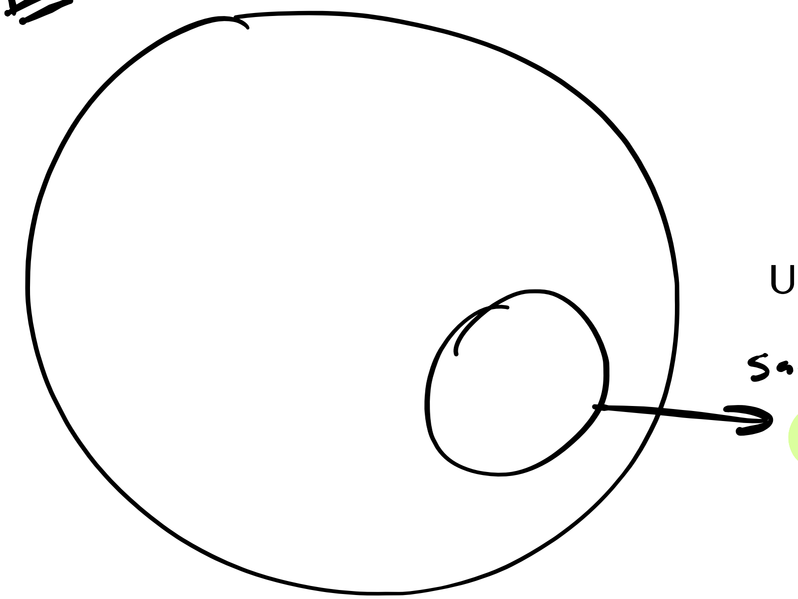
STAT 515 Lec 11 slides

Variance estimation

pop: μ mean
σ^2 variance



Karl Gregory

University of South Carolina

sample of size n

$\bar{X}_n$ = sample mean

S_n^2 = sample variance

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.

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

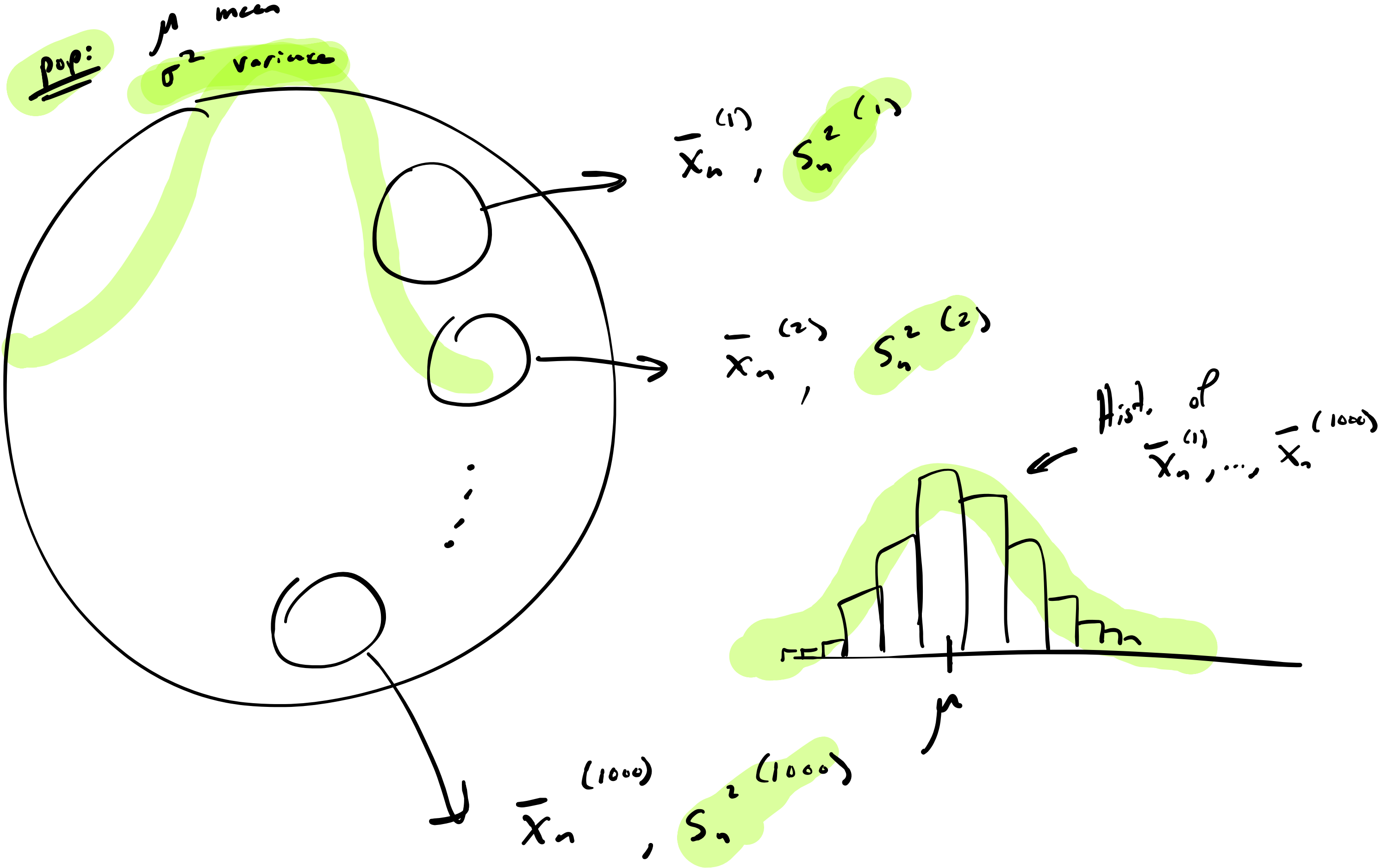
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

