§4.3 Normal Distribution

Definition. Standard Normal Distribution

A RV Z has a standard normal distribution with zero mean and variance of 1 (i.e. $Z \sim N(0,1)$) if its pdf takes the form

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

for $-\infty < z < \infty$.

$\int_{-\infty}^{\infty} f_Z(z)dz = 1.$

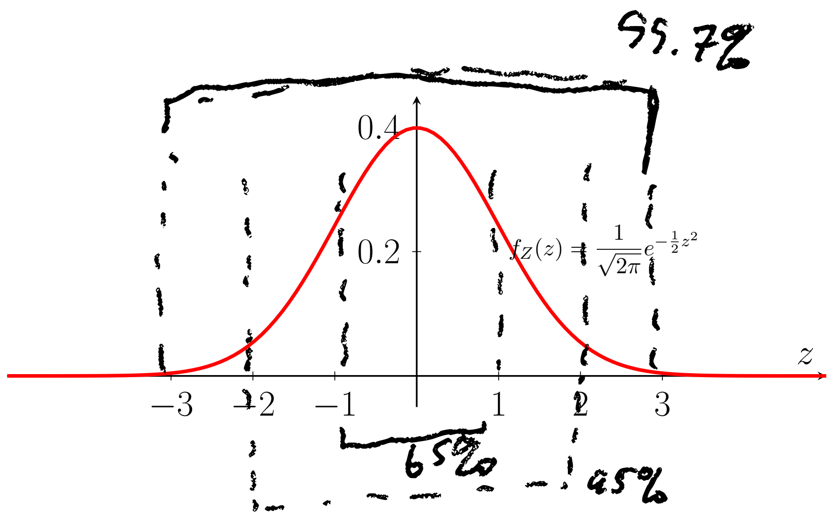
$\mu = E(Z) = 0$

$\sigma^2 = Var(Z) = 1$

$$X \sim N(\mu, \sigma^2)$$

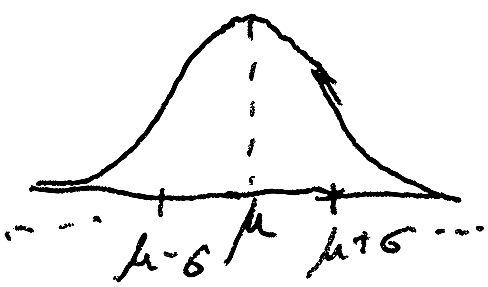
$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

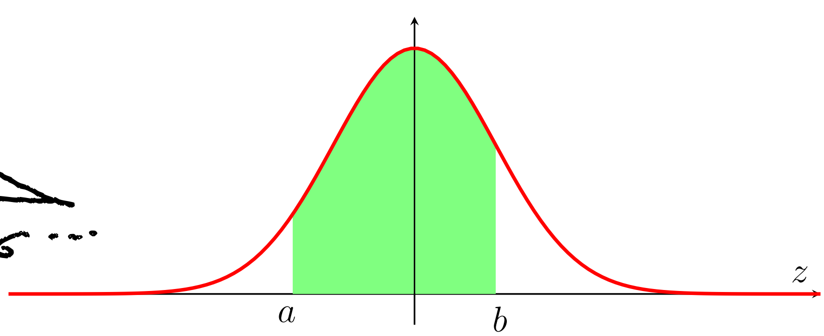
The graph is **Gaussian** curve (bell curve).



