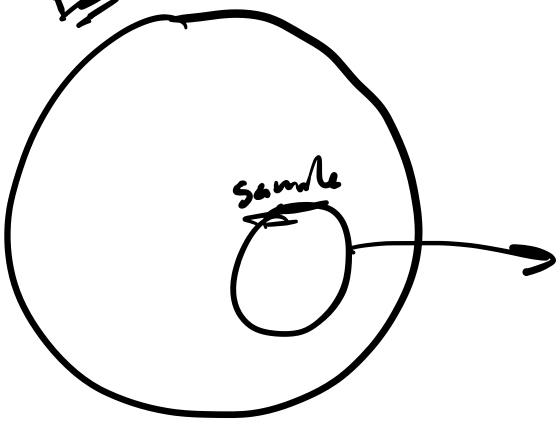
# STAT 515 fa 2023 Lec 12 slides

## Confidence interval for the mean when variance unknown


Karl B. Gregory

University of South Carolina


pop: variance $= \sigma^2$, mean $= \mu$ } unknown




$X_1, \ldots, X_n$

Recall: C.I. for $\mu$

$$\bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

we don't know this value!

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Can I do this: $\bar{X}_n \pm z_{\alpha/2} \frac{S_n}{\sqrt{n}}$ ??

must change this. Need to make it bigger.

**Recall**: If $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, then

$$\bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

is a $(1 - \alpha) \times 100\%$ CI for $\mu$.

*But what if we don't know $\sigma$?*

Using $\bar{X}_n \pm z_{\alpha/2} S_n / \sqrt{n}$ is okay if $n$ is large, but not if $n$ is small...

