Lecture 20 Confidence Interval For The Variance

BIO210 Biostatistics

Xi Chen

Fall, 2025

School of Life Sciences
Southern University of Science and Technology

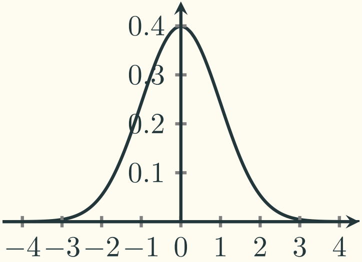
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Interval Estimation For The Variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ Not normally distributed!

Chi-squared Distribution

$$\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, ..., \boldsymbol{Z}_n \sim \mathcal{N}(0, 1)$$



$$\boldsymbol{U}_1 = \boldsymbol{Z}_1^2 \qquad \boldsymbol{U}_1 \sim \boldsymbol{\chi^2}(1)$$

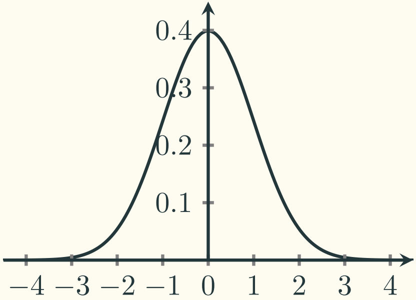
$$\boldsymbol{U}_2 = \boldsymbol{Z}_1^2 + \boldsymbol{Z}_2^2 \qquad \boldsymbol{U}_2 \sim \boldsymbol{\chi^2}(2)$$

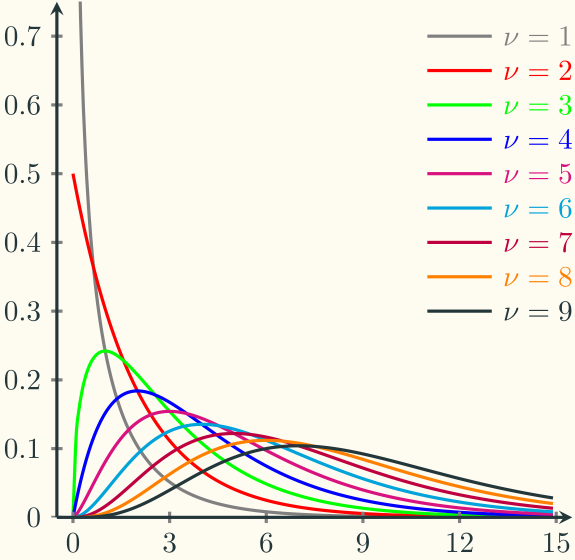
$$\boldsymbol{U}_3 = \boldsymbol{Z}_1^2 + \boldsymbol{Z}_2^2 + \boldsymbol{Z}_3^2 \qquad \boldsymbol{U}_3 \sim \boldsymbol{\chi^2}(3)$$

$$\vdots \qquad \vdots$$

$$\boldsymbol{U}_n = \boldsymbol{Z}_1^2 + \boldsymbol{Z}_2^2 + \boldsymbol{Z}_3^2 + \cdots \boldsymbol{Z}_n^2 \qquad \boldsymbol{U}_n \sim \boldsymbol{\chi^2}(n)$$

Chi-squared Distributions of Different ν





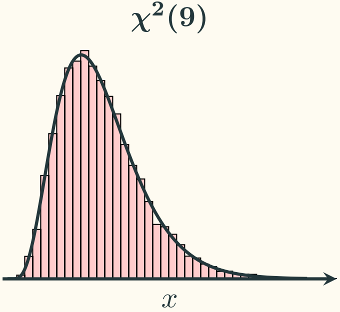
Sampling Distribution of The Sample Variance

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be a sample independently drawn from a normal distribution $\mathcal{N}(\mu, \sigma^2)$, and we define the sample mean and variance as:

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i\,, \quad \boldsymbol{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^2$$

Then we have:

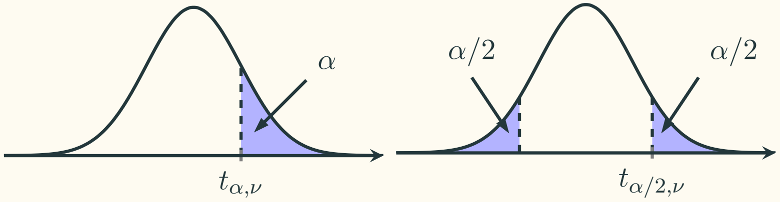
$$\frac{(n-1)\boldsymbol{S}^2}{\sigma^2} \sim \boldsymbol{\chi}^2(n-1)$$



Confidence Interval For The Mean

Recall: construct $(1-\alpha)\times 100\%$ confidence interval for the mean.