Z = # std. devs from mean

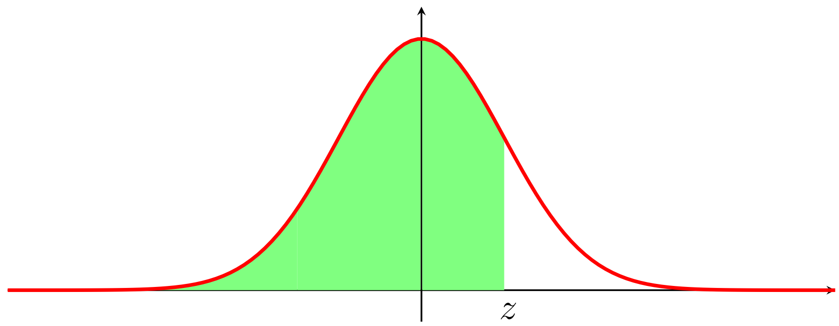
$$P(a \le Z \le b) = \int_a^b f_Z(z)dz.$$





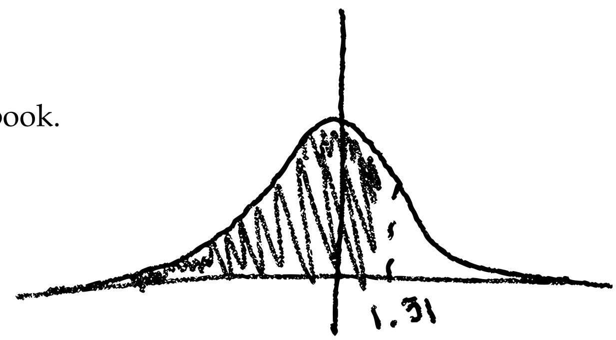
The **cdf** function is

$$F_Z(z) = \int_{-\infty}^{z} f_Z(u)du.$$



TI-83/TI-84: $\boxed{\text{2ND}} \to \boxed{\text{VARS}} \to \boxed{\text{2:normalcdf(}}$

We can also use the table in the **Appendix table A.1** of the book.



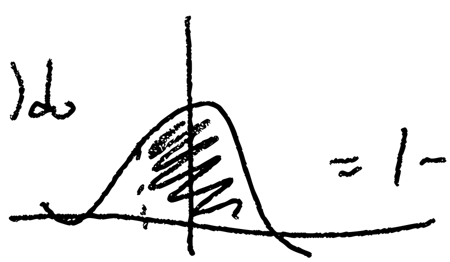
Example 1. Let Z be the standard normal ~~distribution~~ RV.

(1) Find $P(Z \le 1.31)$ or find $\int_{-\infty}^{1.31} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$ approx. $-\infty$

$= \text{normalcdf}(-10^{99}, 1.31, 0, 1)$

$= 0.9049$

(2) Find $P(Z \ge -0.45)$ or find $1 - \int_{-\infty}^{-0.45} f_Z(u)du$



$= 1 - P(Z < -0.45$

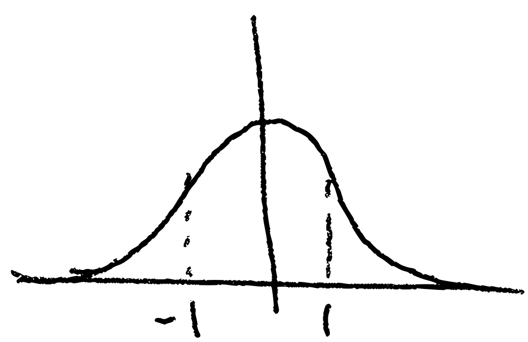
$= 1 - 0.32635$

$= 0.6736$

(3) Find $P(-1 < Z < 1)$ or find $\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

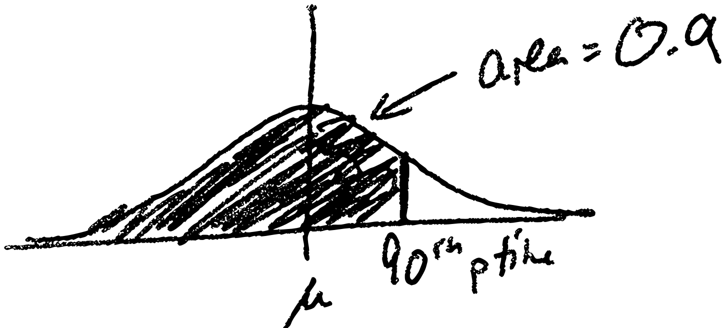
$= \text{normalcdf}(-1, 1, 0, 1)$

$= 0.68269$



Percentile:

The p^th **percentile** is a number ***a*** such that p% of the data fall below ***a***.



