = ?$ — n-1



$m = \lfloor np+p \rfloor = ?$ — n

ex For each n, what is the smallest p such that n is the most likely number of successes (i.e part of the mode)?

n is mode if

$$n \le np+p < n+1$$

When $np+p = n$ we have both $n-1$ and n the mode.

For p any smaller, $np+p<n$, n won't be the mode

Hence solve $n = np+p$ for p.

$$\Rightarrow n = p(n+1)$$

$$\Rightarrow \boxed{p = \frac{n}{n+1}}$$

The mode is a measure of the center of your data.

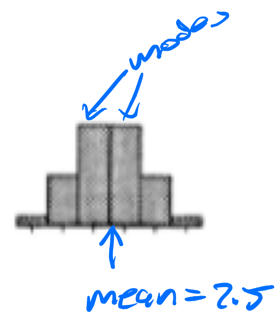
The true center of your data is the expectation or mean

shown in chap 3

Fact The expected (mean) number of successes is $\boxed{\mu = np}$

This isn't usually an integer

ex $n=5$, $p=1/2$ $\quad \mu = np = 5/2$



If the mean is an integer is it the mode?

Yes, if $np \in \mathbb{Z}$ then $np+p \notin \mathbb{Z}$ so have one mode and $m = \lfloor np+p \rfloor = np$ is the mode.

shown in Chap 3

Fact the average spread around the mean (standard deviation) is $\sigma = \sqrt{npq}$ where $q = 1-p$

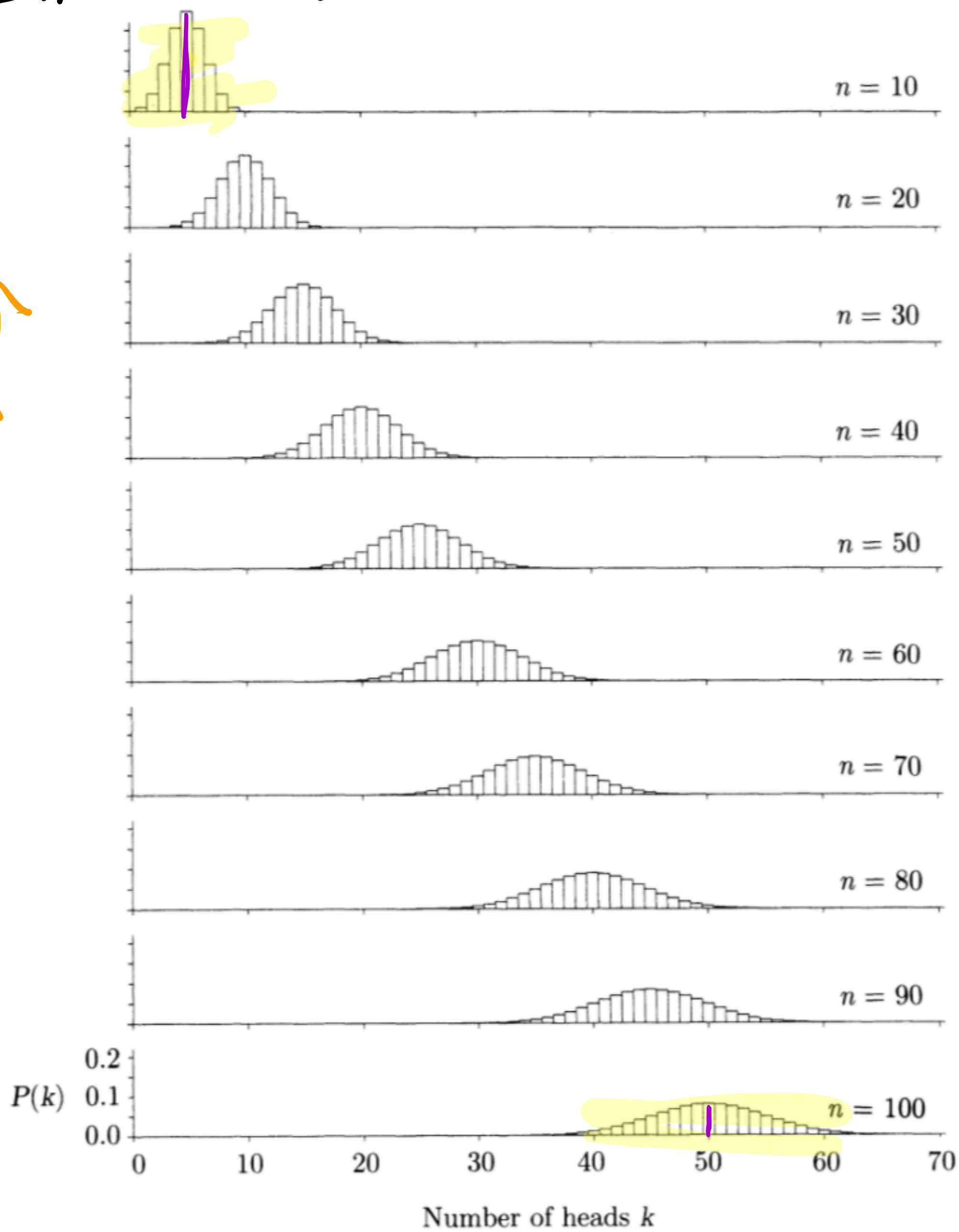
Notice that the spread around the mean gets larger as n gets bigger.

