warm up 7:00-7:10 Stat 134 lec 5

2. A well shuffled deck is cut in half. Five cards are dealt off the top of one half deck and five cards are dealt off the top of the other half 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

when deck is cut in half, there is a different number of diamonds in each half deck (since 13 is odd) so probability of getting a diamond is different.

Also, not all trials are independent.

Last time

Announcement: exact Quiz logistics is still being worked out to give you max flexibility.

Sec 2.1 The Binomial Distribution

A Bernoulli trial has 2 outcomes, success (Prob p) and failure (Prob $q = 1-p$). (think of tossing a coin having prob p of landing head)

n independent Bernoulli trials, each with prob p of success, has a Binomial distribution written $\mathrm{Bin}(n,p)$.

For $k = 0, 1, 2, \ldots, n$ the Binomial formula says

$$P(k) = \binom{n}{k} p^k q^{n-k}$$

$P(k)$: the prob of getting exactly k out of n successes.

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

ex $n = 5$, $p = 1/2$

$P(0) = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = 1/32$

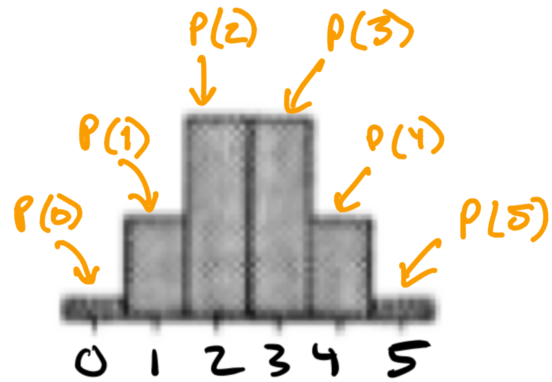
$P(1) = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = 5/32$

$P(2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = 10/32$

$P(3) = 10/32$

$P(4) = 5/32$

$P(5) = 1/32$



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.

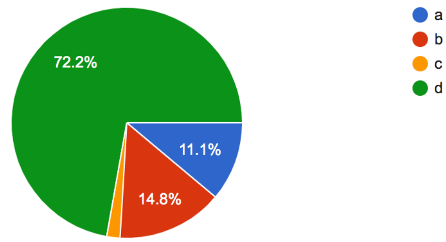
1 Ten cards are dealt off the top of a well shuffled deck. The binominal formula doesn't apply to find the chance of getting exactly three diamonds because:

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	After taking any card, the probability of getting a diamond changes. That also means the trials aren't independent, since P(diamond) will depend on how many diamonds are left in the deck

b	Probabilities are not independent because cards are not replaced; however, the probability of a card being a diamond is unconditional (same Bernoulli p for each card, not conditioning on events occurring before or after a draw) so the probability of a trial succeeding does not change.

② Binomial Dist — The mode and mean

What does the binomial distribution look like for different n, p?

Def^n (Mode)
The mode of the binomial distribution is the k such that $P(k) = \binom{n}{k} p^k (1-p)^{n-k}$ is largest
(i.e. the mode is the most likely outcome)

* ← Proof at end of notes using $\frac{P(k)}{P(k-1)}$, the consecutive odds ratio,

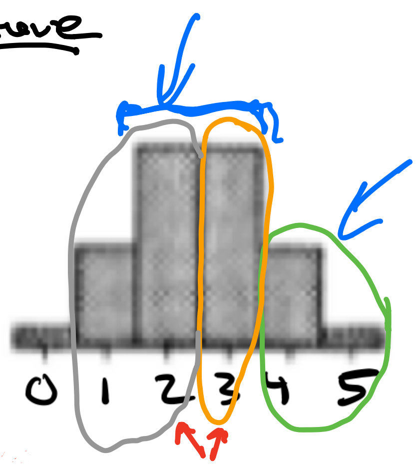
Thm* For $k \in \{1, 2, \dots, n\}$

given by Binomial Formula,

- $k < np + p$ iff $P(k-1) < P(k)$
- $k > np + p$ iff $P(k-1) > P(k)$
- $k = np + p$ iff $P(k-1) = P(k)$

Picture

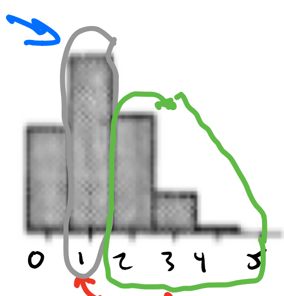
ex



$n = 5$
$k = 0, 1, 2, 3, 4, 5$
$p = 1/2$

$np + p = 5(1/2) + 1/2 = \boxed{3}$

ex



ex $n = 5$, $p = 1/4$, $k = 0, 1, 2, 3, 4, 5$

$np + p = 5 \cdot \frac{1}{4} + \frac{1}{4} = \boxed{1.5}$

The mode for the binomial distribution has 2 cases:

integers.