(4) Find the 70th percentiles. (Find a such that $P(Z \le a) = 0.7$)

$a = \text{invnorm}\left(\frac{p}{100}, \mu, \sigma\right)$

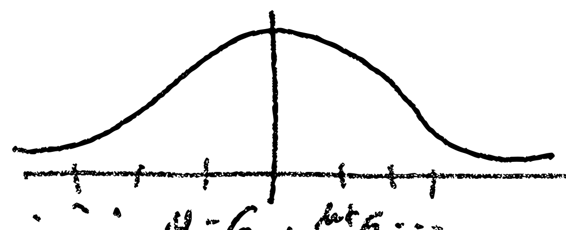
$= \text{invnorm}(0.7, \mu, \sigma)$ $(\mu = 0, \sigma = 1$ for std norm$)$

$= 0.5244$

(5) Find numbers a and b such that $P(a \le Z \le b) = 0.95$, assume a, b symmetric abt μ

$\{a, b\} = \text{invnorm}(0.95, 0, 1, \text{CENTER})$

$\boxed{a, b = \pm 1.96}$



Definition. Normal Distribution (a.k.a. Gaussian Distribution)

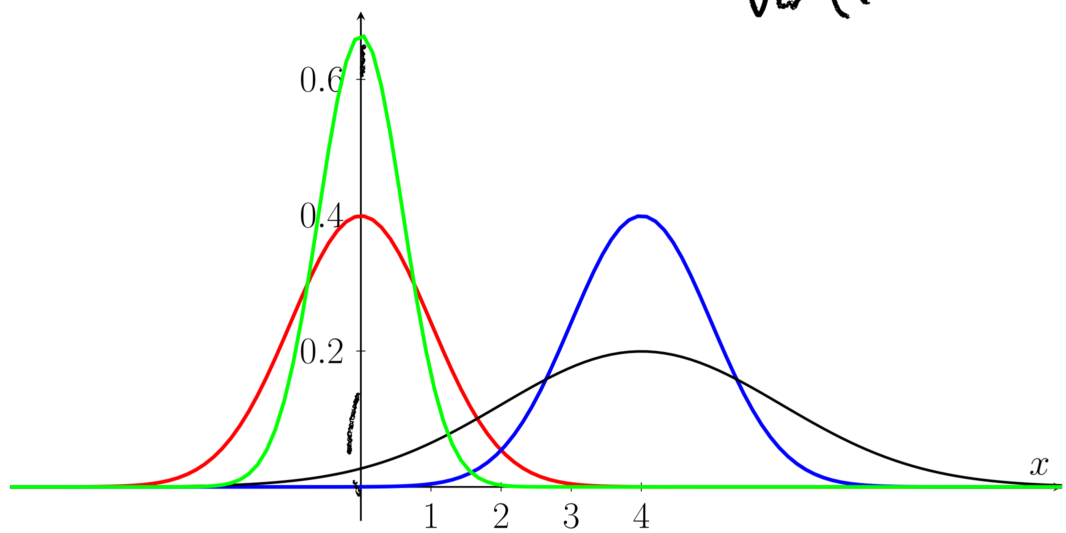
A RV X has a normal distribution with mean μ and variance σ^2 (i.e. $X \sim N(\mu, \sigma^2)$) if its pdf takes the form

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

for $-\infty < x < \infty$.

$\int_{-\infty}^{\infty} f_X(x)dx = 1.$

$\mathrm{Var}(X) = \sigma^2$



Red: $\mu = 0, \sigma = 1$. Green: $\mu = 0, \sigma = 0.6$. Blue: $\mu = 4, \sigma = 1$. Black: $\mu = 4, \sigma = 2$.

