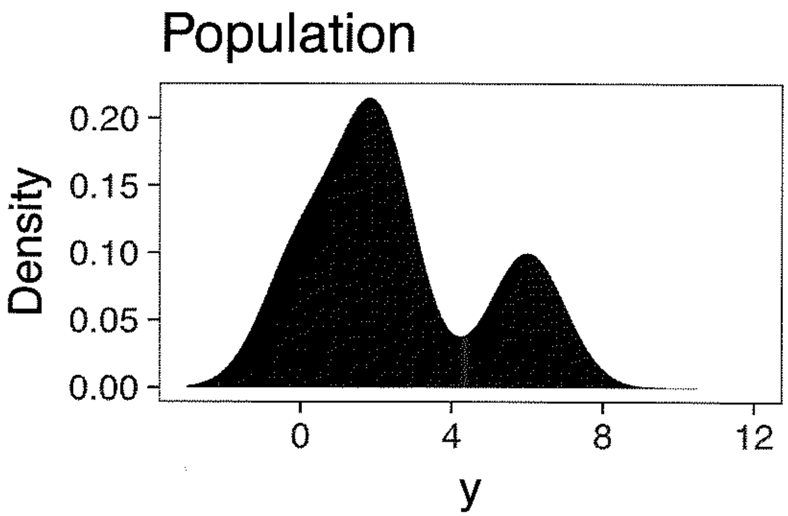
Your Turn

Consider the following population distribution.



The following three histograms represent:

1. A sample of size 1000 from the population B
2. 1000 sample means of samples of size 10 from the population C $Var(\bar{Y}) \approx \frac{\sigma^2}{10}$
3. 1000 sample means of samples of size 100 from the population A $Var(\bar{Y}) \approx \frac{\sigma^2}{100}$

Shape
Variance

Your turn

Using the CLT to approximation the sampling distribution

Population: $\sim (\mu = 20, \sigma^2 = 4)$

Sample: $n = 16$ i.i.d from population, $Y_1, \ldots, Y_{16}$

Sample statistic: Sample mean, $\overline{Y} = \frac{1}{n}\sum_{i=1}^{16} Y_i$

What is the approximate distribution for $\overline{Y}$?

$$\overline{Y} \mathrel{\dot\sim} N\left(20, \frac{4}{16}\right)$$

$$\overline{Y} \mathrel{\dot\sim} N\left(20, \frac{1}{4}\right)$$

Var

Applying CLT

Same setup: we have sample of size, $n = 16$ from a population with population mean $\mu = 20$ and population variance $\sigma^2 = 4$.

What is the probability the sample mean is less than 20.5?

CLT says:

$$\bar{Y} \overset{\cdot}{\sim} N\left(20, \frac{1}{4}\right)$$

$$P\left(\bar{Y} \leq 20.5\right)$$

continuous

$\leq$, $<$ } doesn't matter which

Using a Standard Normal Probability Table

To use a Standard Normal table, we first need to transform our random variable ($\overline{Y}$) to a Standard Normal.

We can convert a probability for $\overline{Y}$ to a standard Normal probability by:

1. Subtracting the mean of $\overline{Y}$ from both sides, then
2. Dividing both sides by the standard deviation of $\overline{Y}$ (square root of the variance)