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

$\lfloor 3.2 \rfloor = 3$ the integer part of $np+p$

$\lfloor 3.7 \rfloor = 3$

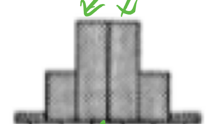
The mode is a measure of the center of the data.

However, the true center of your data is the
<u>expectation</u> (i.e the <u>mean</u>).

shown in chap 3

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

mode



mean = 2.5

This isn't usually an integer

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

If the mean is an integer is it the mode?

<u>Yes</u> $np \in \mathbb{Z} \Rightarrow np+p \notin \mathbb{Z}$ (assume $p \neq 0,1$)

element in $\quad \Rightarrow$ single mode with $\text{mode} = \lfloor np+p \rfloor$

$A \subseteq B$

$a \in B$

$= np = \text{mean}$

shown in chap 3

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

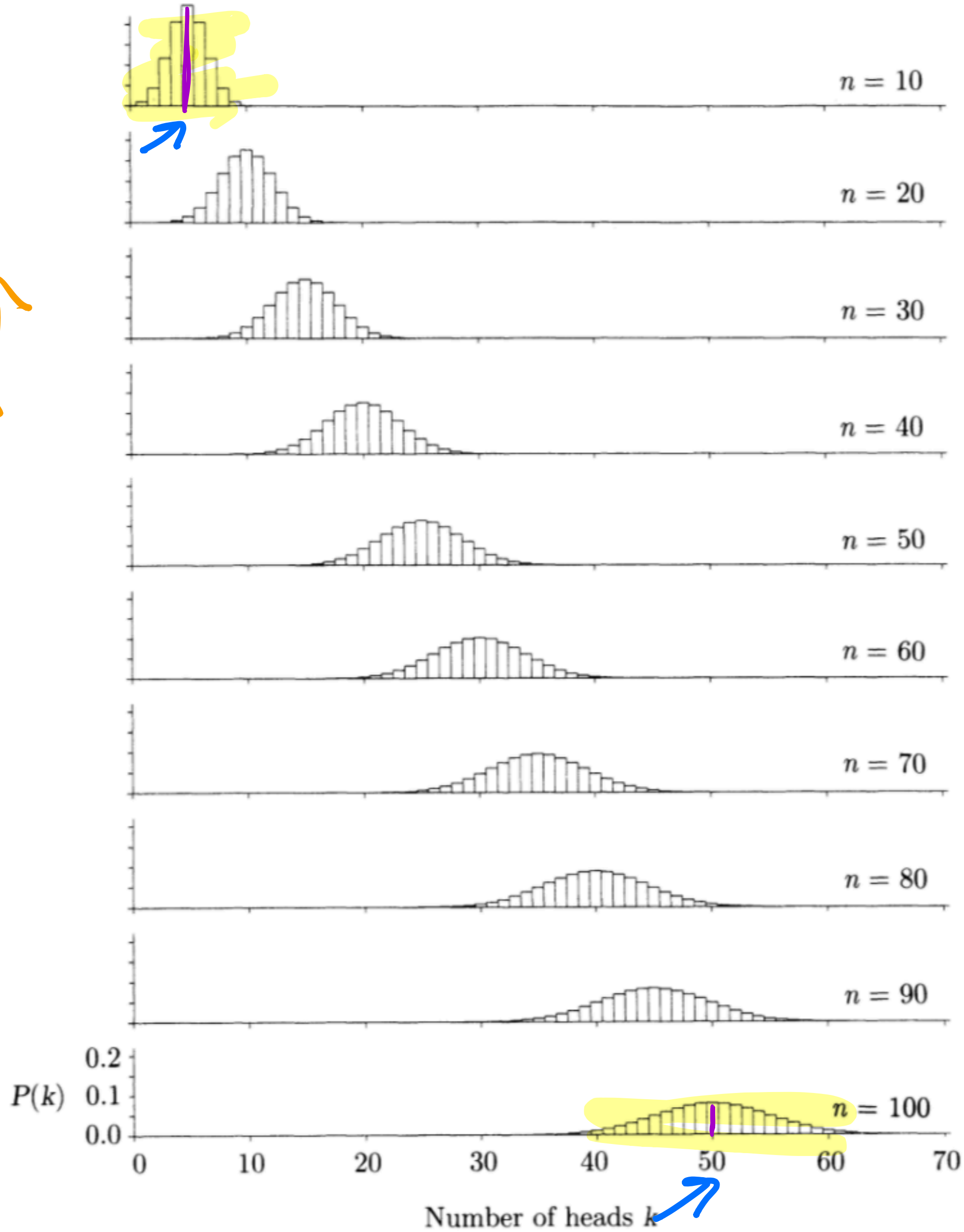
Notice that the spread around the mean gets larger as n gets bigger,

