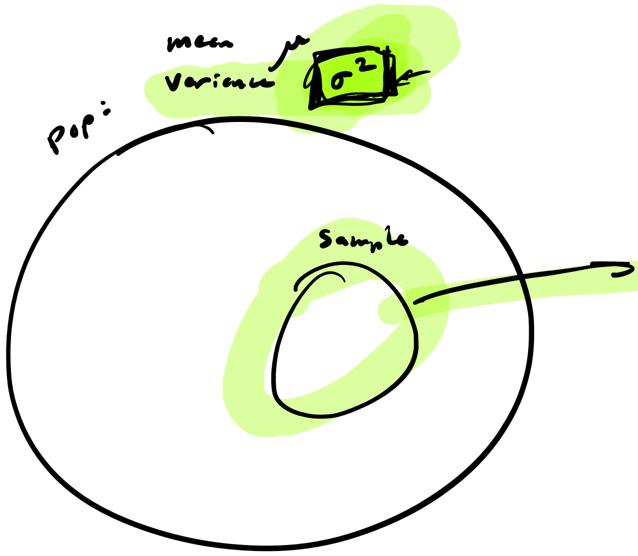
STAT 515 fa 2023 Lec 11 slides

Variance estimation

Karl B. Gregory

University of South Carolina

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.



$x_1, \dots, x_n$

$\bar{x}_n = \frac{x_1 + \dots + x_n}{n}$

sample variance: $\boxed{S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$

↑

Best guess of the value of σ^2.

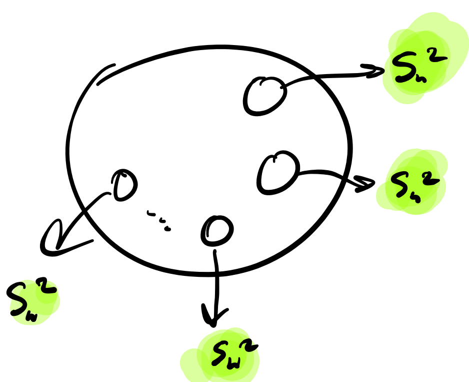
$S_n^2 = \text{var}(\text{data})$

sample std. dev $\quad S_n = \text{sd}(\text{data})$

for μ: $\quad \bar{x}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

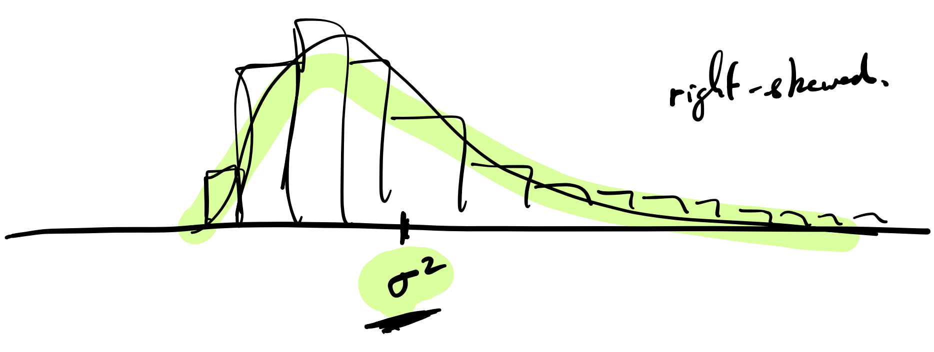
for p: $\quad \hat{p}_n \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$

Goal: Build a C.I. for σ^2.



$$\frac{(n-1) S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

chi-squared distribution.



right-skewed.