P88

Binomial $(n, \frac{1}{2})$

$$\sigma = \sqrt{npq} = \sqrt{n \frac{1}{2} \cdot \frac{1}{2}} = \frac{\sqrt{n}}{2}$$

Notice spread ↑ as n ↑



Stat 134

Chapter 2 Friday February 1 2019

1. 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

As n increases the spread increases and the probabilities go down. Think of finding a needle in a growing haystack.

The binomial formula gives:

$$P(5) = \binom{10}{5}\left(\frac{1}{2}\right)^{10} = .246$$

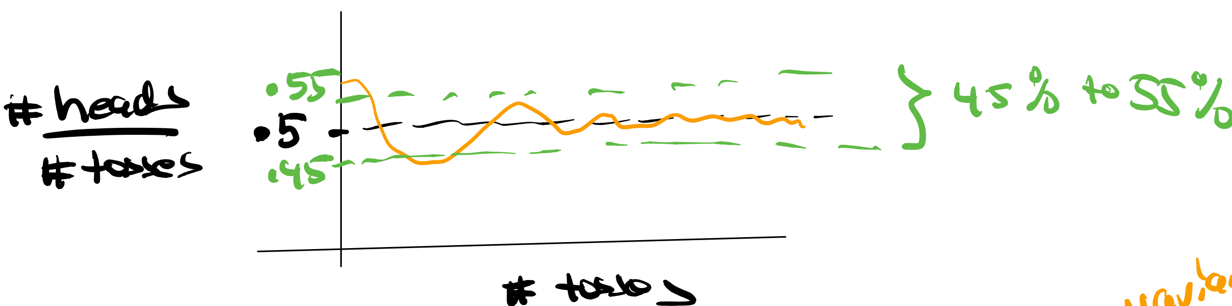
$$P(50) = \binom{100}{50}\left(\frac{1}{2}\right)^{100} = .08$$

2. A fair coin is tossed, and you win a dollar if you get between 45% an 55% heads. Which is better?

   a 10 tosses

   b 100 tosses

you saw in sec 1.2 that the relative frequency of heads decreases with more coin tosses



explained chapter 3

reason:

$$\underset{\text{variance}}{\mathrm{var}}\left(\frac{\#\text{heads}}{n}\right) = \frac{\mathrm{var}(\#\text{heads})}{n^2} = \frac{\overset{\text{variance of } Bin(n,p)}{npq}}{n^2}$$