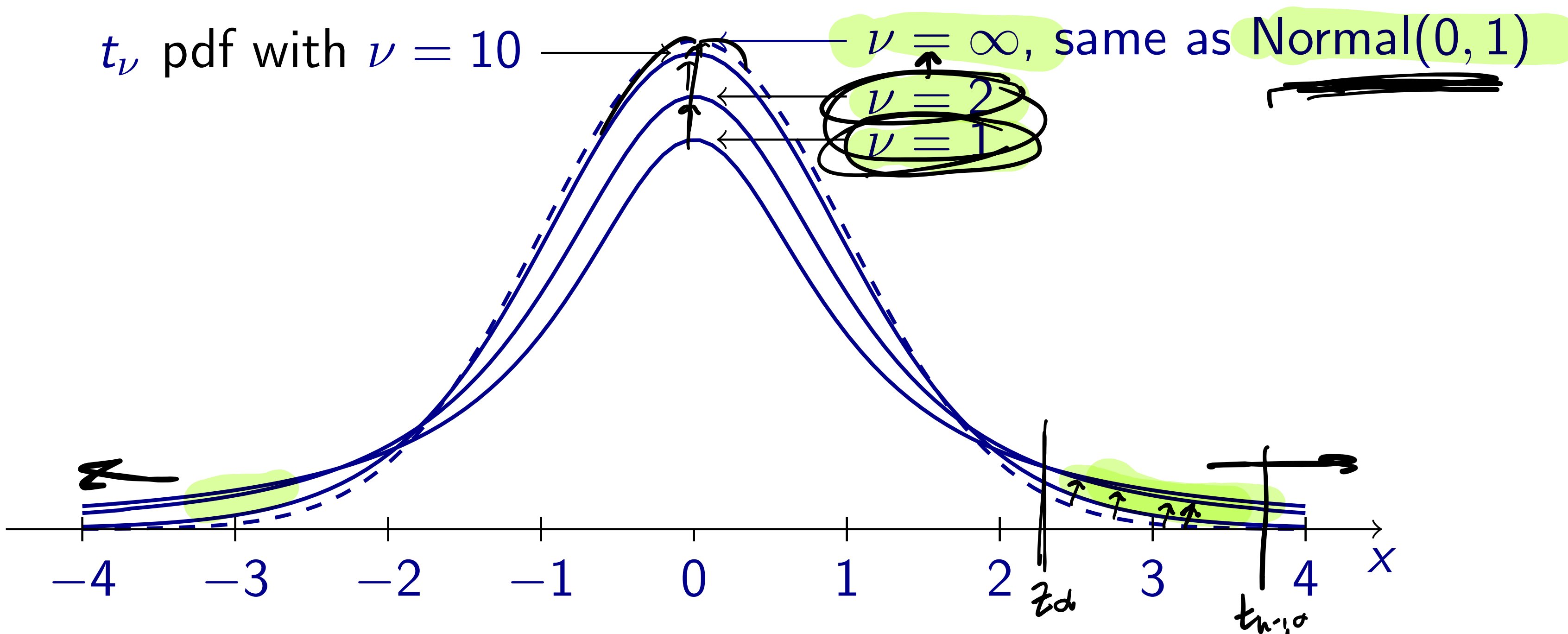
## The $t$-distributions

The probability distribution with pdf given by

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\nu\pi\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad \text{where } \Gamma(z) = \int_0^\infty u^{z-1}e^{-u}du,$$