> **Theorem**
>
> The relationship between a standard normal distribution $Z \sim N(0, 1)$ and a normal distribution $X \sim N(\mu, \sigma^2)$ is that
>
> $$X = \mu + \sigma Z$$
>
> Z-score (# of std devs. from the mean)
>
> or write it in another way
>
> $$Z = \frac{X - \mu}{\sigma}.$$

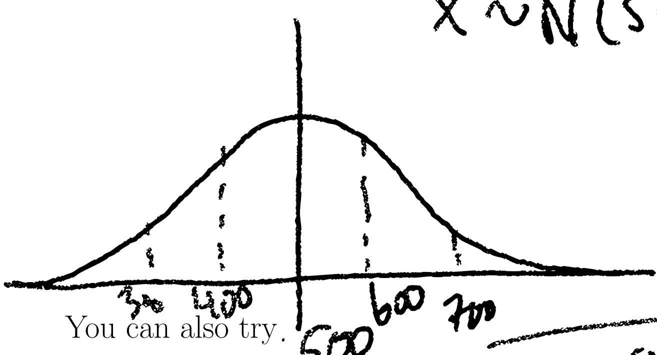
Example 2. Suppose the national Mathematics SAT scores is normally distributed with mean of 500 and a standard deviation 100. What percentage score between 400 and 600?

$\mu = 500, \ \sigma = 100$

$X \sim N(500, 100^2)$

Method 1:

$$\int_{400}^{600} f_X(x)\,dx = \text{normalcdf}(400, 600, 500, 100) = 68.2689\%$$



You can also try.

$P(300 \le X \le 700) = 0.9545$

$P(200 \le X \le 800) = 0.9973$

or... convert X to Z

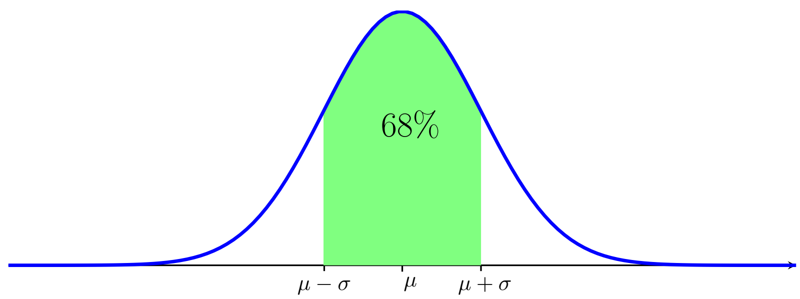
Z-score of 400 = −1

600 = +1

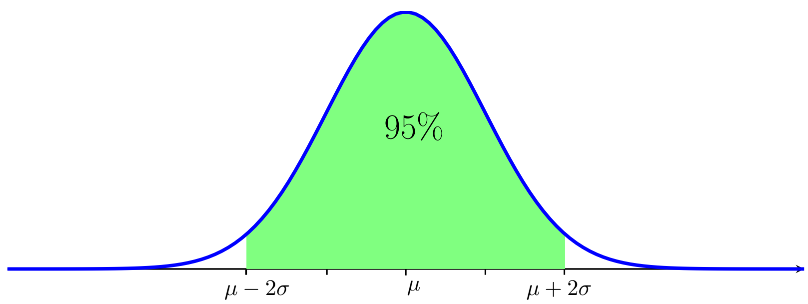
$\text{normalcdf}(-1, 1, \underbrace{0, 1}_{Z \text{ is std normal}})$