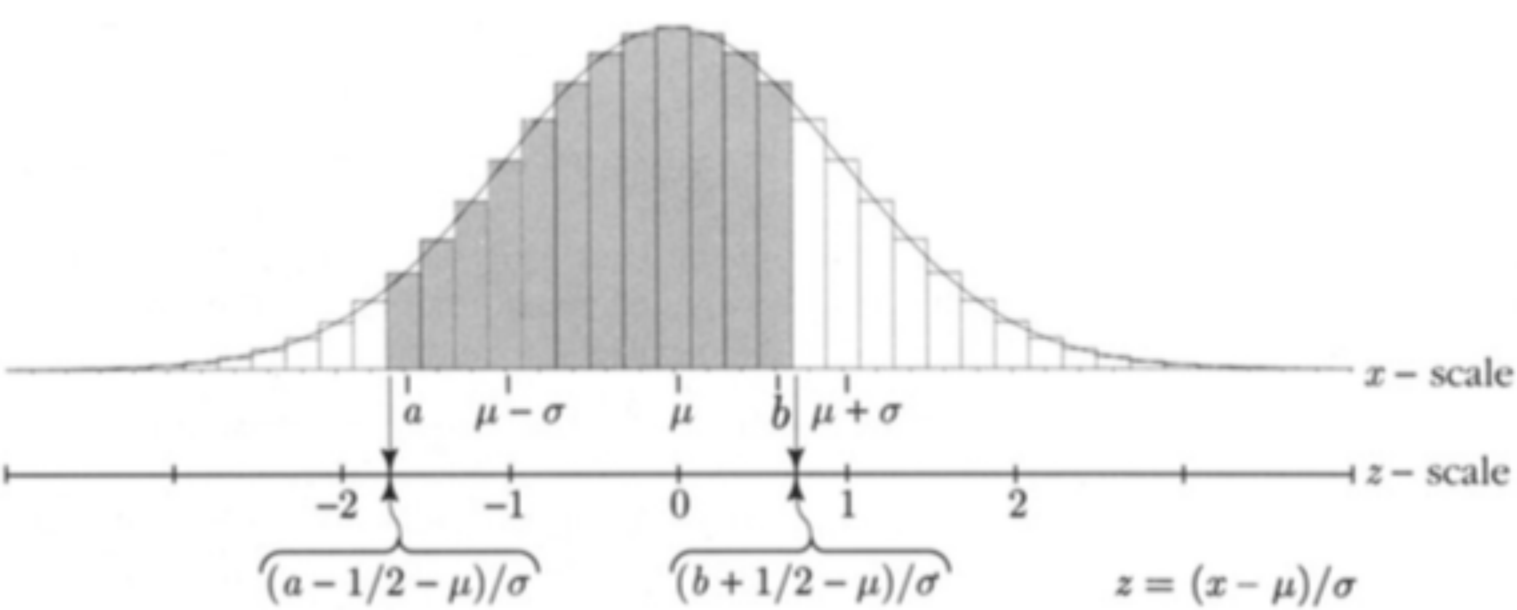
$$= \frac{pq}{n} \to 0 \text{ as } n \to \infty.$$

(3) Sec 2.2 Normal approx to binomial

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

p98

FIGURE 4. A binomial histogram, with the normal curve superimposed. Both the x scale (number of successes) and the z scale (standard units) are shown.



The normal curve is

$$y = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Notice:

