

— FRI, JAN 29 —

§2 REPEATED TRIALS AND SAMPLING

*WE CONSIDER AN EXPERIMENT THAT CONSISTS OF A SERIES OF INDEPENDENT BERNOUlli TRIALS.

* LET $X = \begin{cases} 1 & \text{IF TRIAL SUCCESS} \\ 0 & \text{IF TRIAL FAILS} \end{cases}$

$X \sim \text{Bernoulli}(p)$, i.e.

$$P(X=k) = \begin{cases} p & k=1 \\ q = 1-p & k=0 \\ 0 & \text{o/w} \end{cases}$$

* RECALL, IF $N = \#$ OF TRIALS UNTIL
A SUCCESS, THEN $N \sim \text{Geometric}(p)$:

$$P(N=k) = \begin{cases} q^{k-1} p & k=1, 2, \dots \\ . & \text{o/w} \end{cases}$$

* ANOTHER NATURAL QUESTION:
SUPPOSE WE RUN n TRIALS.
LET $Y = \#$ OF n TRIALS THAT
ARE SUCCESSES. THEN,

$$(\S 2.1) \quad Y \sim \text{Binomial}(n, p)$$

$$P(Y = k) = \begin{cases} \binom{n}{k} p^k q^{n-k} & k=0, 1, \dots, n \\ 0 & \text{o/w} \end{cases}$$

SEE APPENDIX 1:

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

$m!$ = "m FACTORIAL" = $m(m-1) \cdots 2 \cdot 1$

ONE WAY TO SEE THAT $\binom{n}{k}$ IS

WAYS TO CHOOSE k FROM n

OBJECTS IS BY INDUCTION,

NOTING:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Choose 1st and
k-1 from rest

→
Don't choose 1st and
k from rest.

GROUPS

A FAIR COIN IS TOSS'D
20 TIMES. GIVEN THAT THERE
WERE 12 H's, WHAT IS THE
CHANCE THAT THE 1ST TOSS
WAS A H?

Ans: $Y = \# H's$. $Y \sim \text{Binomial}(20, \frac{1}{2})$.

$$P(H_1 | Y=12)$$

$$= \frac{P(H_1 \wedge Y=12)}{P(Y=12)}$$

$$= \frac{\binom{12}{2} P(Y' = 11)}{P(Y=12)}, \quad Y' \sim \text{Binomial}(19, \frac{1}{2})$$

$$= \frac{\binom{12}{2} \left(\binom{19}{11}\right) \left(\frac{1}{2}\right)^{11} \left(\frac{1}{2}\right)^{19-11}}{\binom{20}{12} \left(\frac{1}{2}\right)^{12} \left(\frac{1}{2}\right)^{20-12}}$$

$$= \binom{19}{11} / \binom{20}{12}.$$

EASIER WAY :

$$\frac{\binom{19}{11}}{\binom{20}{12}}$$

SINCE ALL COMBINATIONS
WITH 12 H'S ARE EQUIVALENT
LIKELY.

EVEN EASIER:

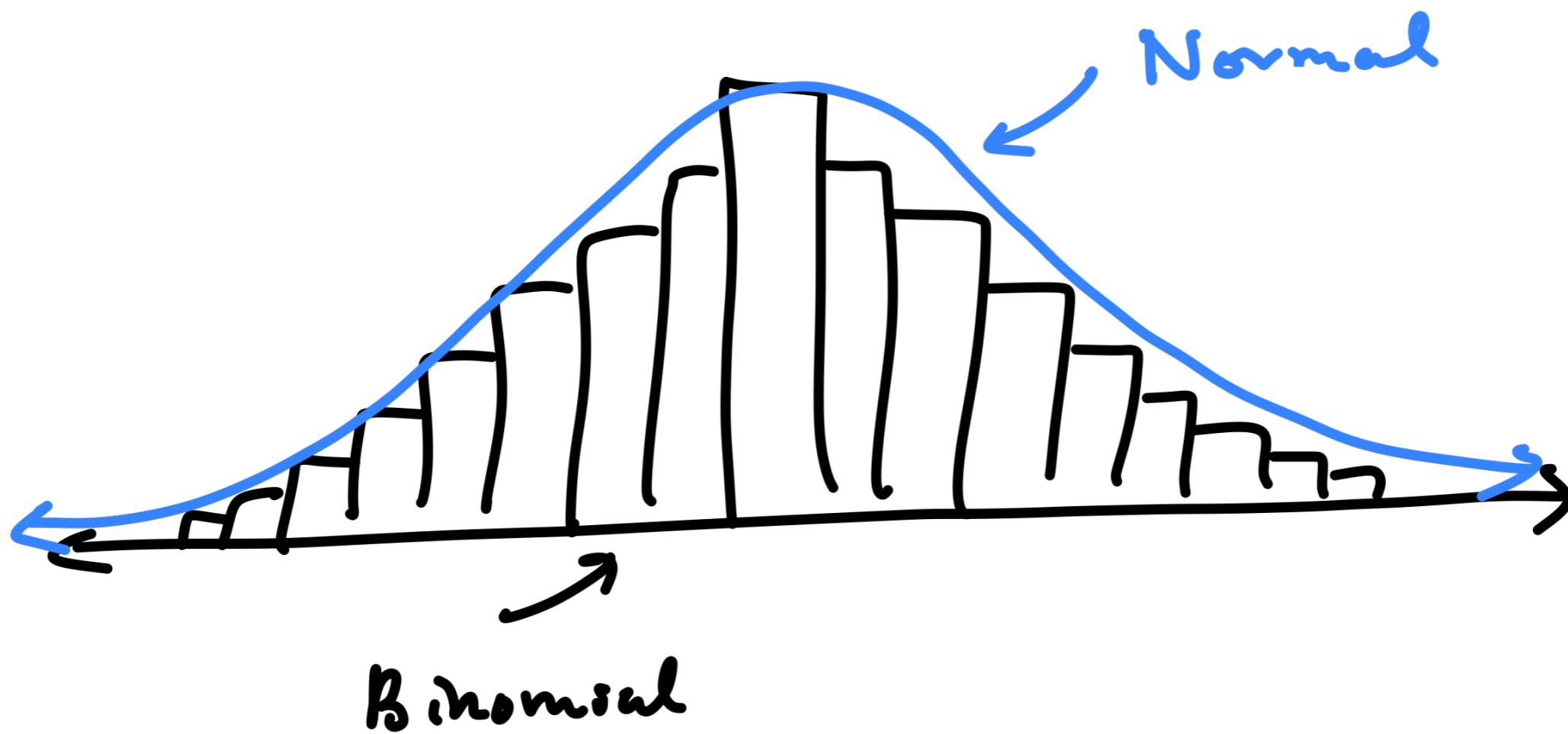
$$\frac{12}{20} \left[= \frac{\binom{19}{11}}{\binom{20}{12}} \right]$$

SINCE SUPPOSE

THERE ARE 12 H'S, AND
WE REVEAL CARDS ONE
AT A TIME. IF (1st CARD
REVEALED IS H) = $\frac{12}{20}$.

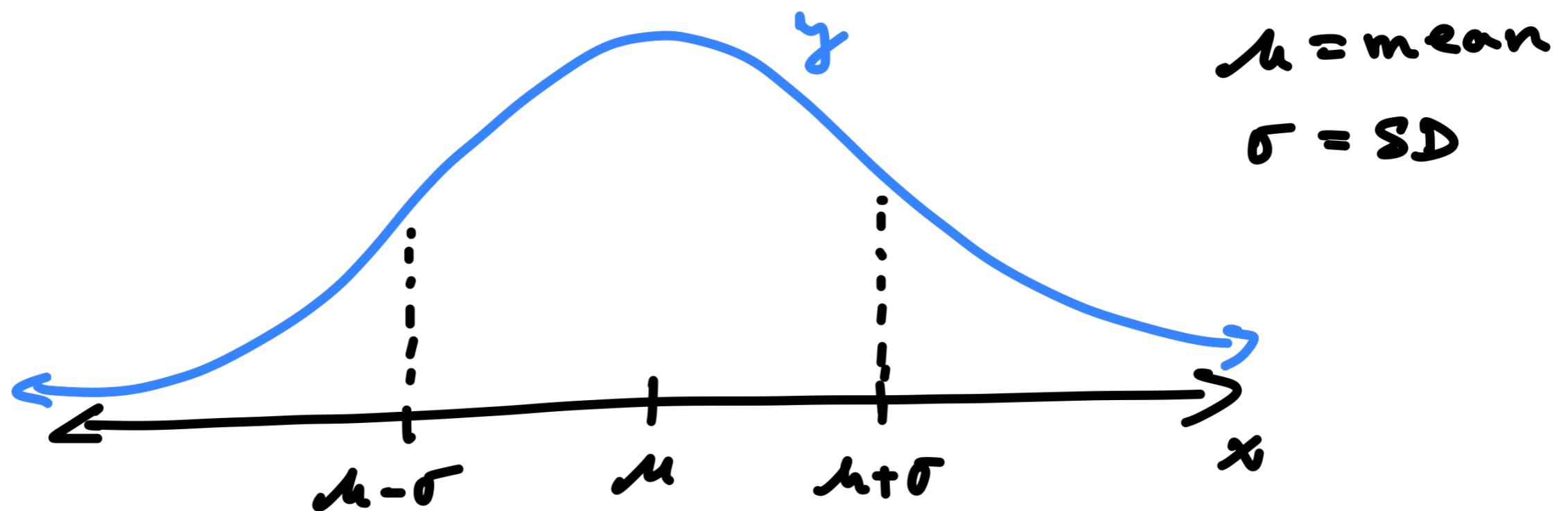
§2.2 NORMAL APPROXIMATION.