The chi-squared distributions

The probability distribution with pdf given by

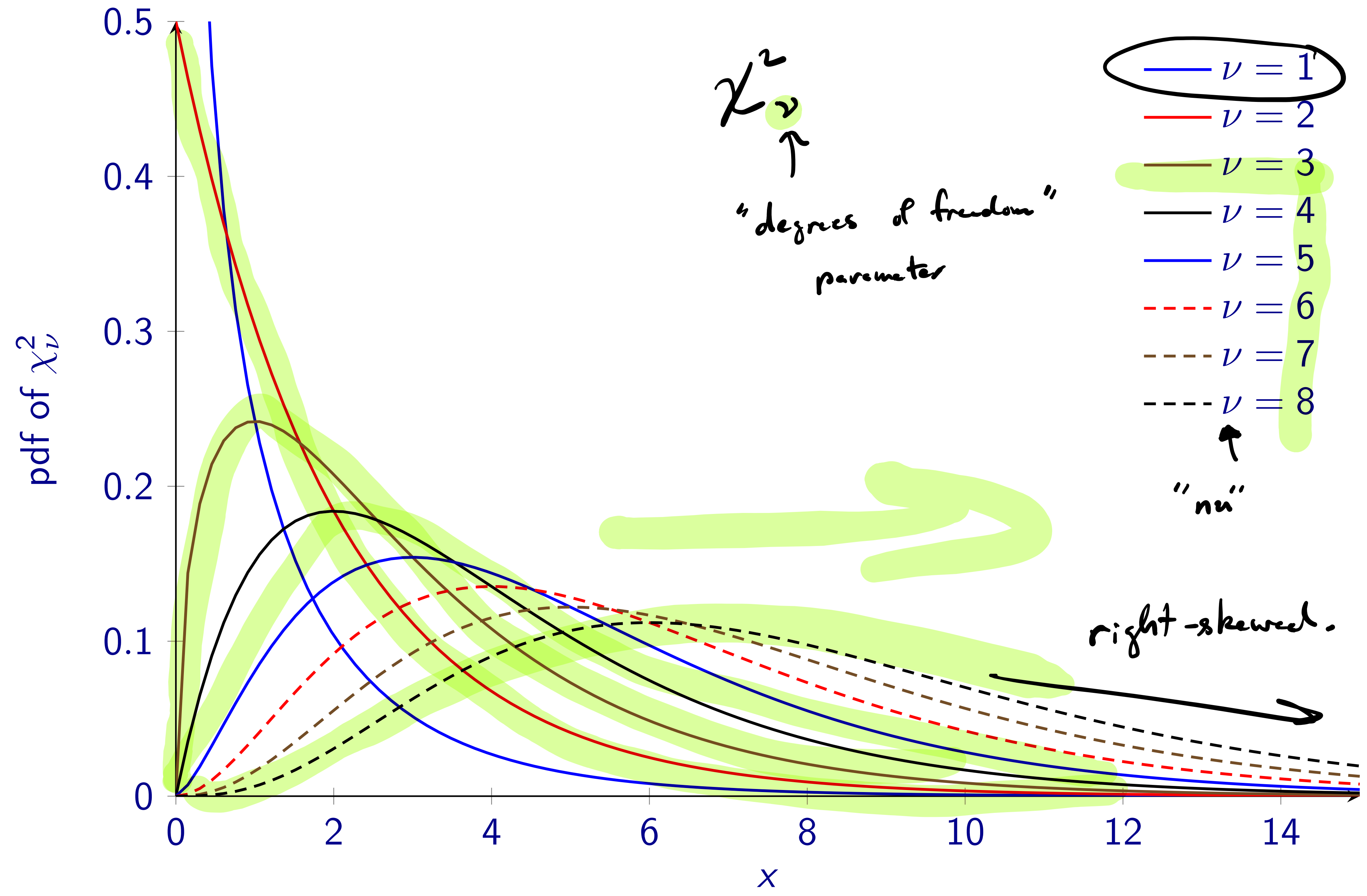
$$f(x) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} \exp\left(-\frac{x}{2}\right), \quad x > 0,$$

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

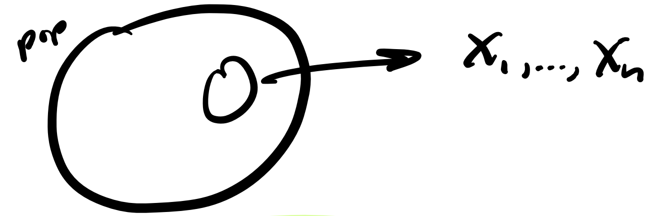
For a rv W with this distribution, we write $W \sim \chi^2_\nu$.