P88

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

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

Notice spread ↑ as n ↑



tinyurl.com/sept4-1

Stat 134

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 probability of getting any particular value goes down.

The binomonial formula gives

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

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

$X = \#$ heads

$\frac{X}{n} \rightarrow p = \frac{1}{2}$

③ sec 2.2 The normal distribution

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

Notice:

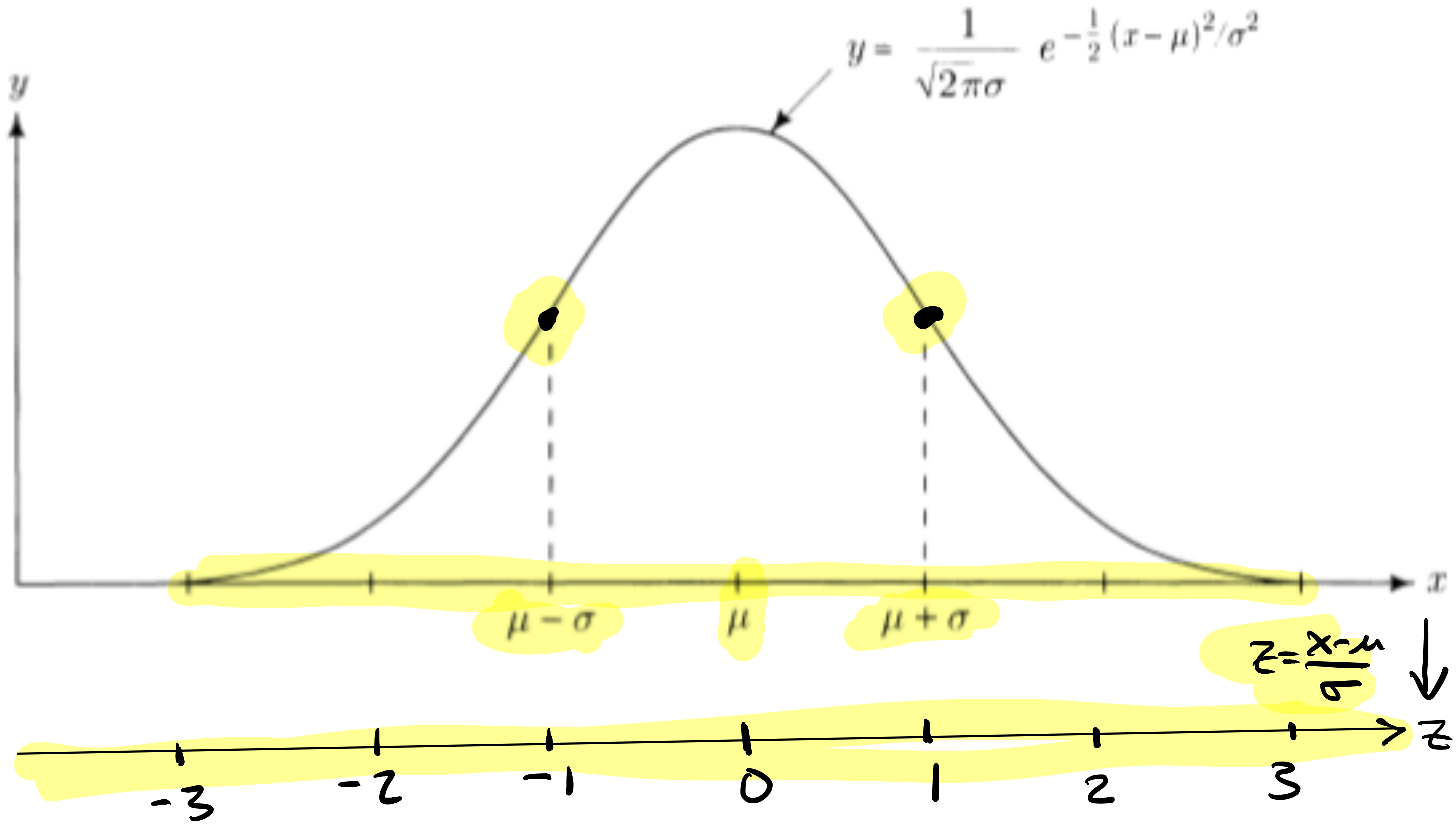
① two param μ = mean
σ = std dev

② inflection pts $\mu \pm \sigma$

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

P93

FIGURE 1. The normal curve.