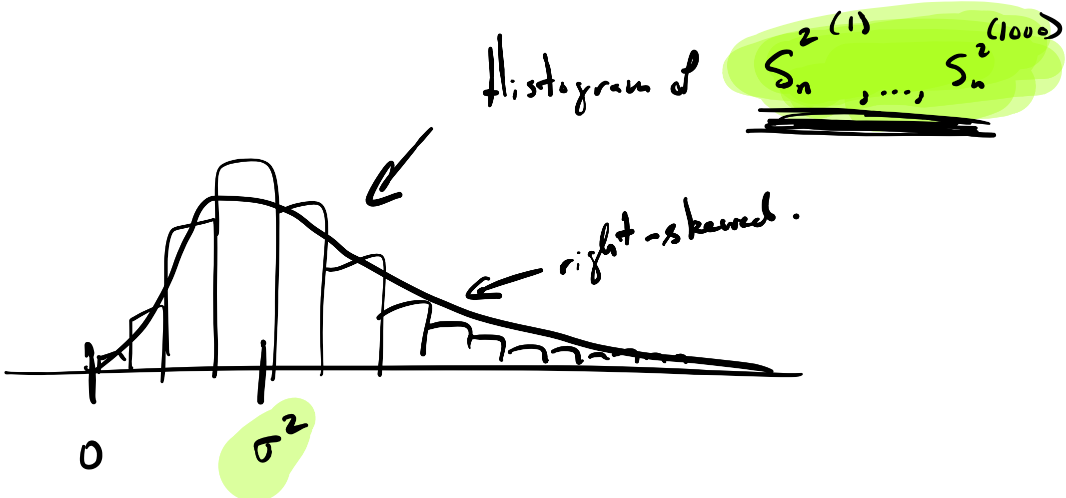
deviations

n=14

	X	Xbar	X - Xbar	(X - Xbar)^2
1	28.72	27.442	1.278	1.633
2	29.07	27.442	1.628	2.65
3	30.21	27.442	2.768	7.661
4	25.66	27.442	-1.782	3.176
5	28.66	27.442	1.218	1.483
6	25.99	27.442	-1.452	2.109
7	27.16	27.442	-0.282	0.08
8	27.9	27.442	0.458	0.21
9	25.45	27.442	-1.992	3.969
10	28.97	27.442	1.528	2.334
11	27.93	27.442	0.488	0.238
12	26.54	27.442	-0.902	0.814
13	26.08	27.442	-1.362	1.855
14	25.85	27.442	-1.592	2.535

sum





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

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

χ : "chi"

Sampling distribution of the sample variance

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

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

In the above χ^2_{n-1} denotes the chi-squared dist. with $n-1$ degrees of freedom...