How to build a χ^2_k random variable

If $W = Z_1^2 + \cdots + Z_k^2$, where $Z_1, \ldots, Z_k \overset{\text{ind}}{\sim} \text{Normal}(0, 1)$, then $W \sim \chi^2_k$.



pdf of χ^2_ν
0.5
0.4
0.3
0.2
0.1
0
2
4
6
8
10
12
14
x
$\nu = 1$
$\nu = 2$
$\nu = 3$
$\nu = 4$
$\nu = 5$
$\nu = 6$
$\nu = 7$
$\nu = 8$
χ^2_ν
"degrees of freedom" parameter
"nu"
right-skewed.



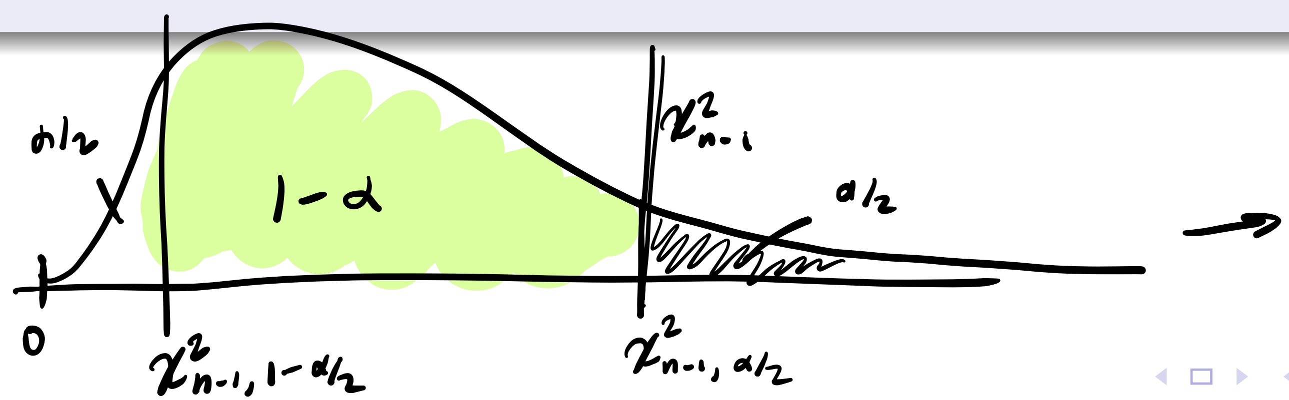
Recall: For a rs $X_1, \ldots, X_n$ the *sample variance* is the quantity

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

Sampling distribution of the sample variance

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

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}.$$



$$P\left(\chi^2_{n-1,\,1-\alpha/2} < \frac{(n-1)S_n^2}{\sigma^2} < \chi^2_{n-1,\,\alpha/2} \right) = 1-\alpha$$

rearrange to get σ^2 alone in the middle.

$a < b$

$\frac{1}{a} > \frac{1}{b}$

$$\hookrightarrow P\left(\frac{1}{\chi^2_{n-1,\,1-\alpha/2}} > \frac{\sigma^2}{(n-1)S_n^2} > \frac{1}{\chi^2_{n-1,\,\alpha/2}} \right) = 1-\alpha$$

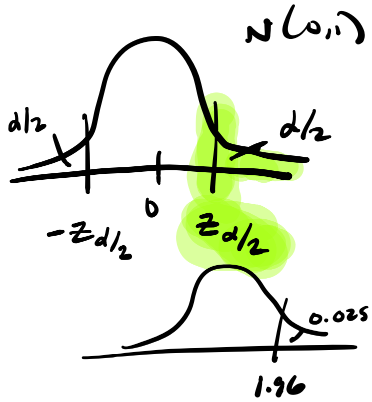
$$\hookrightarrow P\left(\frac{(n-1)S_n^2}{\chi^2_{n-1,\,1-\alpha/2}} > \sigma^2 > \frac{(n-1)S_n^2}{\chi^2_{n-1,\,\alpha/2}} \right) = 1-\alpha$$