- One-sided (lower bound): $\mathbb{P}\left(\mu \geqslant \bar{X} - t_{\alpha,\nu}\cdot\frac{s}{\sqrt{n}}\right) = 1-\alpha$
- One-sided (upper bound): $\mathbb{P}\left(\mu \leqslant \bar{X} + t_{\alpha,\nu}\cdot\frac{s}{\sqrt{n}}\right) = 1-\alpha$
- Two-sided: $\mathbb{P}\left(\bar{X} - t_{\alpha/2,\nu}\cdot\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + t_{\alpha/2,\nu}\cdot\frac{s}{\sqrt{n}}\right) = 1-\alpha$



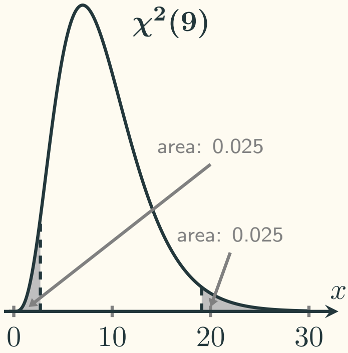
Goal: for a population with unknown variance σ^2, find a and b, such that $\mathbb{P}\left(a \leqslant \sigma^2 \leqslant b\right) = 0.95$

$$\mathbb{P}\left(\chi^2_{0.975,\nu} \leqslant \chi^2 \leqslant \chi^2_{0.025,\nu}\right) = 0.95$$

$$\mathbb{P}\left[\chi^2_{0.975,\nu} \leqslant \frac{(n-1)\boldsymbol{S}^2}{\sigma^2} \leqslant \chi^2_{0.025,\nu}\right] = 0.95$$

$$\mathbb{P}\left[\frac{1}{\chi^2_{0.025,\nu}} \leqslant \frac{\sigma^2}{(n-1)\boldsymbol{S}^2} \leqslant \frac{1}{\chi^2_{0.975,\nu}}\right] = 0.95$$

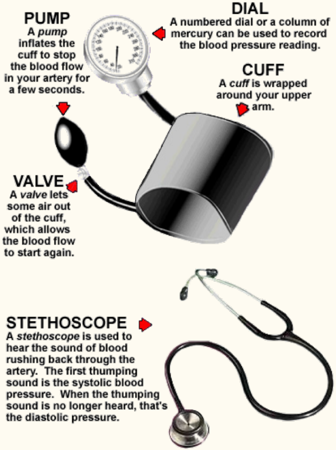
$$\mathbb{P}\left(\frac{n-1}{\chi^2_{0.025,\nu}} \cdot \boldsymbol{S}^2 \leqslant \sigma^2 \leqslant \frac{n-1}{\chi^2_{0.975,\nu}} \cdot \boldsymbol{S}^2\right) = 0.95$$



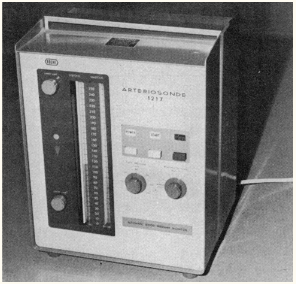
95% CI for σ^2 :

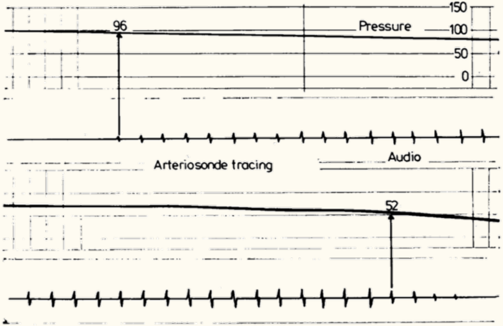
$$\left[\frac{n-1}{\chi^2_{0.025,\nu}} \cdot \boldsymbol{S}^2 , \frac{n-1}{\chi^2_{0.975,\nu}} \cdot \boldsymbol{S}^2\right]$$

Sphygmomanometer



Hypertension: An Arteriosonde machine "prints" blood-pressure readings on a tape so that the measurement can be read rather than heard. A major argument for using such a machine is that the variability of measurements obtained by different observers on the same person will be lower than with a standard blood-pressure cuff.