The chi-squared distributions

The probability distribution with pdf given by

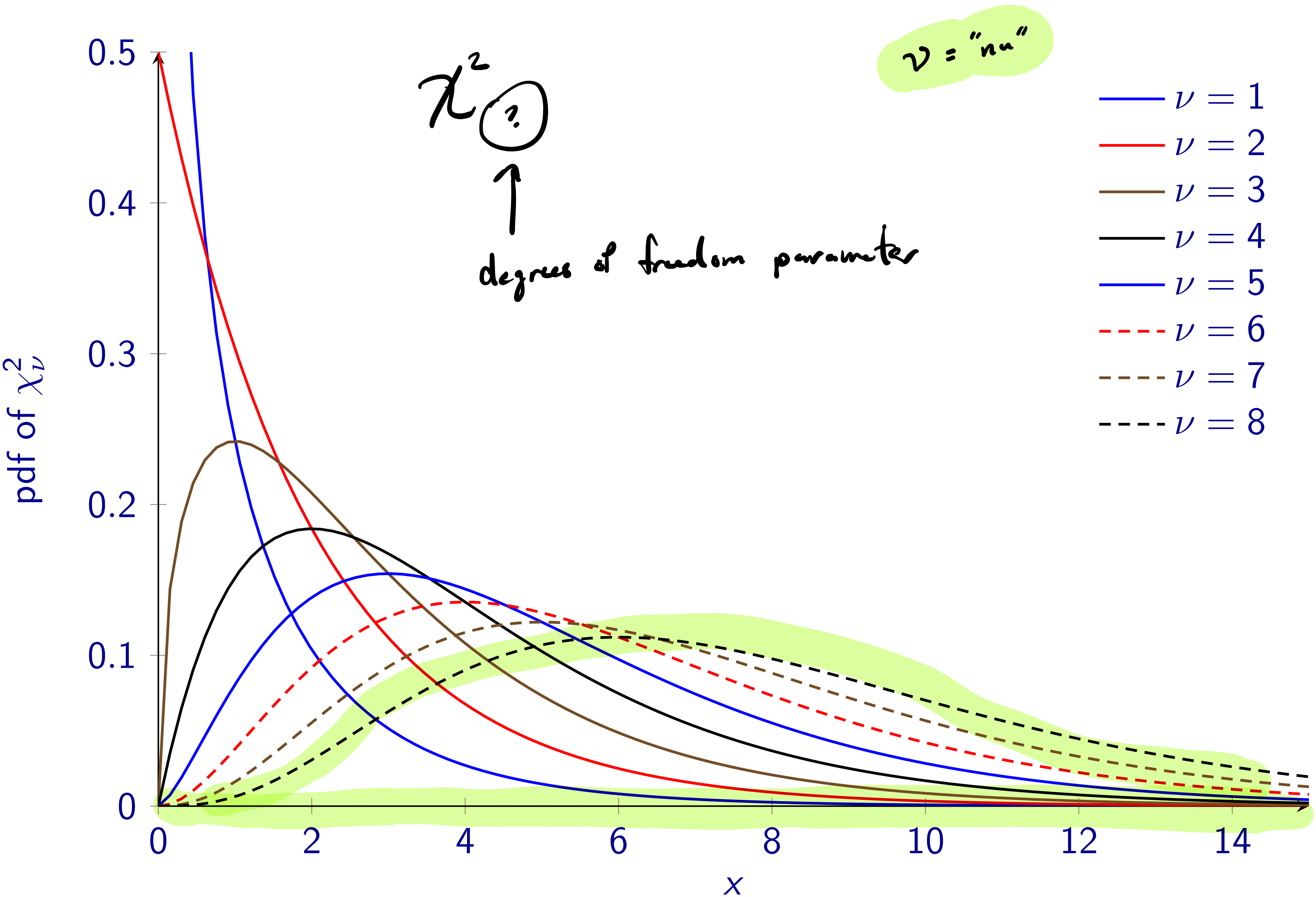
$$f(x) = \frac{1}{\underbrace{\Gamma(\nu/2)}_{\text{Gamma function.}}2^{\nu/2}} x^{\nu/2-1} \exp\left(-\frac{x}{2}\right), \quad x > 0,$$

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

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

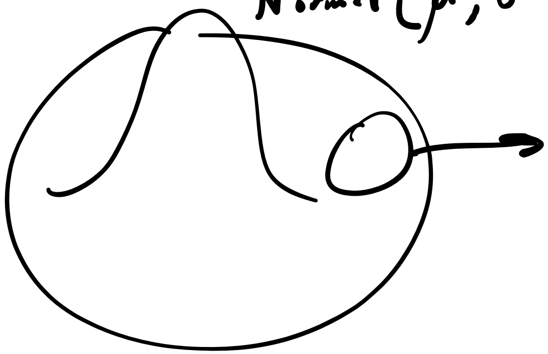
How to build a χ^2_k random variable

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



ν = "nu"
$\chi^2_{?}$
degrees of freedom parameter
$\nu = 1$
$\nu = 2$
$\nu = 3$
$\nu = 4$
$\nu = 5$
$\nu = 6$
$\nu = 7$
$\nu = 8$
0.5
0.4
0.3
0.2
0.1
0
pdf of χ^2_ν
0
2
4
6
8
10
12
14
x