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

then $\frac{Y-\mu}{\sigma} \sim N(0,1)$

$$P(\overline{Y} \leq 20.5) =$$

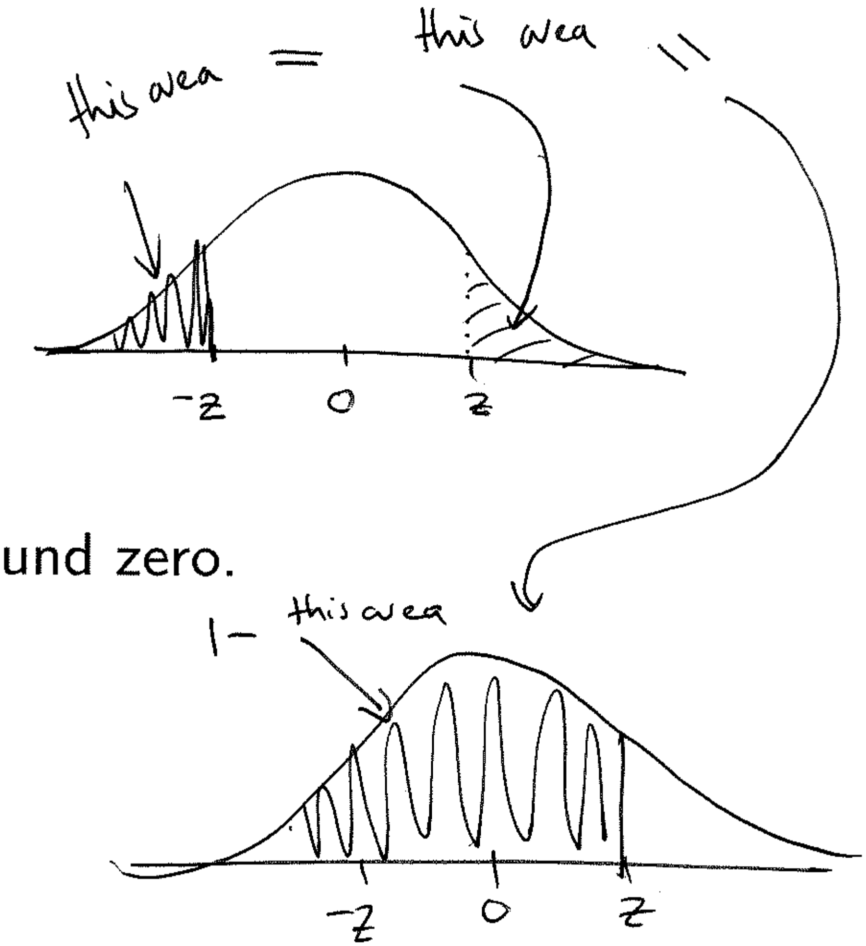
$$P\left(\frac{\bar{Y} - 20}{\sqrt{1/4}} \leq \frac{20.5 - 20}{\sqrt{1/4}} \right)$$

$\bar{Y} \dot{\sim} N(20, \frac{1}{4})$

$$P(Z \leq 1)$$

$Z \sim N(0, 1)$

Using a Standard Normal Probability Table



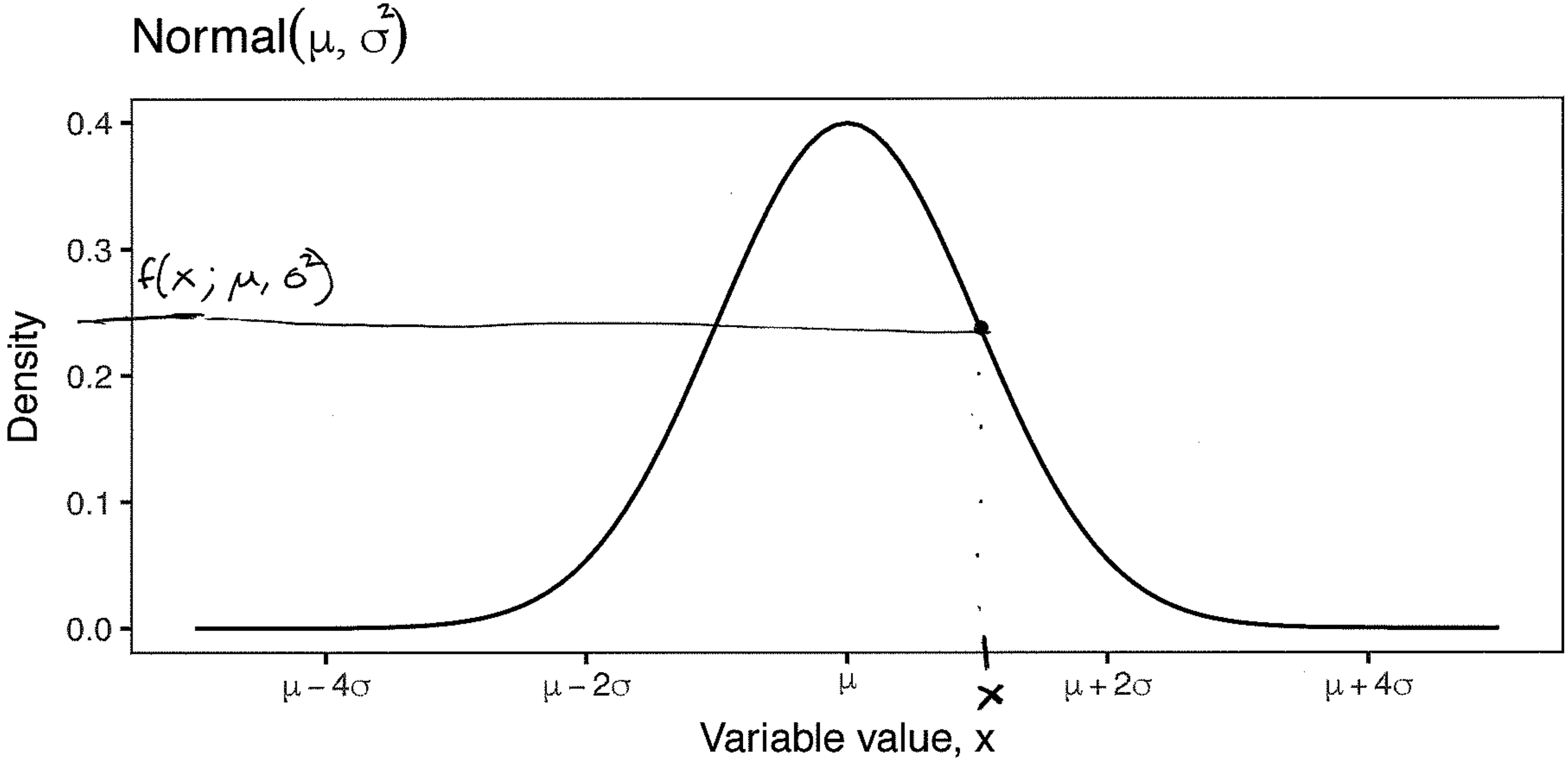
What if z is negative?