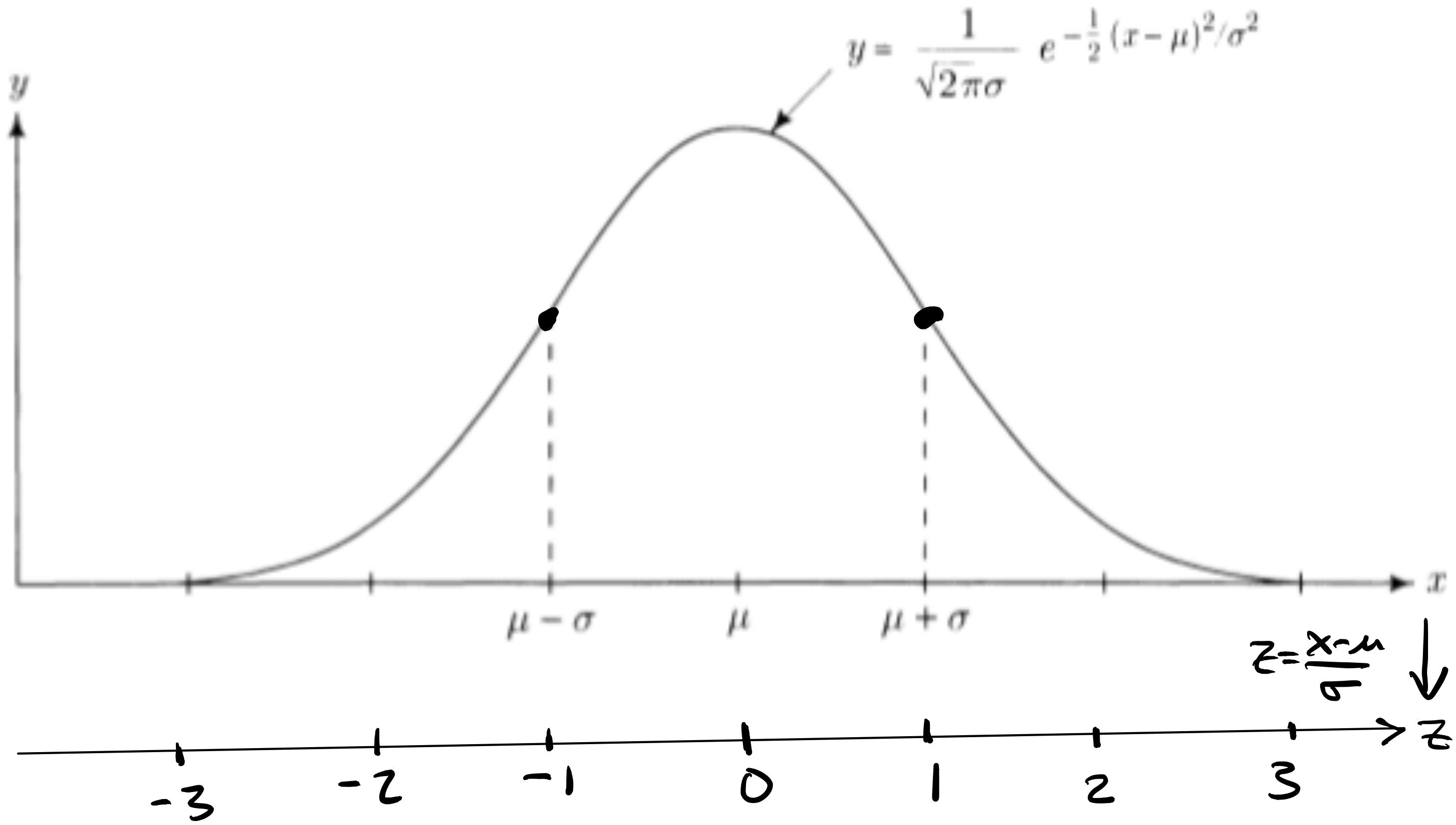
① two param μ = mean
σ = std dev

② inflection pts $\mu \pm \sigma$

③ almost all data between $\mu \pm 3\sigma$

P93

FIGURE 1. The normal curve.