for $\nu > 0$ is called the *$t$-distribution* with *degrees of freedom $\nu$*.

We write $T \sim t_\nu$ if a rv $T$ has this distribution.



$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad T = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1}$$

## Sampling distribution of studentized mean

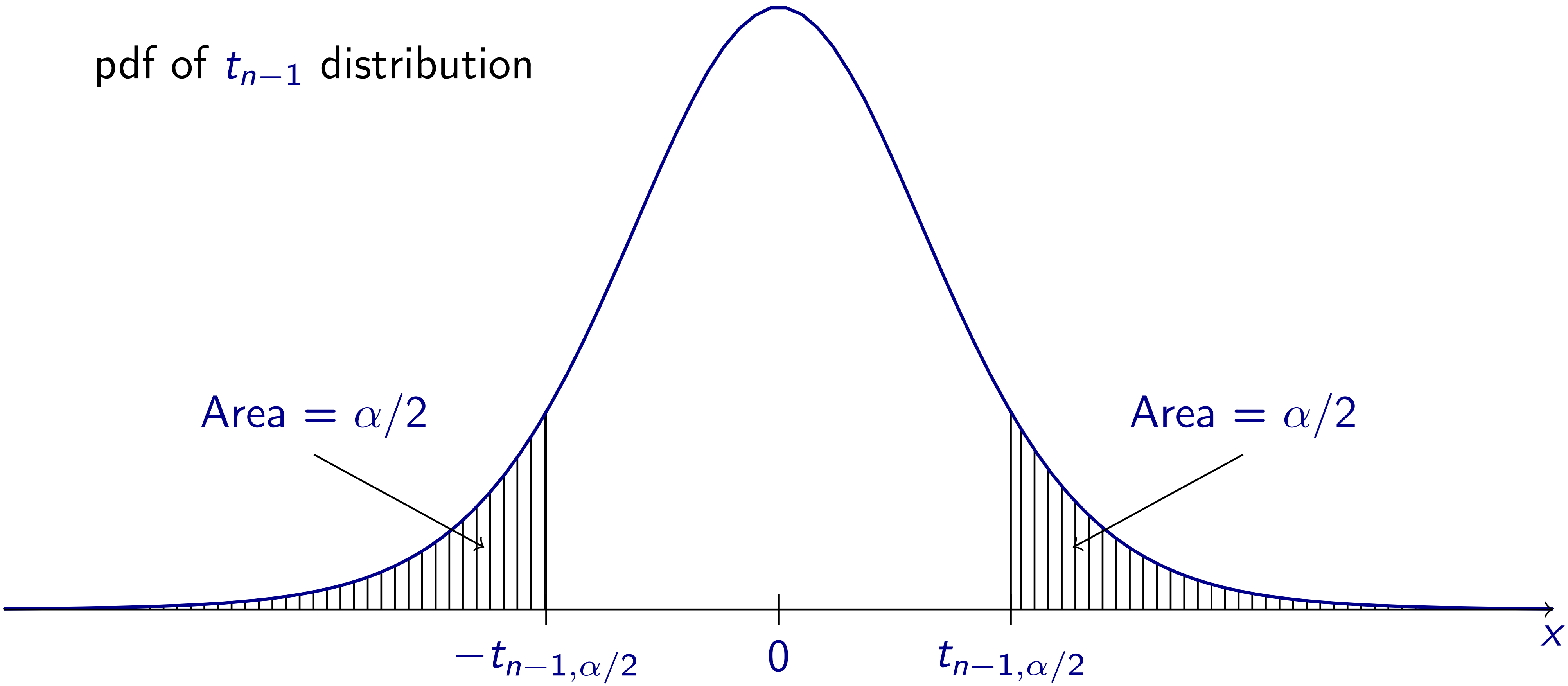
W. Gosset.