To find the area under the curve it is convenient to make a change of coordinates

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

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

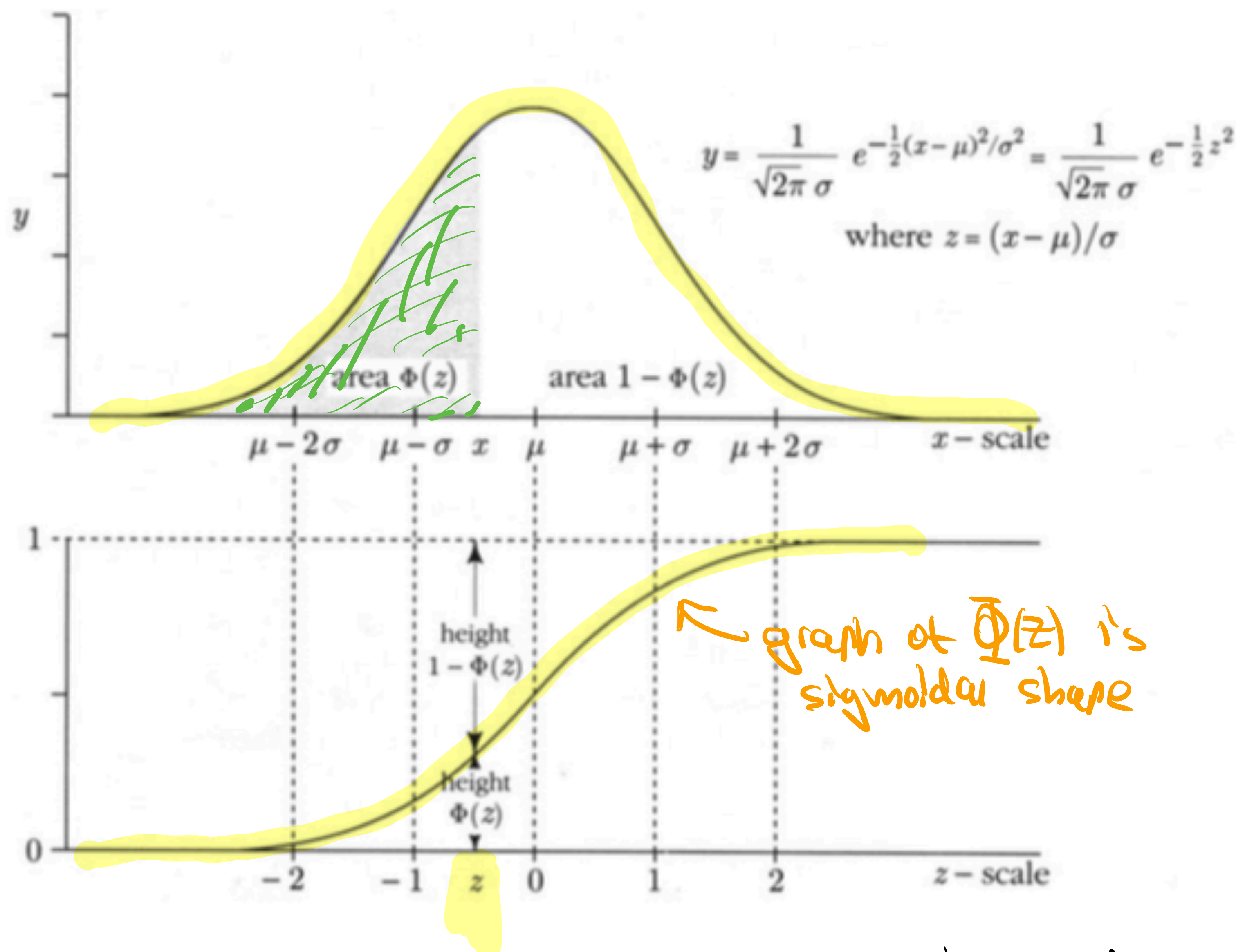
density function

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

<u>Define</u> cummulative distribution function (cdf)