To find area under curve convenient to make a change of coords.

$$z = \frac{x-\mu}{\sigma}$$

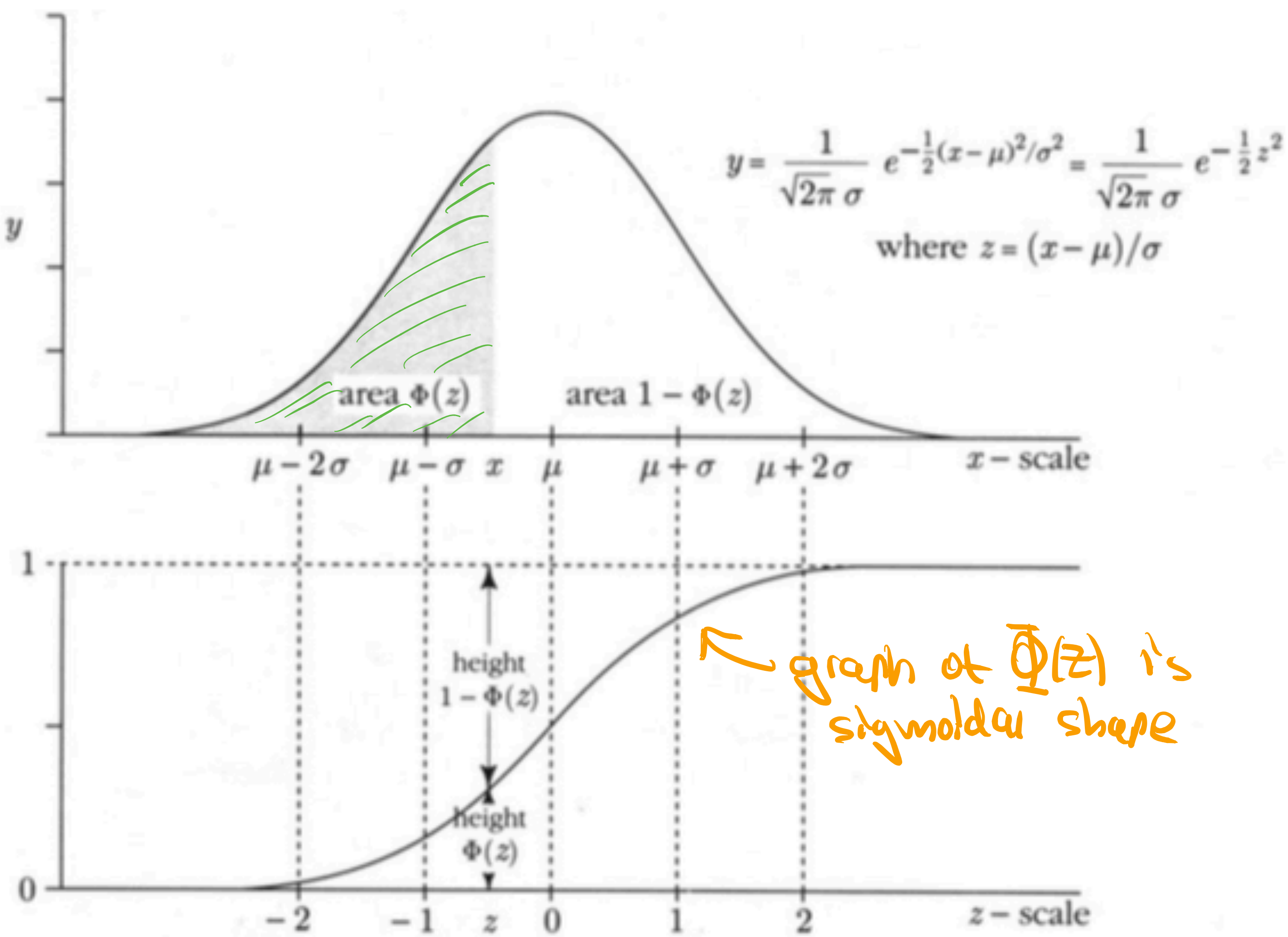
This makes $\mu = 0$ and $\sigma = 1$

density function

$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ std normal curve

Define cumulative distribution function (cdf) as $\Phi(z) = \int_{-\infty}^{z} \phi(z)\,dz$ ← area between $-\infty$ and z