Standard Normal is symmetric around zero.

$$P(Z \leq -z) = 1 - P(Z \leq z)$$

Density

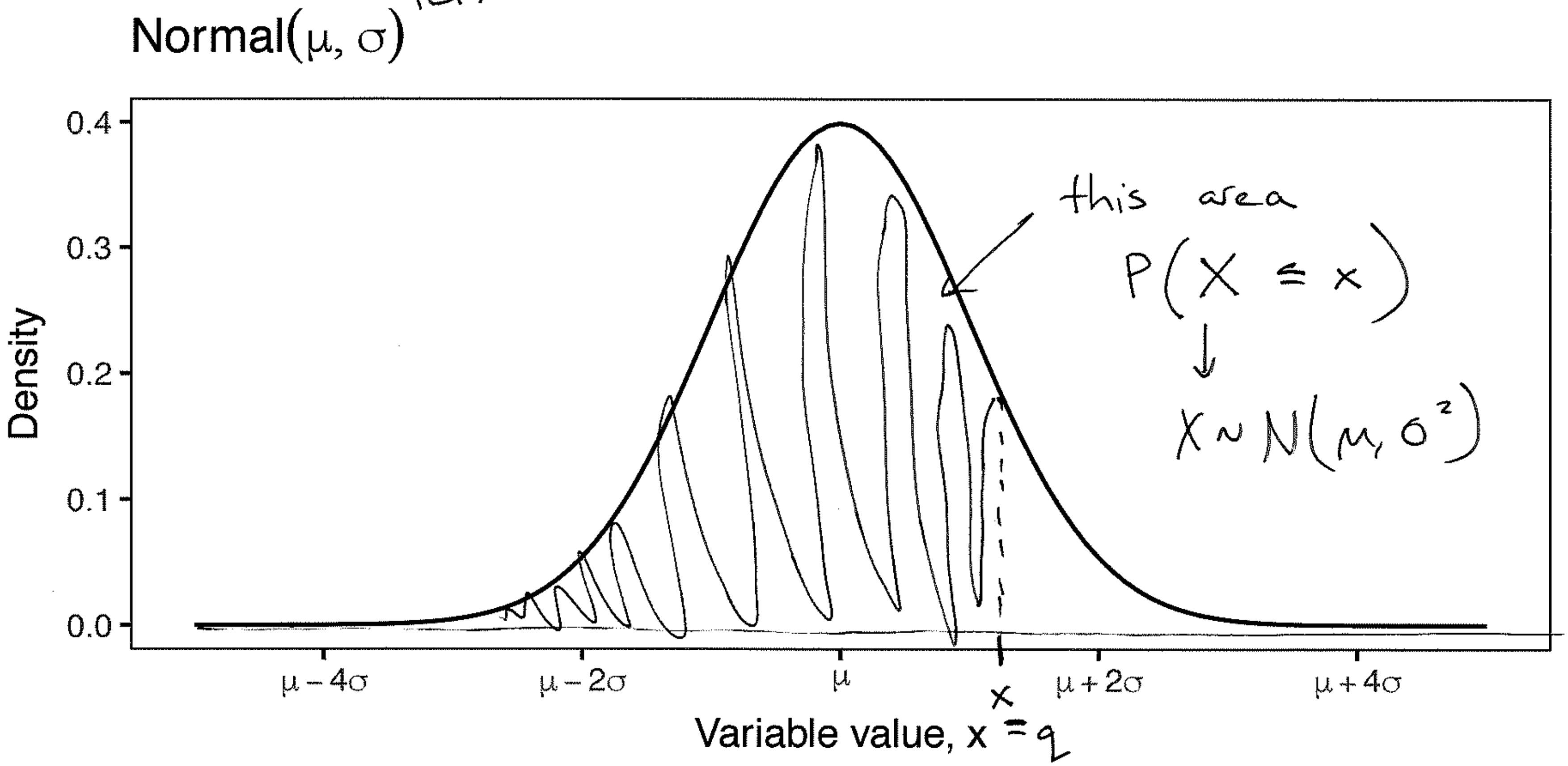
Density: height of probability density function at x: $f(x; \mu, \sigma^2)$