If $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, then

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}.$$

**Exercise:** Show above using this: If $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$ are ind. then

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu.$$




Can use function `qt()` or a t-table to look up the values, e.g.

$$t_{19,0.025} = \texttt{qt(.975,19)} = 2.093024$$
$$t_{19,0.005} = \texttt{qt(.995,19)} = 2.860935.$$

## Confidence interval for mean of a Normal population with $\sigma$ unknown

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$. Then a $(1-\alpha) \times 100\%$ CI for $\mu$ is

$$\bar{X}_n \pm t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}}.$$

*Show where this CI comes from.*

Build C.I.s for ratio $\frac{B}{A}$ (finger data).

Does C.I. contain 1.618? (1.618000)

$\bar{X}_n \pm t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}}$

95%   $\alpha = 0.05$, $n = 27$ → $t_{27-1,\,.025} = 2.0555$

$\bar{X}_n = 1.565$ estimate of $\mu$

$S_n = 0.148$ estimate of $\sigma$

95% : $1.565 \pm 2.0555 \frac{0.148}{\sqrt{27}} = (1.506, 1.623)$

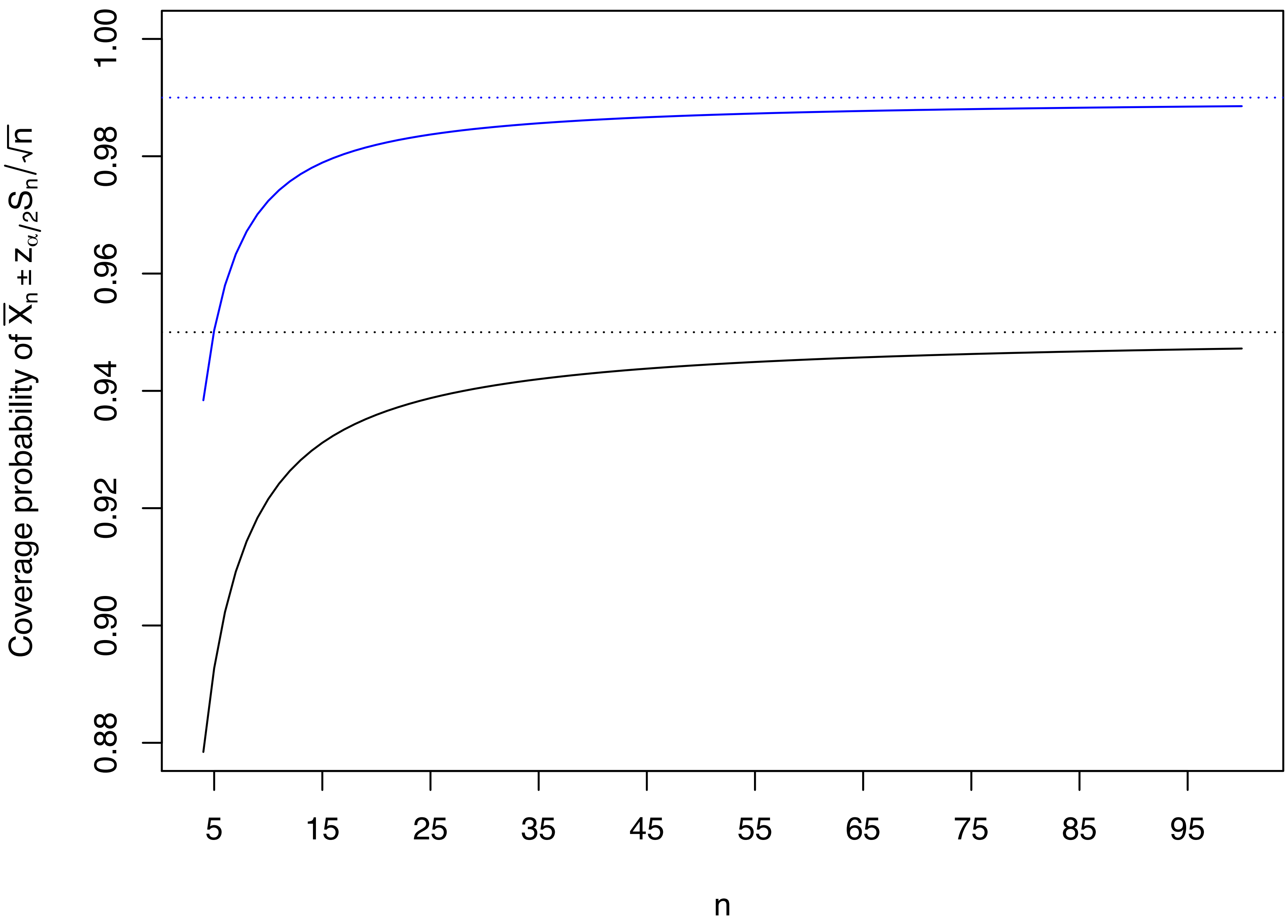
90% $= 1.565 \pm 1.7056 \frac{0.148}{\sqrt{27}} = (1.516, 1.613)$

$\alpha = 0.10$   $t_{27-1,\,0.05}$

**Exercise:** These are the commute times (sec) to class of a sample of students.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1832 | 1440 | 1620 | 1362 | 577 | 934 | 928 | 998 | 1062 | 900 |
| 1380 | 913 | 654 | 878 | 172 | 773 | 1171 | 1574 | 900 | 900 |

1. Make a Q-Q plot to check Normality of the population.
2. Construct a 95% confidence interval for the mean commute time of all students.
3. Construct a 99% confidence interval for the mean commute time of all students.
4. What if the intervals $\bar{X}_n \pm z_{\alpha/2} \cdot S_n/\sqrt{n}$ are used? How are they different?



Coverage probability of $\overline{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$
1.00
0.98
0.96
0.94
0.92
0.90
0.88
5
15
25
35
45
55
65
75
85
95
n