as $\Phi(z) = \int_{-\infty}^{z} \phi(t)\,dt$ ← area between $-\infty$ and z

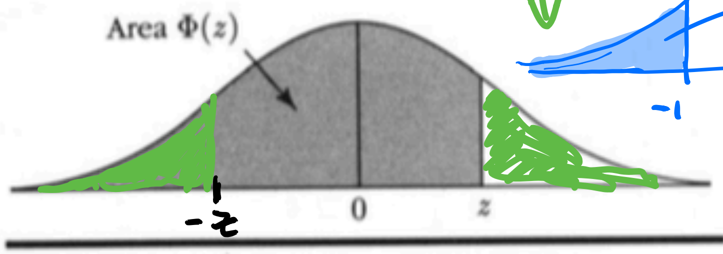
P95

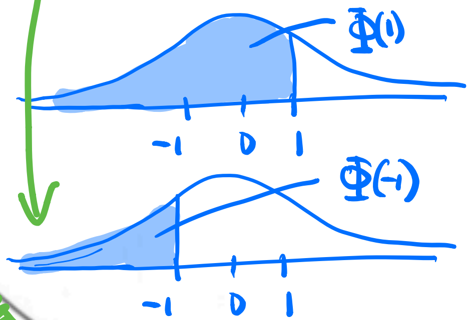


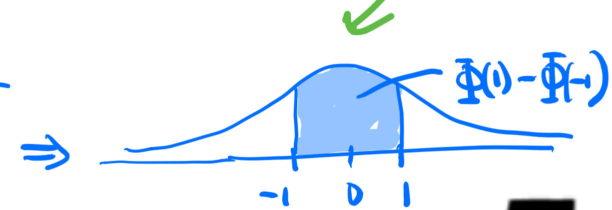
we can't solve integral $\Phi(z)$ but instead use look-up table.

Notice

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







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.

$2\Phi(2) - 1$

$2(.9772) - 1 = \boxed{.95}$

Find area between 3 and -3.

$= \boxed{.997}$

Empirical rule.

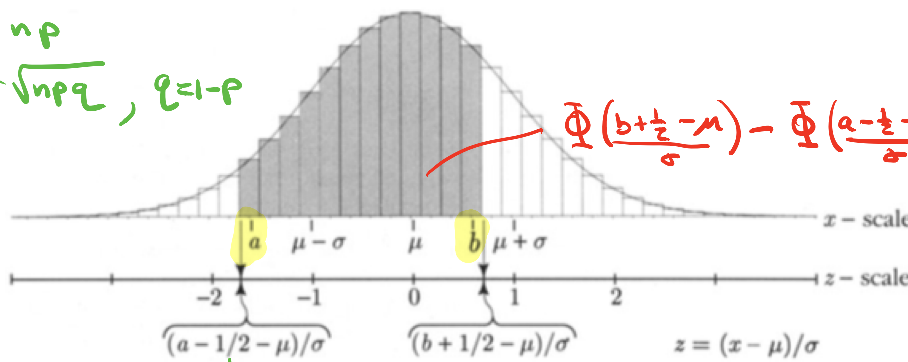
Normal Approx to binomial

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

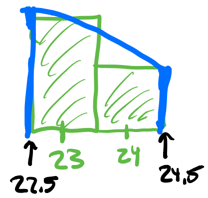
$\mu = np$

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

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



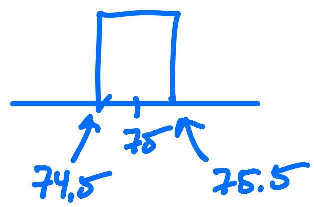
why $\frac{1}{2}$?



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.



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

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

$a = b = 75$

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

← putting x into std units

Exact value

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

Appendix

Thm

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 $P(k-1) = P(k)$

Pf/

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

Called the consecutive odds ratio

binomial formula

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

$$P(k-1) < P(k)$$

$$\iff 1 < \frac{P(k)}{P(k-1)}$$

$$\iff 1 < \frac{n-k+1}{k} \cdot \frac{p}{1-p}$$

$$\iff k(1-p) < (n-k+1)p$$

$$\iff k - \cancel{kp} < np - \cancel{pk} + p$$

$$\iff k < np+p$$

so $P(k-1) < P(k) \iff k < np+p$

similarly for $>$ or $=$ instead of $<$ $\square$