In R: `dnorm(x, mean = mu, sd = sigma)`

Cumulative Probability

Cumulative Probability: the area under the probability density function to the ~~right~~ left of x: $F(x; \mu, \sigma)$

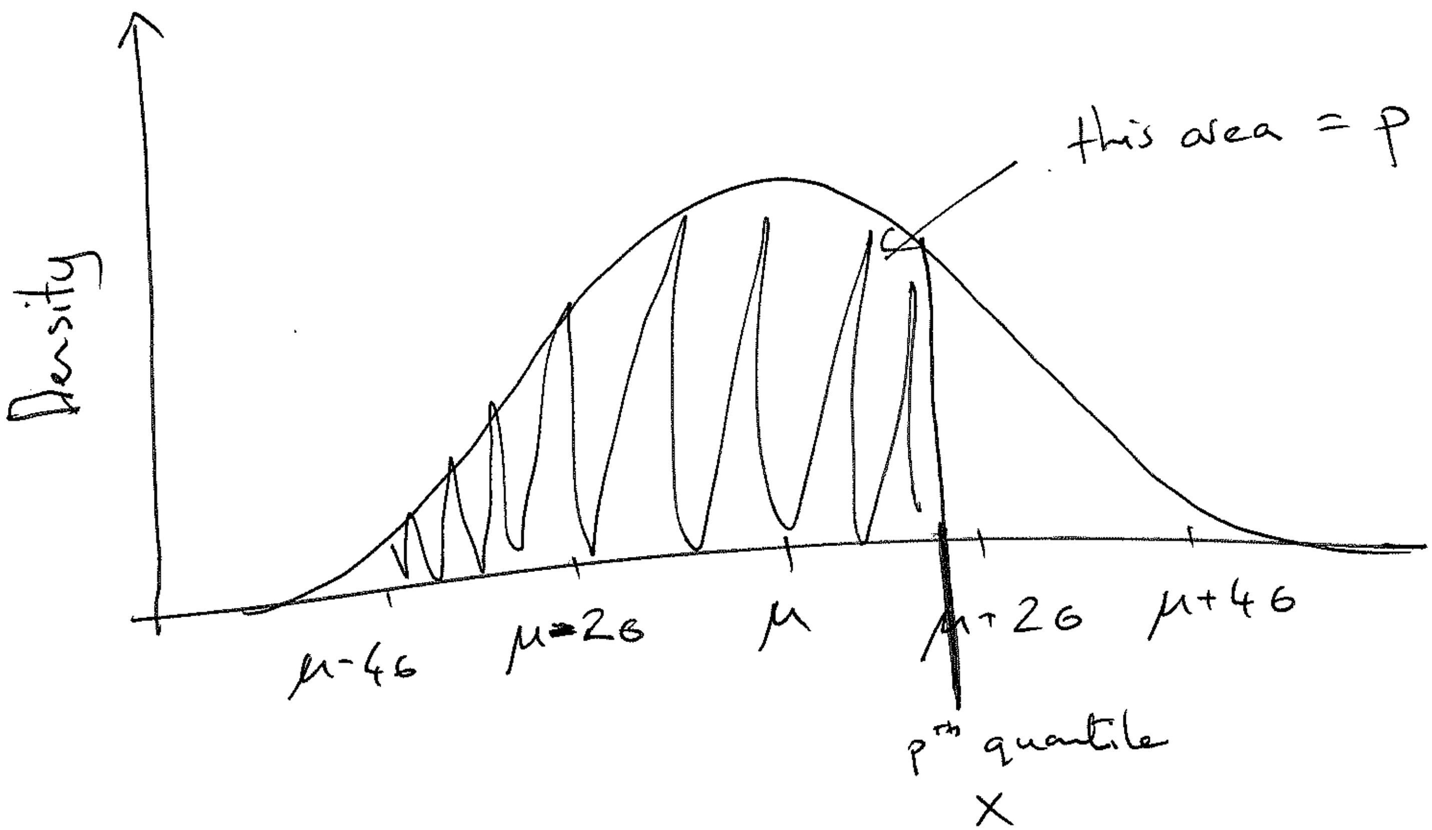


In R: `pnorm(q, mean = mu, sd = sigma)`

Area to ~~left?~~ right? $1 - \text{pnorm}(q, \text{mean}, \text{sd})$

~~Qua~~ Quantiles

F^{-1}



Exercise: Find probability

Return to example: $\mu = 20$ $\sigma^2 = 4$, $n = 16$

What is $P(\overline{Y} < 20.5)$**? What is** $P(\overline{Y} < 21)$

$\sqrt{\frac{4}{16}}$

pnorm(20.5, mean = 20, sd = $\frac{1}{2}$)

= 0.8413

$P(\bar{Y} < 21) = 0.977$

Exercise: Find sample size for specific variance in sample mean

Similar setup

