

STAT 512 su 2021 Lec 07 slides

Large-sample pivot quantities and central limit theorem

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.

Recall:

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

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1), \quad \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}, \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$



What if X_1, \dots, X_n are not sampled from a Normal distribution?

Central limit theorem informally stated

If X_1, \dots, X_n are iid with mean μ and variance $\sigma^2 < \infty$, then

$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ behaves more and more like $Z \sim N(0, 1)$ for larger and larger n .

Same is true of $\frac{\bar{X}_n - \mu}{\hat{\sigma}_n/\sqrt{n}}$, where $\hat{\sigma}_n \xrightarrow{p} \sigma$.

Can use to build large- n CI for μ when population non-Normal.

Convergence in distribution

A sequence of rvs Y_1, Y_2, \dots with cdfs F_{Y_1}, F_{Y_2}, \dots *converges in distribution* to the rv $Y \sim F_Y$ if

$$\lim_{n \rightarrow \infty} F_{Y_n}(y) = F_Y(y),$$

for all $y \in \mathbb{R}$ at which $F_Y(y)$ is continuous.

Formalizes “behaves more and more like”.

Write $Y_n \xrightarrow{D} Y$.

Refer to dist. with cdf F_Y as the *asymptotic distribution* or *limiting distribution*.

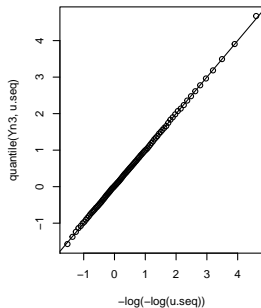
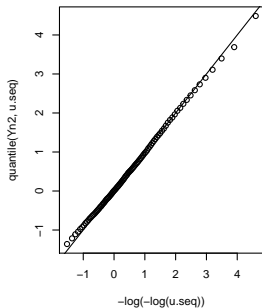
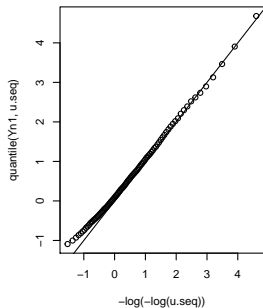
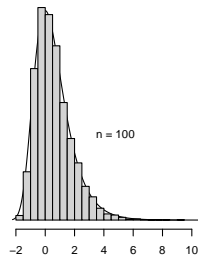
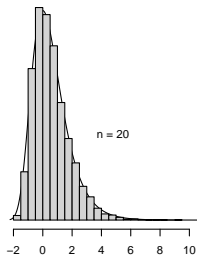
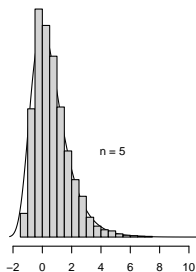
Exercise: For $n \geq 1$, let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda)$ and let

$$Y_n = (X_{(n)} - \lambda \log n) / \lambda \quad \text{and} \quad Y \sim F_Y(y) = e^{-e^{-y}} \text{ for } y \in \mathbb{R}.$$

Show that $Y_n \xrightarrow{D} Y$.



Y_n is an *asymptotic pivot quantity*: limiting dist. known and parameter-free.



Theorem (Central limit theorem)

If X_1, \dots, X_n are i.i.d. from a dist. with mgf defined in a neighborhood of zero, then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim \text{Normal}(0, 1) \text{ as } n \rightarrow \infty,$$

where μ and σ are the population mean and standard deviation.