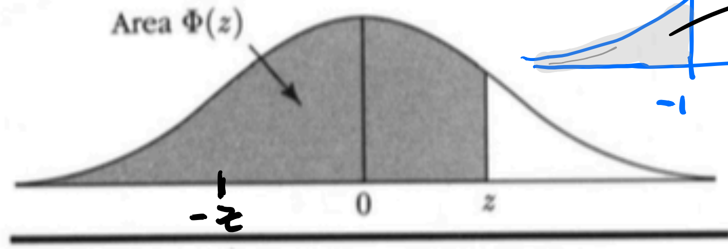
p95

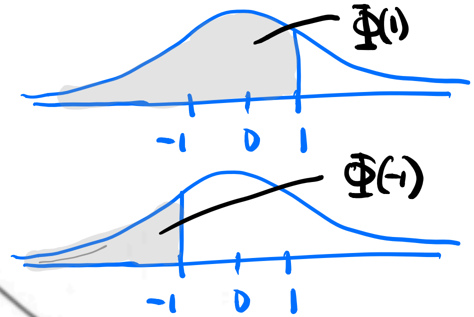


we can't solve integral $\Phi(z)$ but instead use look-up table.

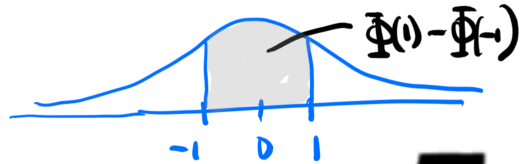
Notice table only gives values for $z \geq 0$

$\Phi(-z) = 1 - \Phi(z)$





$\Rightarrow$



Appendix 5

Normal Table

Table shows values of $\Phi(z)$ for z from 0 to 3.59 by steps of .01. Example: to find $\Phi(1.23)$, look in row 1.2 and column .03 to find $\Phi(1.2 + .03) = \Phi(1.23) = .8907$. Use $\Phi(z) = 1 - \Phi(-z)$ for negative z.

	.0	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Find area between 1 and −1 in std normal curve:

$\Phi(1) - \Phi(-1)$

$= \Phi(1) - (1 - \Phi(1))$

$= 2\Phi(1) - 1$

$2(.8413) - 1$

$= \boxed{.68}$

Find area between 2 and −2.

$2\Phi(2) - 1$

$= 2(.9772) - 1$

$= .95$

Find area between 3 and −3.

$= .997$

Empirical rule

Appendix *

Fact

① $k < np + p$ iff $P(k-1) < P(k)$

② $k > np + p$ iff $P(k-1) > P(k)$

③ $k = np + p$ iff $P(k-1) = P(k)$

Proof of Fact * above

First note that
$$\frac{\binom{n}{k}}{\binom{n}{k-1}} = \frac{\frac{\cancel{n!}}{k!(n-k)!}}{\frac{\cancel{n!}}{(k-1)!(n-k+1)!}} = \boxed{\frac{n-k+1}{k}}$$

$$\frac{P(k)}{P(k-1)} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \boxed{\frac{n-k+1}{k} \cdot \frac{p}{1-p}}$$

$$P(k-1) \;\text{☺}\; P(k)$$ where ☺ is $<$, $>$, or $=$

$$\Leftrightarrow \quad 1 \;\text{☺}\; \frac{P(k)}{P(k-1)}$$

$$\Leftrightarrow \quad 1 \;\text{☺}\; \frac{n-k+1}{k} \cdot \frac{p}{1-p}$$

$$\Leftrightarrow \quad k(1-p) \;\text{☺}\; (n-k+1)p$$

$$\Leftrightarrow \quad k - \cancel{kp} \;\text{☺}\; np - \cancel{pk} + p$$

$$\Leftrightarrow \quad k \;\text{☺}\; np + p$$

so $P(k-1) \;\text{☺}\; P(k) \Leftrightarrow k \;\text{☺}\; np + p$

where ☺ is $<$, $>$ or $=$