68-95-99.7 rule in the normal distribution $X \sim \text{Normal}(\mu, \sigma^2)$

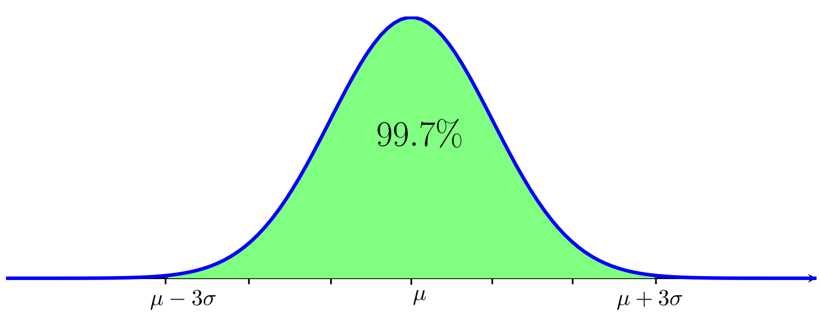
$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx \mathbf{68}.27\%$$



$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx \mathbf{95}.45\%$$



$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx \mathbf{99.7}3\%$$



Applications: Central Limit Theorem!

Suppose random variables $X_1, X_2, \cdots, X_n$ are independent and identically distributed (**IID**) from any distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

Recall:

Sample sum $X = X_1 + X_2 + \cdots + X_n$

Sample mean $\overline{X} = \frac{X}{n} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

From §3.9, $E(X) = n\mu$, $Var(X) = n\sigma^2$

Theorem: (**Central Limit Theorem**) for a sample mean

Under above assumption **IID** and when n is large enough, $n \geq 30$

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

Theorem: **CLT as a standard normal distribution**

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

or

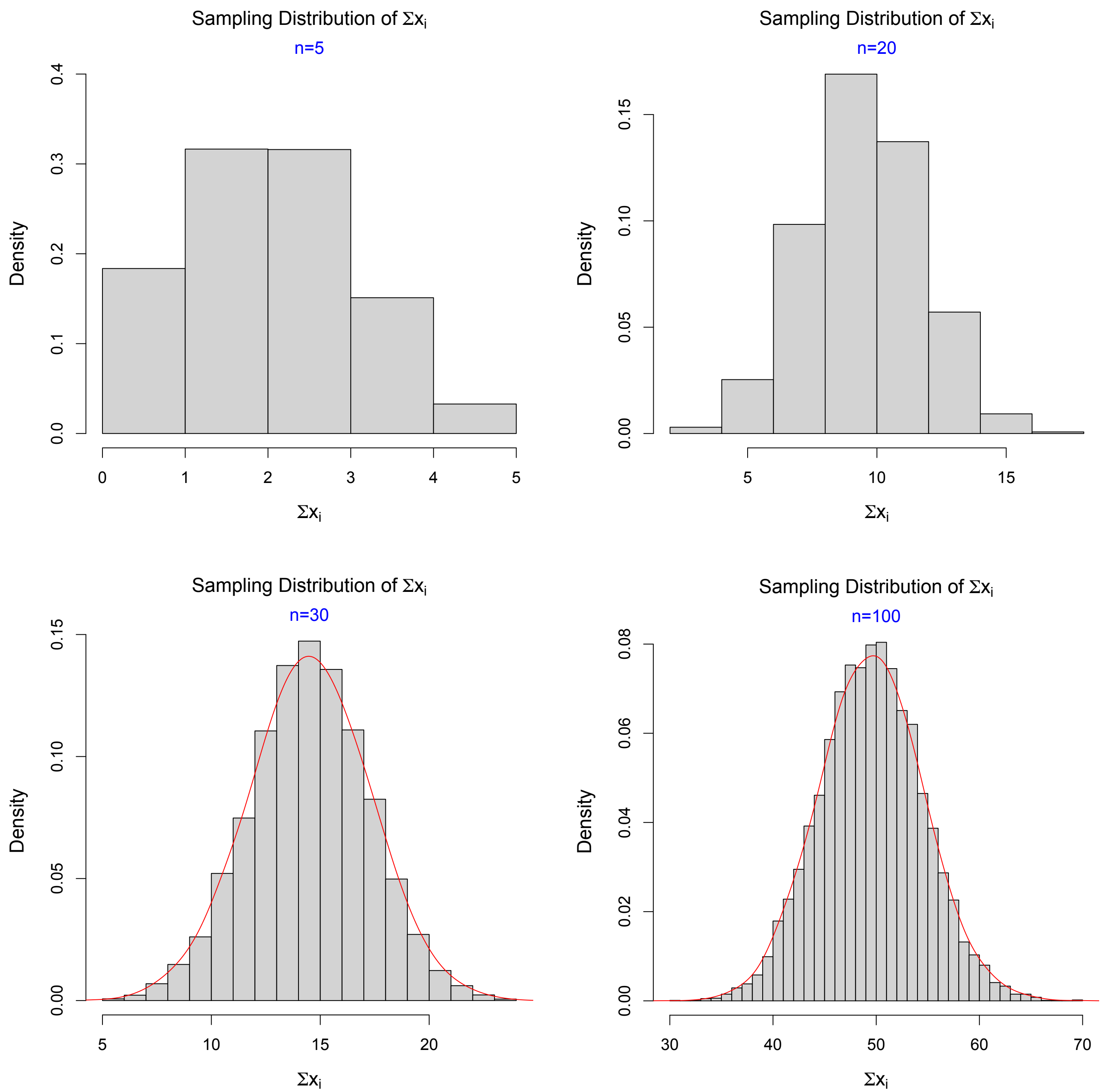
$$\frac{X - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