$\text{Normal}(\mu, \sigma^2)$



$\bar{X}_n, S_n^2$

$n = 9$

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

$$\frac{(9-1)S_n^2}{\sigma^2} \sim \chi^2_{9-1}$$

Exercise: Let $X_1, \dots, X_9 \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$ and use the fact that

$$P\left(2.18 < \frac{(9-1)S_9^2}{\sigma^2} < 17.53\right) = 0.95$$

to build a 95% confidence interval for σ^2.

$$P\left(\frac{1}{2.18} > \frac{\sigma^2}{(9-1)S_9^2} > \frac{1}{17.53}\right) = 0.95$$

$$\Leftrightarrow \quad P\left(\frac{(9-1)S_9^2}{2.18} > \sigma^2 > \frac{(9-1)S_9^2}{17.53}\right) = 0.95$$

$\Leftrightarrow$

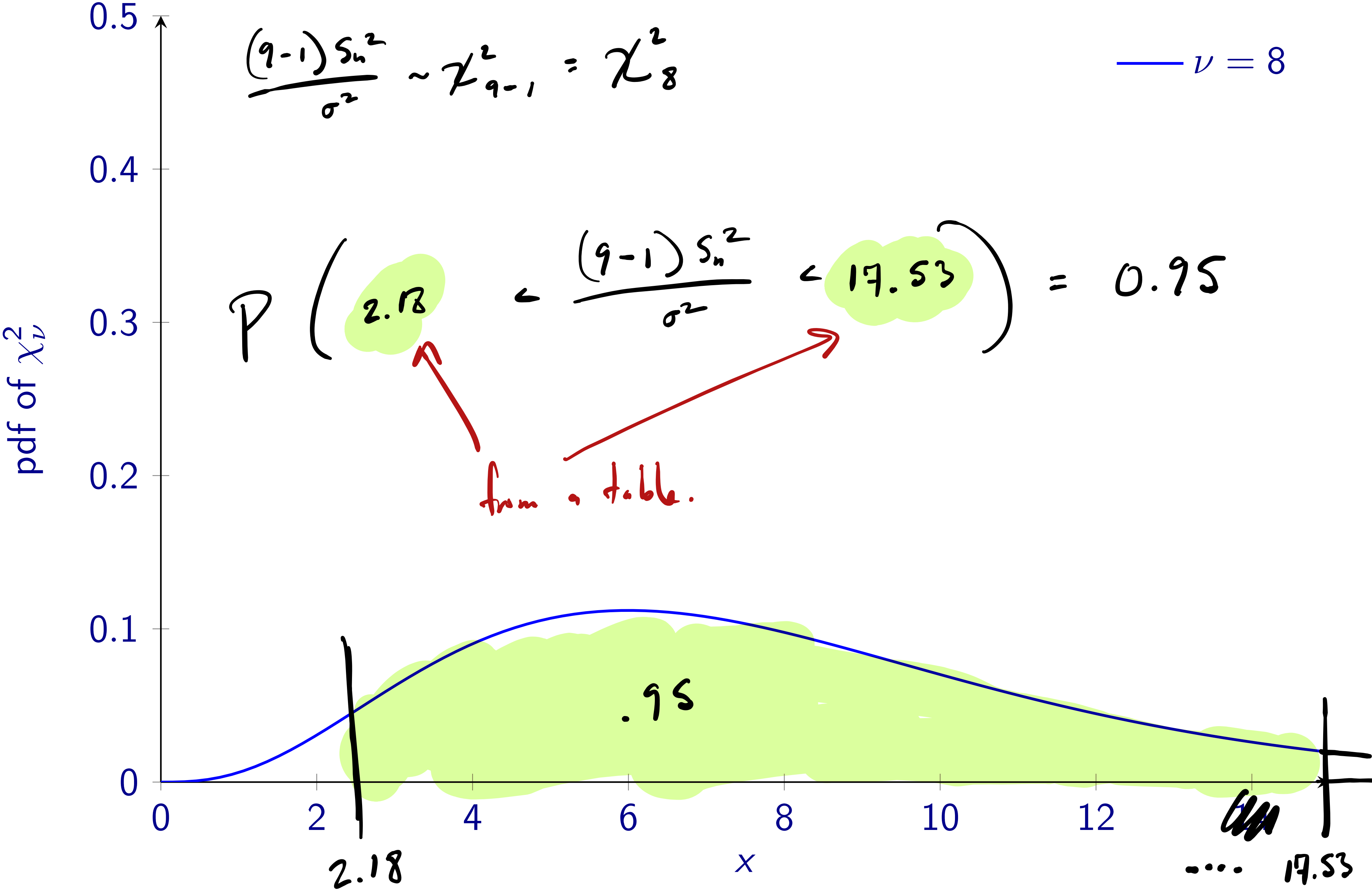
$$P\left(\frac{(9-1)S_n^2}{17.53} < \sigma^2 < \frac{(9-1)S_n^2}{2.18}\right) = 0.95$$