A $(1-\alpha)100\%$ C.I. for σ^2 is

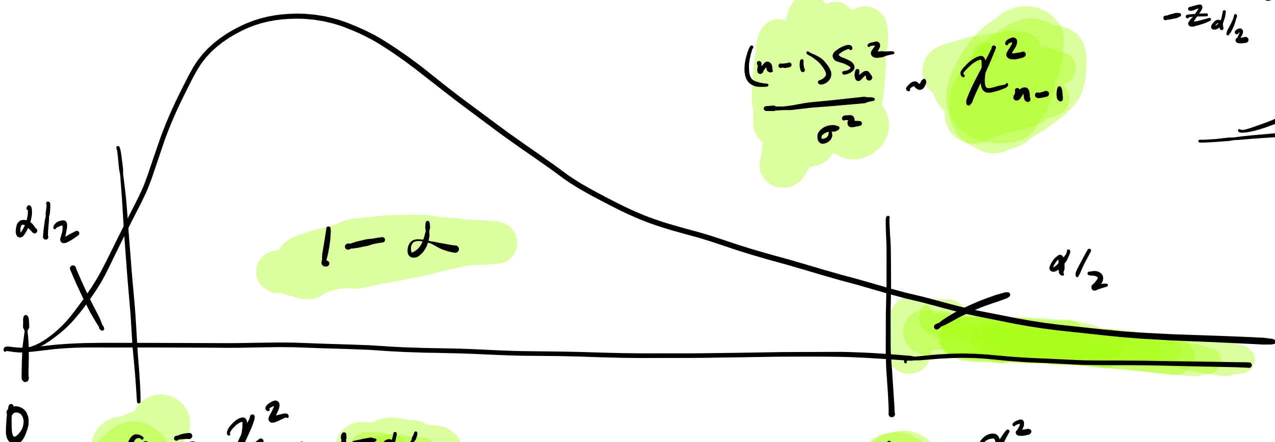
$$\left(\frac{(n-1)S_n^2}{\chi^2_{n-1,\,\alpha/2}} , \frac{(n-1)S_n^2}{\chi^2_{n-1,\,1-\alpha/2}} \right).$$

Want to build a $(1-\alpha)100\%$ C.I. for σ^2.

Draw sample of size n. Compute S_n^2.