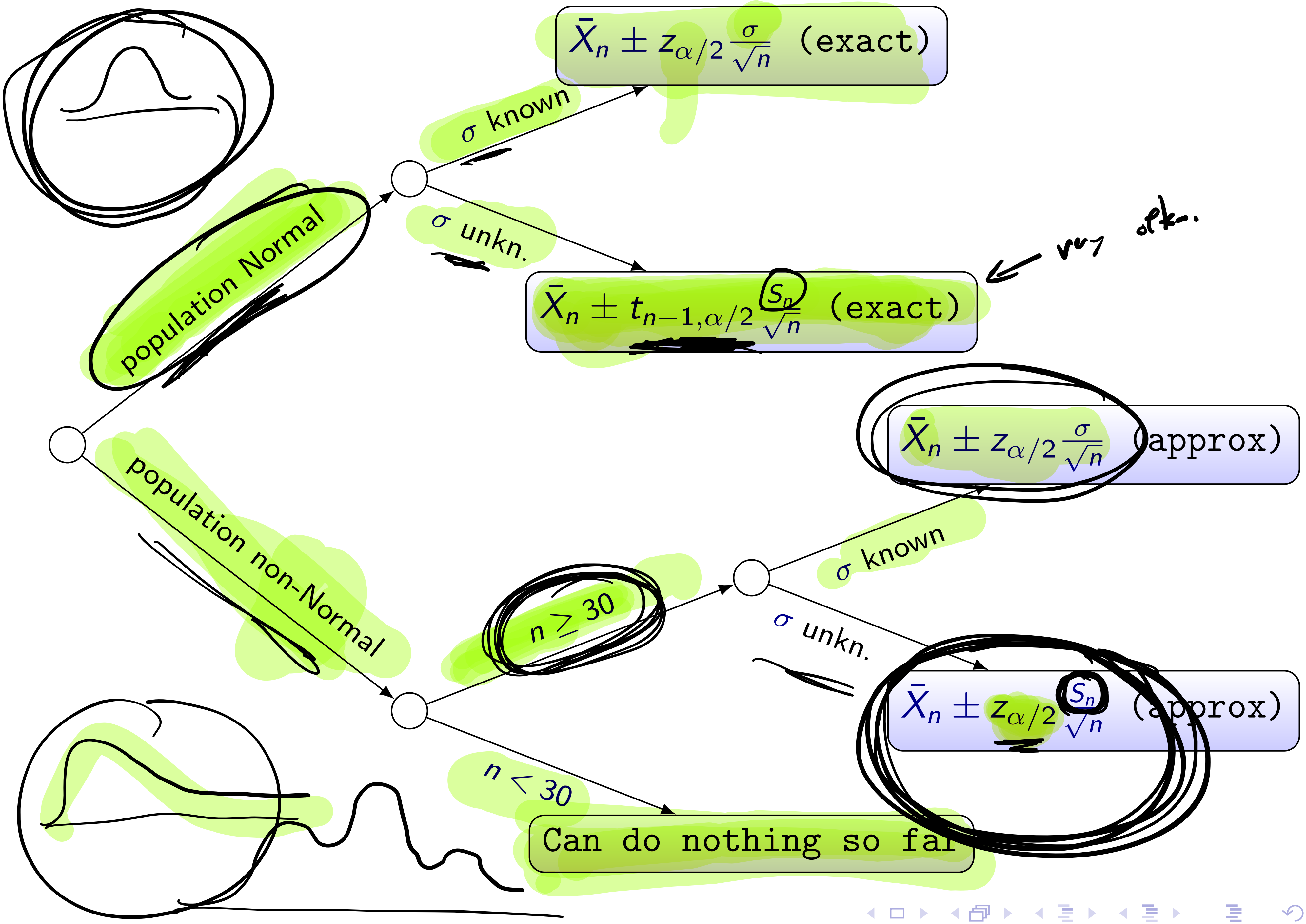
## CI for mean of non-Normal population with $\sigma$ unknown

Let $X_1, \dots, X_n$ be a rs from a pop. with mean $\mu$, and with $\mu_4 < \infty$, then

$$\bar{X}_n \pm z_{\alpha/2} \cdot \frac{S_n}{\sqrt{n}}$$

is an approximate $(1-\alpha) \times 100\%$ CI for $\mu$ when $n$ is large ($\geq 30$, say).

In the above $\mu_4 = \mathbb{E}|X_1|^4$. This limits the heavy-tailedness of the population.


population Normal
σ known
$\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (exact)
σ unkn.
$\bar{X}_n \pm t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}$ (exact)
population non-Normal
$n \geq 30$
σ known
$\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (approx)
σ unkn.
$\bar{X}_n \pm z_{\alpha/2}\frac{S_n}{\sqrt{n}}$ (approx)
$n < 30$
Can do nothing so far