Central Limit Theorem (CLT) for Sample Sum

Let $\sum X$ be the total number of heads in n flips of a fair coin. Then

$$\sum X = \sum X_i = X_1 + X_2 + \cdots + X_n \text{ where } X_i \sim \mathit{Bernoulli}(0.5).$$

Make 10000 simulations for each case of n=5, n=20, n=30, n=100



The more times you flip the coin, the more likely the shape of the distribution of the sample means tends to look like a normal distribution curve.

We know that the sum of n *Bernoulli*(p) variables is *Binomial*(n, p). If we plot the pdf for different values of n, we see that when n is large, the binomial distribution tends to a normal distribution.

Theorem: CLT for a sample sum

$\sum X \sim N(n\mu, n\sigma^2)$, $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

More precisely,

$$\lim_{n\to\infty} P\Big(a \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le b\Big) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz$$

Summary of the CLT

1. If the sample size is sufficiently large (n $\geq$ 30) then the distribution of the sample means is approximately normally distributed.

2. If the population is normally distributed, then the result of CLT holds for samples of any size (i.e., the sampling distribution of the sample means is approximately normal even for samples of size less than 30).

Example 3. Time between text message arrivals
Suppose the time X between text message arriving on your cell phone is distributed with mean 25 minutes and standard deviation 25 minutes. You record the times between your next 50 messages. What is the probability that their average exceeds 21 minutes?

↳ CLT applies

X = time between 2 texts

$\overline{X}$ = sample mean, $\overline{X} \sim N\left(25, \frac{25^2}{50}\right)$

$\mu = 25$, $\sigma = 25$

$$P(\overline{X} > 21) = \int_{21}^{\infty} P_{\overline{X}}(\bar{x})\,d\bar{x}$$

$= \text{normalcdf}\left(21, 10^{99}, 25, \frac{25}{\sqrt{50}}\right)$

$= 0.8711$

Example 4. In a recent study, the mean age of tablet users is 35 years. Suppose the standard deviation is ~~ten years~~ and the sample size is 39.

(a) What are the mean and standard deviation for the sum of the ages of tablet users? What is the distribution?

(b) Find the probability that the sum of the ages is between 1,400 and 1,500 years.