$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$



$a = \chi^2_{n-1,\,1-\alpha/2}$ $\qquad$ $b = \chi^2_{n-1,\,\alpha/2}$

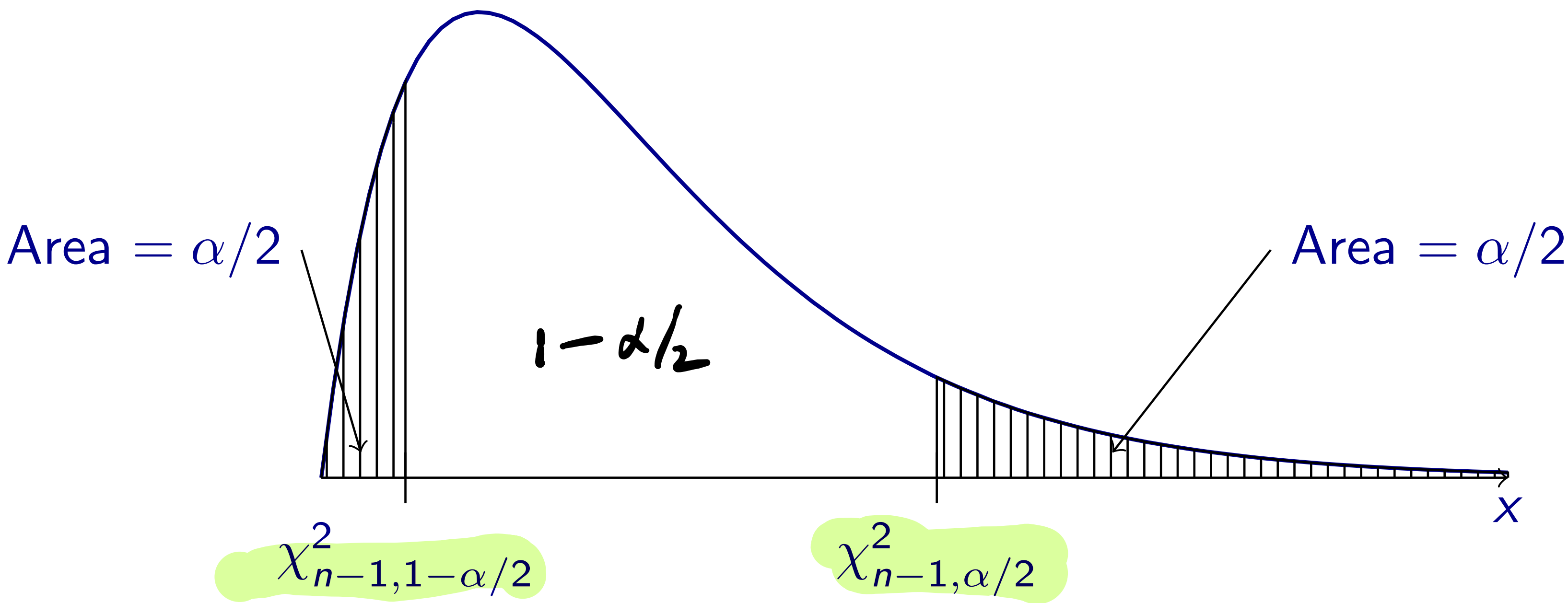
$$P\left(a < \frac{(n-1)S_n^2}{\sigma^2} < b \right) = 1-\alpha$$

$$P\left(\frac{1}{a} > \frac{\sigma^2}{(n-1)S_n^2} > \frac{1}{b} \right) = 1-\alpha$$

$$P\left(\underbrace{\frac{(n-1)S_n^2}{a}}_{U} > \sigma^2 > \underbrace{\frac{(n-1)S_n^2}{b}}_{L} \right) = 1-\alpha$$

A $(1-\alpha)100\%$ C.I. for σ^2 is $\left(\frac{(n-1)S_n^2}{b}, \frac{(n-1)S_n^2}{a} \right)$.

What about a general $(1-\alpha) \times 100\%$ CI for any $\alpha \in (0,1)$?



Example: For $\alpha = 0.05$, $n = 9$, can get

$$\chi^2_{9-1,0.975} = \texttt{qchisq(0.025,9-1)} = 2.179731$$
$$\chi^2_{9-1,0.025} = \texttt{qchisq(0.975,9-1)} = 17.53455$$

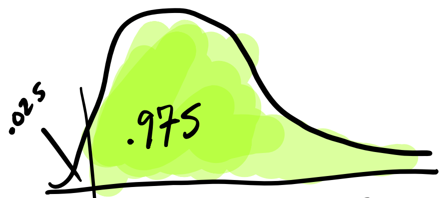
Can also use chi-square-table.

Confidence interval for variance of a Normal population

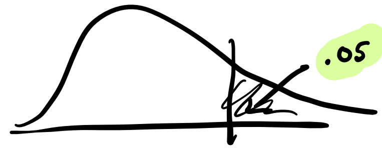
Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$. Then