$\mu = 20, \sigma^2 = 4.$ Sample n iid Sample mean

What should n be so $Var(\overline{Y}) = 0.5$?

$$Var(\overline{Y}) = \frac{\sigma^2}{n}$$

$$= \frac{4}{n} = 0.5$$

$$\Rightarrow \frac{4}{0.5} = n$$

$$n = 8$$

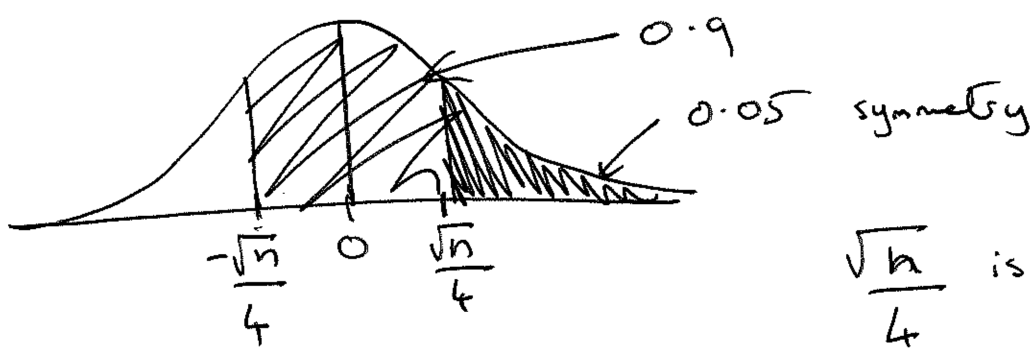
Exercise: Find sample size for an interval with desired probability

What should n be, so that $P(19.5 < \overline{Y} < 20.5) = 0.9$?

Transform to a statement about $Z \sim N(0, 1)$

$$P\left(\frac{19.5 - 20}{2/\sqrt{n}} < \frac{\overline{Y} - 20}{2\sqrt{n}} < \frac{20.5 - 20}{2\sqrt{n}} \right) = 0.9$$

$$P\left(-\sqrt{n}\,\frac{1}{4} < Z < \frac{\sqrt{n}}{4} \right) = 0.9$$



$\frac{\sqrt{n}}{4}$ is .95 ~~percentile~~ quantile

$\text{qnorm}(0.95) = 1.64$

$n = 43. \ldots$