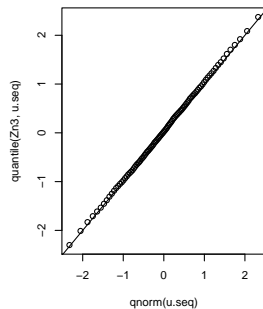
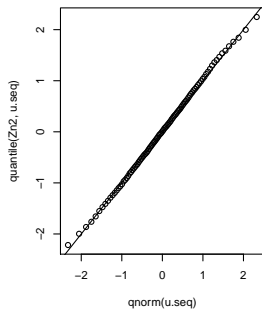
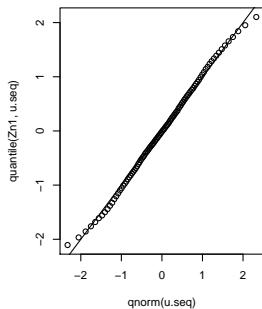
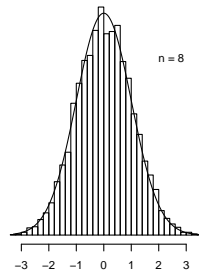
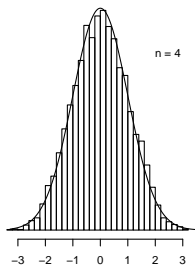
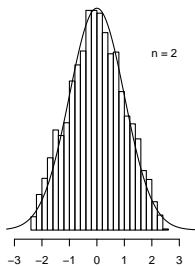


Exercise: Prove using mgfs.

Mgf assumption not needed. Can prove CLT requiring only $\sigma^2 < \infty$ (wow hard).

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Uniform}(a, b)$.

- 1 Give a function of \bar{X}_n which is asymptotically standard Normal.
- 2 Run some simulations to investigate the rate of convergence.



Theorem (Slutzky's theorem)

If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{p} a$, then

$$X_n + Y_n \xrightarrow{D} X + a.$$

$$X_n Y_n \xrightarrow{D} Xa$$

$$X_n/Y_n \xrightarrow{D} X/a, \quad \text{provided } a \neq 0.$$

Theorem (Corollary of Slutsky's theorem and central limit theorem)

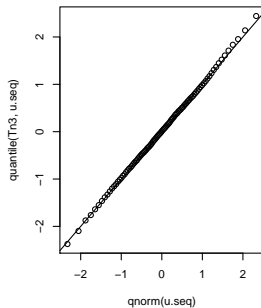
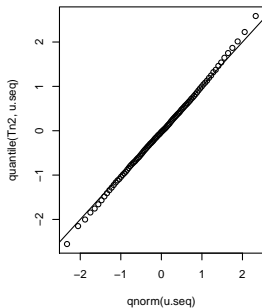
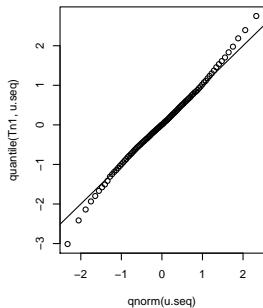
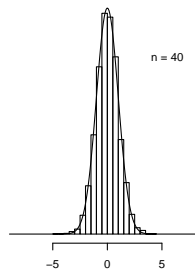
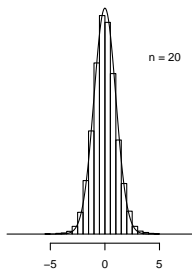
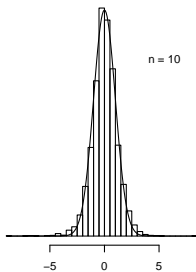
Let X_1, \dots, X_n be a rs from a dist. with mean μ and variance $\sigma^2 < \infty$. Then

$$\frac{\bar{X}_n - \mu}{\hat{\sigma}_n / \sqrt{n}} \xrightarrow{D} Z \sim \text{Normal}(0, 1) \text{ as } n \rightarrow \infty,$$

where $\hat{\sigma}_n \xrightarrow{P} \sigma$.

Exercise:

- 1 Show how the CLT and Slutsky's theorem imply the above.
- 2 Illustrate this result in simulation for $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Uniform}(a, b)$.



Exercises: Derive asymptotic confidence intervals for

- ① μ based on X_1, \dots, X_n iid with mean μ and finite 4th moment.
- ② p based on $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$.

Exercise: For $n = 10, 20, 40, 80, 160, 320$, simulate coverage of

$$\bar{X}_n \pm z_{0.025} \cdot S_n / \sqrt{n}$$

for mean of $\text{Gamma}(\alpha = 2, \beta = 4)$ distribution.

n	10	20	40	80	160	320
coverage	0.887	0.925	0.934	0.938	0.947	0.947

The *coverage* of a CI is the probability with which it contains its target.