$$\left(\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$

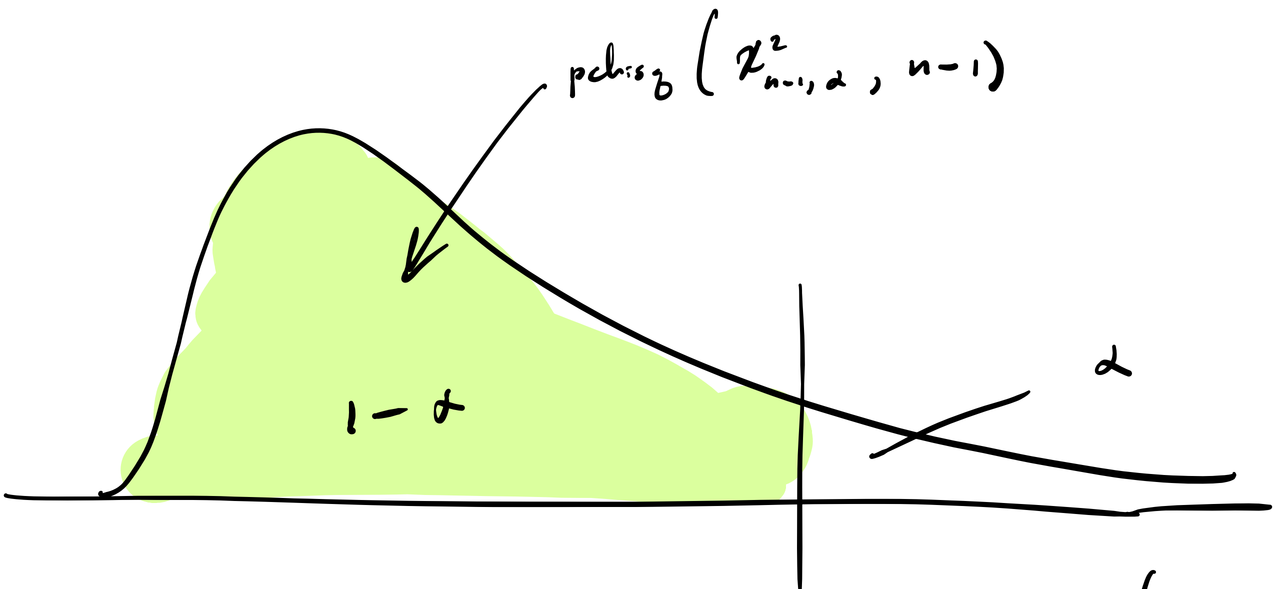
is a $(1-\alpha) \times 100\%$ confidence interval for σ^2.



$\chi^2_{11,0.975} = 3.8157$



$\chi^2_{3,0.05} = 7.8147$



$\chi^2_{n-1,\alpha} = \text{qchisq}(1-\alpha, n-1)$

Example:

$\chi^2_{14-1,0.025} = \text{qchisq}(.975, 14-1) = 24.7356$

$\chi^2_{14-1,0.975} = \text{qchisq}(.025, 14-1) = 5.00875$

Exercise: Run `data(Loblolly)` in R and consider the heights of 10-yr-old trees.

1. Make a Normal Q-Q plot of the data.
2. Build a 95% CI for the variance of the tree heights.
3. Build a 99% CI for the variance of the tree heights.

$$\left(\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$

(2) $n = 14$, $S_n^2 = 2.365$, $\alpha = 0.05$

95% C.I. is $\left(\frac{(14-1)\,2.365}{\chi^2_{14-1,.025}}, \frac{(14-1)\,2.365}{\chi^2_{14-1,.975}} \right)$

$\chi^2_{14-1,\,0.975} = 5.0088$

$\chi^2_{14-1,\,0.025} = 24.7356$

$$= \left(\frac{(14-1)\,2.365}{24.7356},\ \frac{(14-1)\,2.365}{5.0088} \right)$$

$$= (1.24,\ 6.14)$$

(2) $n = 14$, $s_n^2 = 2.365$, $\alpha = 0.01$

99% C.I. is $\left(\frac{(14-1)\,2.365}{\chi^2_{14-1,\,.005}},\ \frac{(14-1)\,2.365}{\chi^2_{14-1,\,.995}} \right)$

$\chi^2_{14-1,\,.005} = 29.8195$

$\chi^2_{14-1,\,.995} = 3.5650$