SEE P 88 - 89 : THE Binomial (n, p)
DISTRIBUTION HAS A (DISCRETE)
"BELL - CURVED" SHAPE.



* WHEN Binomial (n, p) HISTOGRAM IS PROPERLY RESCALED, AND AS $n \rightarrow \infty$, IT APPROACHES THE NORMAL BELLE CURVE:

$$y = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$



* THIS Allows us To APPROXIMATE
 $P(Bin(n, p) = k)$ WHEN n IS
LARGE.

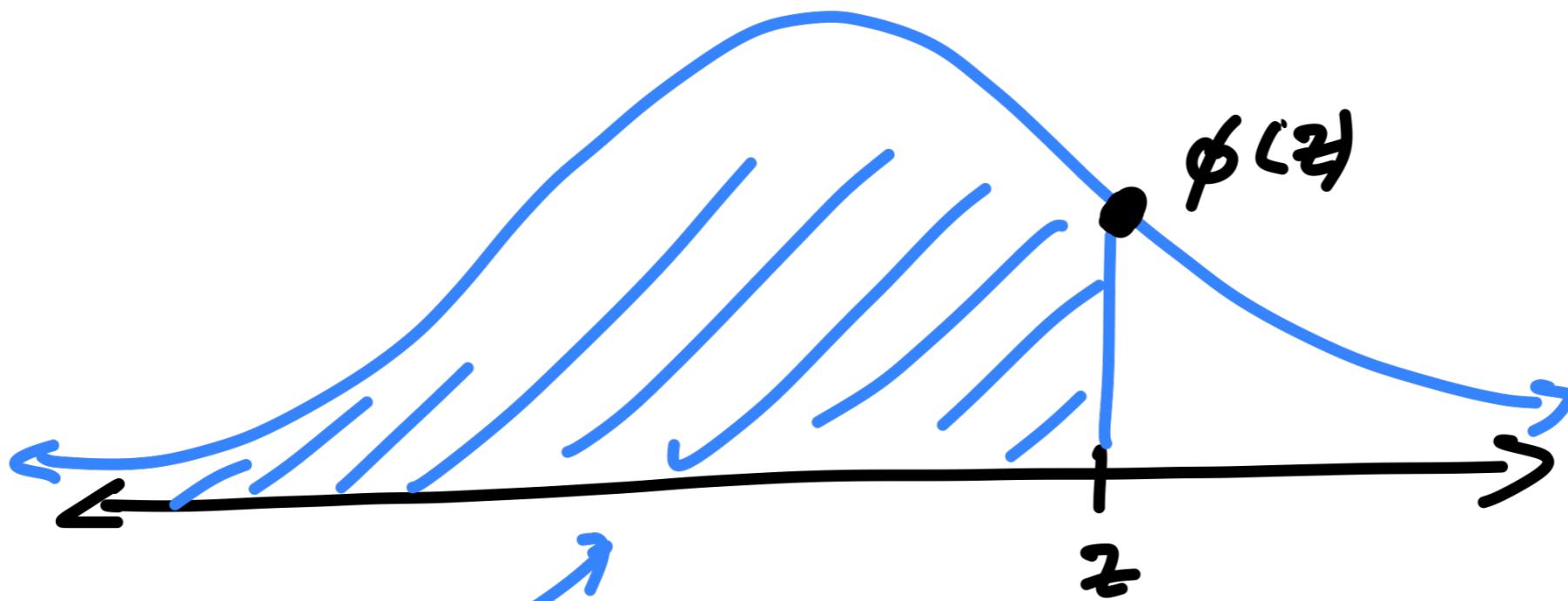
STANDARD NORMAL : $\mu = 0$
 $\sigma = 1$