So a 95% C.I. for σ^2 is $\left(\frac{(9-1)S_n^2}{17.53}, \frac{(9-1)S_n^2}{2.18}\right)$.

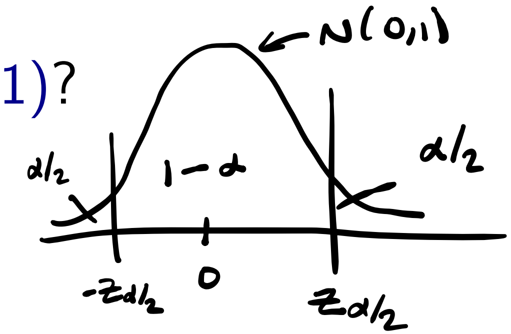
$$\frac{(9-1)S_n^2}{\sigma^2} \sim \chi^2_{9-1} = \chi^2_8$$

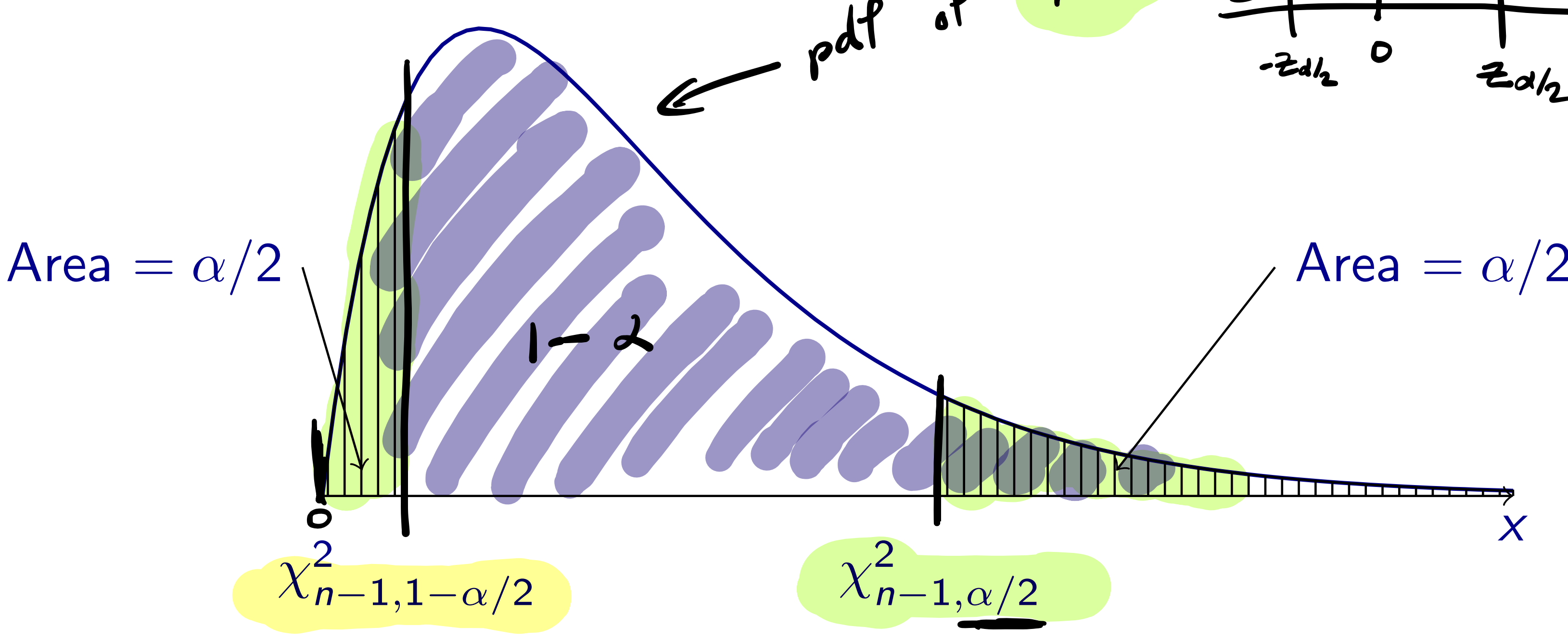
$$P\left(2.18 < \frac{(9-1)S_n^2}{\sigma^2} < 17.53\right) = 0.95$$