Confidence interval for variance

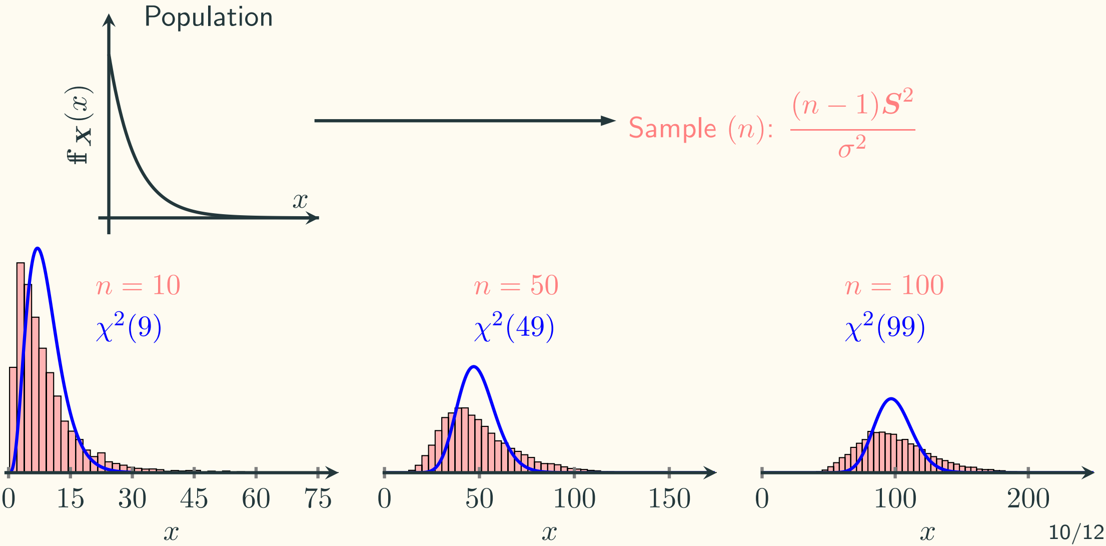
Persion (i)	Observer #1	Observer #2	Difference (d_i)
1	194	200	-6
2	126	123	3
3	130	128	2
4	98	101	-3
5	136	135	1
6	145	145	0
7	110	111	-1
8	108	107	1
9	102	99	3
10	126	128	-2

Does it make sense to construct a 95% CI for the variance?

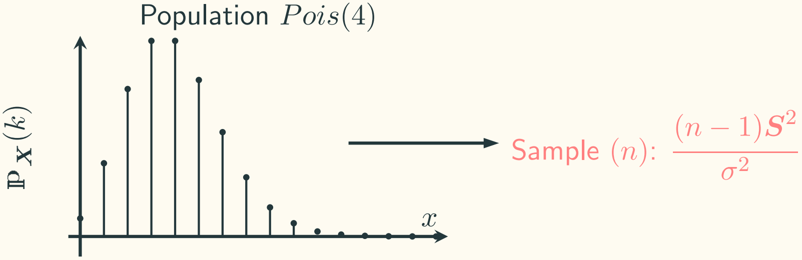
95% CI for σ^2 :

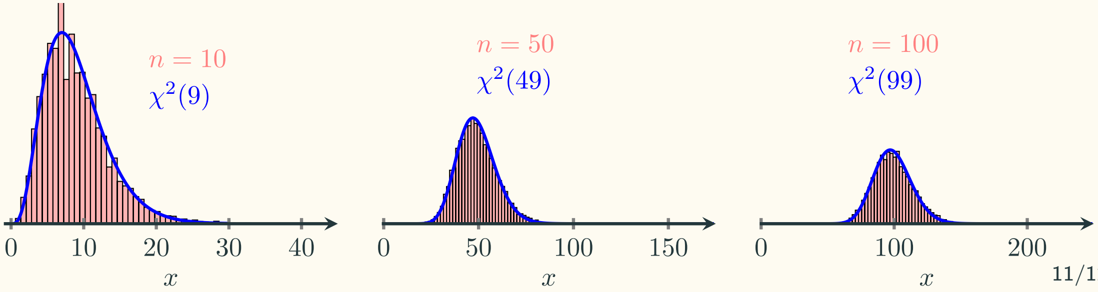
$$\left[\frac{n-1}{\chi^2_{0.025,\nu}} \cdot s^2 \,, \frac{n-1}{\chi^2_{0.975,\nu}} \cdot s^2\right]$$

Sampling Distribution of The Sample Variance



Sampling Distribution of The Sample Variance





Conditions For Valid Confidence Intervals For The Variance

1. Random Samples
2. Independence ($n < 10\%$ population size)
3. Original population distribution must be normal