(c) Find the 90th percentile for the sum of the 39 ages.

a) since $n > 30$, CLT applies

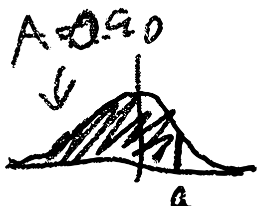
X = age of tablet users

$$\mu\left(\sum_1 X\right) = n\mu = 39 \cdot 35 = 1365$$

$$\sigma\left(\sum_1 X\right) = \sqrt{n}\,\sigma = 62.45$$

b) $\sum_1 X \sim N(1365, 62.45)$

$$P\left(1400 < \sum_1 X < 1500\right) = \text{normalcdf}(1400, 1500, 1365, 62.45) = 0.2723$$

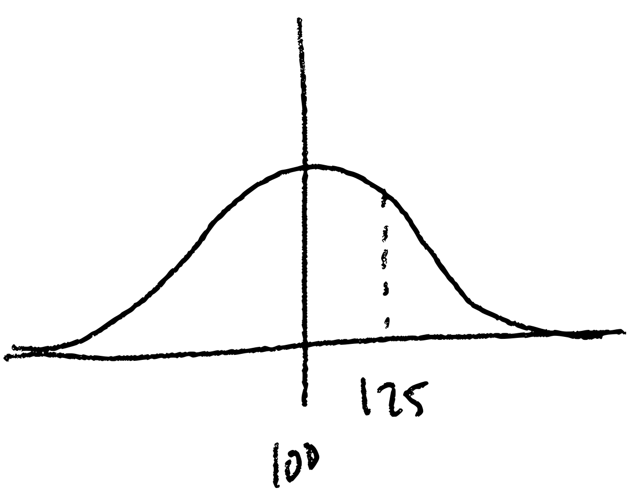
c) 

$a = \text{invnorm}(0.9, 1365, 62.45, \text{LEFT}) = 1445.0329$

Example 5. Standardized IQ tests are designed so that their scores have a normal distribution in the general population with a mean of 100 and the standard deviation 15. Randomly choose 1245 people, person A gets a score 125. How many people have a better score than this person A?

Hint: First find the probability that a person's score is higher than 125, P(score >125). Then multiply with the population 1245.

$\mu = 100$, $\sigma = 15$ let X = IQ score $\sim N(\mu, \sigma^2)$



$$P(X > 125) = \text{normalcdf}(125, 10^{99}, 100, 15) = 0.0478$$

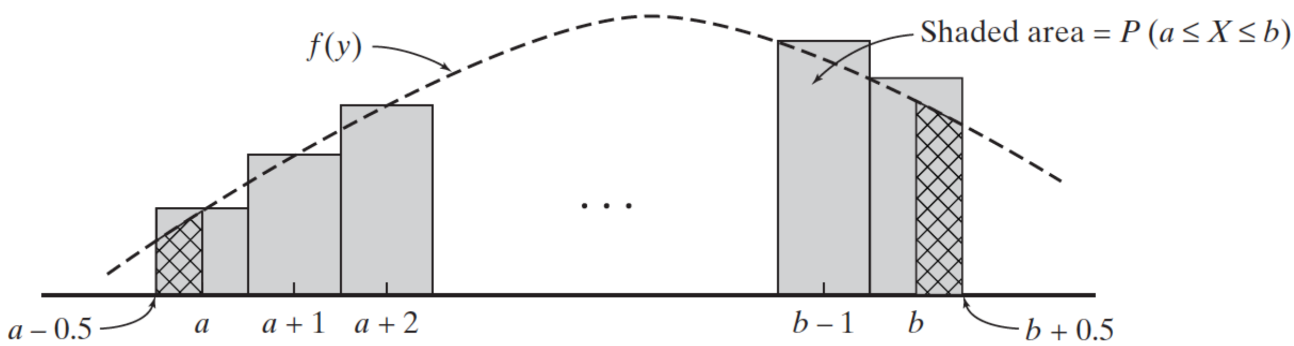
$$0.0478 \cdot 1245 = 59.4989 \approx 59$$

Approximating the Binomial Distribution with the Normal Distribution

In some situations, the normal distribution may be useful in approximating probabilities for a binomial random variable. When n is very large, we find that the normal distribution can be used to provide a close working approximation of binomial probabilities and others.