from a table.



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`.

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

Rearrange

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.

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

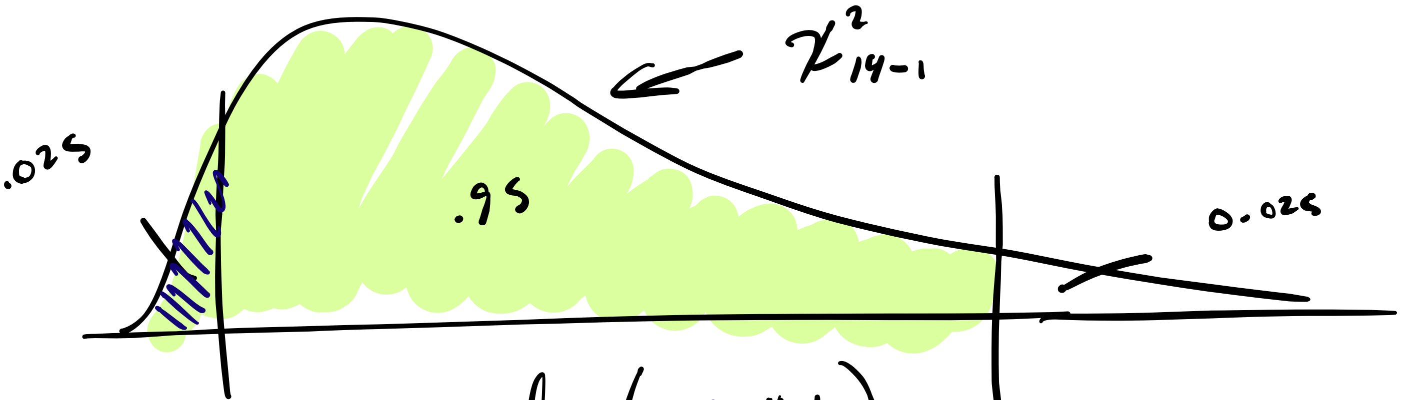
1. Make a Normal Q-Q plot of the data.
2. Build a 95% CI for the variance of the tree heights.] $\alpha = 0.05$
3. Build a 99% CI for the variance of the tree heights.

$n = 14$ $\qquad s_n = 1.54$

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

$$\left(\frac{(14-1)(1.54)^2}{24.74}, \frac{(14-1)(1.54)^2}{5.01}\right) = (1.25, 6.14)$$

$\chi^2_{14-1,0.025} = 24.74$ $\qquad \chi^2_{14-1,.975} = 5.01$



$\chi^2_{14-1,1-0.025}$ = qchisq(0.025, 14-1) = 5.01.

```
data("Loblolly")
X <- Loblolly$height[Loblolly$age == 10]
Sn <- sd(X)
n <- length(X)
alpha <- 0.05
lo <- (n-1)*Sn^2 / qchisq(1-alpha/2,n-1)
up <- (n-1)*Sn^2 / qchisq(alpha/2,n-1)
```

$\chi^2_{14-1,.025}$ = qchisq(0.975, 14-1) = 24.74

area under curve to the **LEFT.**

3 99%, $\alpha = 0.01$

$$\left(\frac{(14-1)(1.54)^2}{29.82}, \frac{(14-1)(1.54)^2}{3.57} \right) = (1.03, 8.62)$$

$\chi^2_{14-1,\ \underbrace{0.005}_{\alpha/2}} = 29.82$

$\chi^2_{14-1,\ \underbrace{1-0.005}_{.995}} = 3.57$

$1-\alpha/2$