Density Function

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

Cumulative Distribution Function
(CDF)

$$F(z) = \int_{-\infty}^z \phi(y) dy$$



$$\Phi(z) = \int_{-\infty}^z \phi(y) dy$$

BY CHANGE OF VARIABLES

(LET $z = \frac{x-\mu}{\sigma}$),

$$P(a \leq \text{NORMAL}(\mu, \sigma) \leq b)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

* NO CLOSED FORM FOR $\Phi(z)$, BUT

VALUES ARE TABULATED

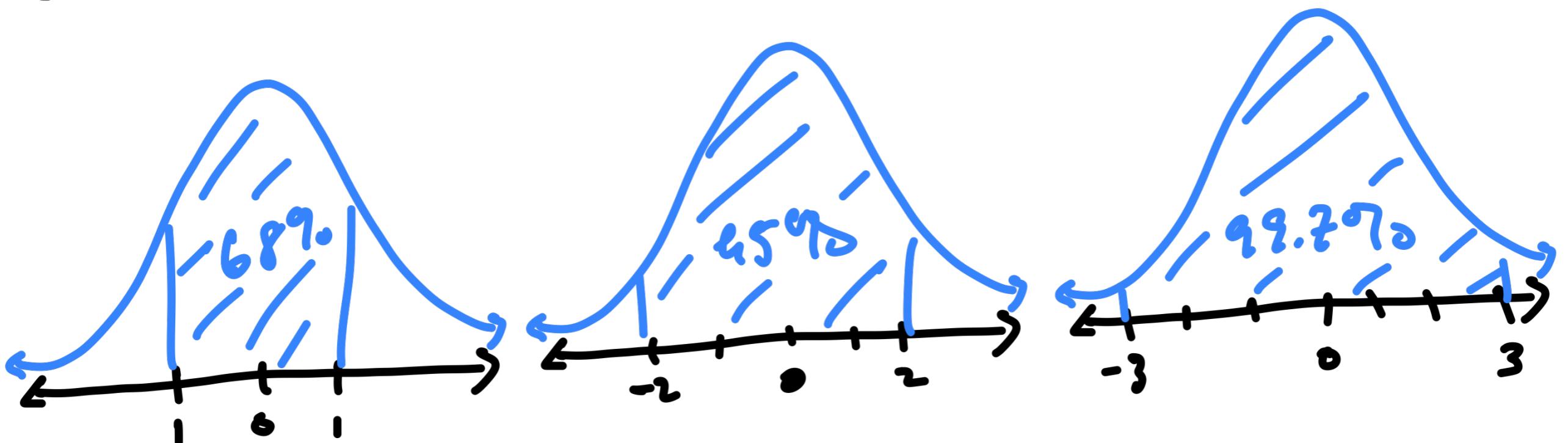
IN APPENDIX 5. [FOR $z \geq 0$, BUT

BY SYMMETRY, $\Phi(-z) = 1 - \Phi(z)$.]

AN IMPORTANT RULE OF THUMB:

$$\Phi(a, b) = \Phi(b) - \Phi(a)$$

$$\Phi(-1, 1) \approx 68\% \quad \Phi(-2, 2) = 95\% \quad \Phi(-3, 3) = 99.7\%$$



* WE'LL SEE IN § 3 THAT THE
EXPECTED VALUE = MEAN OF
 $X \sim \text{Binomial}(n, p)$ IS $\mathbb{E}X = np$.
* AND ITS $SD(X) = \sqrt{npq}$.

IT TURNS OUT (SEE § 2.3 - NOT COVERED)
THAT IF \sqrt{npq} IS LARGE, THEN
 $\text{Binomial}(n, p) \approx \text{Normal}(\mu, \sigma)$
WHERE $\mu = np$ AND $\sigma = \sqrt{npq}$.

NORMAL APPROXIMATION:

LET X = # SUCCESSES IN n
Bernoulli(p) TRIALS.

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

WHERE $\mu = E(X) = np$, $\sigma = SD(X) = \sqrt{npq}$.

* THE " $\frac{1}{2}$ " IS A "CONTINUITY CORRECTION"

THAT MAKES THE APPROX BETTER
WHEN \sqrt{npq} IS SMALL.

* RECALL $\Phi(-3, 3) = 99.7\%$

* $\Phi(-4, 4) = 99.999\%$, ...

LAWS THAT HOLD FOR LARGE n :

Square Root Law: # successes in n Bernoulli (p) trials will, with high probability, be in a small interval of width $O(\sqrt{n})$ centred around np .

Law of Large Numbers: The proportion of successes, with high prob., will be very close to p .