Using the normal distribution in this case involves a little adjustment of our X-values. Because we are approximating a discrete distribution with a continuous one, we need to add 0.5 to our X-value before we compute a Z. This is called a **continuity correction.**

One problem is that the binomial distribution is discrete. We need to use the **Continuity Correction** to fix the difference.



$$P_{Bin}(a \le X \le b) \approx P_{Norm}(a - 0.5 \le X \le b + 0.5)$$

Recall! If X is a binomial RV with parameters, then $\boldsymbol{E(X) = np}$ and $\boldsymbol{Var(X) = np(1-p)}$.

$$\boxed{P\Big(a \le X \le b\Big) \approx \textbf{normalcdf}(a - 0.5, b + 0.5, np, \sqrt{np(1-p)})}$$

Correction for Continuity:

- $P_{Bin}(X = a) \approx P_{Norm}(a - .5 \le X \le a + .5)$
- $P_{Bin}(X \le a) \approx P_{Norm}(X \le a + .5)$
- $P_{Bin}(X < a) = P_{Bin}(X \le a - 1) \approx P_{Norm}(X \le a - 1 + .5) = P_{Norm}(X \le a - .5)$
- $P_{Bin}(X \ge a) \approx P_{Norm}(X \ge a - .5)$
- $P_{Bin}(X > a) = P_{Bin}(X > a + 1) \approx P_{Norm}(X \ge a + 1 - .5) = P_{Norm}(X \ge a + .5)$
- $P_{Bin}(a \le X \le b) \approx P_{Norm}(a - .5 \le X \le b + .5)$

Example 6. Eighty percent of older Americans wear glasses. What is the probability that 45 or more out of a random sample of 64 older Americans will wear glasses? Use the continuity correction.

$p = 0.8,\ n = 64,\ \sigma = 3.2$

$X \sim N(51.2,\ 51.2(0.2)),\quad Y = B(64, 0.8)$

$P(Y \geq 45) \approx P(X > 44.5)$

$= \text{normalcdf}(44.5,\ 10^{99},\ 51.2,\ 3.2)$

$\boxed{= 0.98186}$

Example 7. State Tech's basketball team, the Fighting Logarithms, have 70% foul-shooting percentage. Approximate the probability that out of their next one hundred free throws, they will make between seventy-five and eighty, inclusive. Use the continuity correction.

$p = 0.7,\ n = 100,\ \mu = 70,\ \sigma = \sqrt{np(1-p)}$

$= \sqrt{70(0.3)} = 4.582$

$X \sim N(70,\ 4.582^2),\quad Y \sim B(100,\ 0.7)$

$P(75 \leq Y \leq 80) \approx P(74.5 < X < 81.5)$

$= \text{normalcdf}(74.5,\ 81.5,\ 70,\ 4.582)$

$= 0.15698$

$\boxed{P(75 \leq Y \leq 80) \approx 0.15698}$

Example 8. The exam score of all students (8 sections, 530 students) are recorded. Assuming the distribution of the score is normal with a mean of 83 and a standard deviation of 7.

(1) Let $\overline{Y}$ be the average score in section 1 (with 71 students). What is the probability that the average score will exceed 84?

$n > 30$, CLT applies

let X = avg in a classroom, by CLT $X \sim N(83, \frac{49}{71})$

$P(X > 84) = \text{normalcdf}(84, 10^{99}, 83, \sqrt{\frac{49}{71}})$

$\boxed{= 0.11435}$

(2) Randomly choose a student's score Yi. What is the probability that the score will exceed or equal 90 ?

Y_i is a ...

$Y_i \sim N(83, 7^2)$

$P(Y \geq 90) = \text{normalcdf}(90, 10^{99}, 83, 7) = \boxed{0.15865}$

(3) What is the probability that more than 5 of the student's scores will exceed or equal 90 in section 1 with 71 students?

Z = # of students w/ ≥ 90.

$Z \sim B(71, 0.15865)$

$P(Z > 5) = 1 - P(Z \leq 5)$

$= 1 - \text{binomcdf}(71, 0.1587, 5) = 0.9773$