Asymptotics: Limiting results

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Today's outline

Asymptotics: Limiting results

Outline

Focus on the sample average

Definition (Weak Law of Large Numbers)

Let $\{X_n\}$ be a sequence of random variables with finite expected values $\mathrm{E}(X_n) = \mu_n$. We say that $\{X_n\}$ obeys the weak law of large numbers (WLLN), if

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) = \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

(May generalize with a more generic sequence b_n instead of n.)

For $\mathrm{E}(X_i) = \mu_i = \mu \ \forall i$ the WLLN would imply that

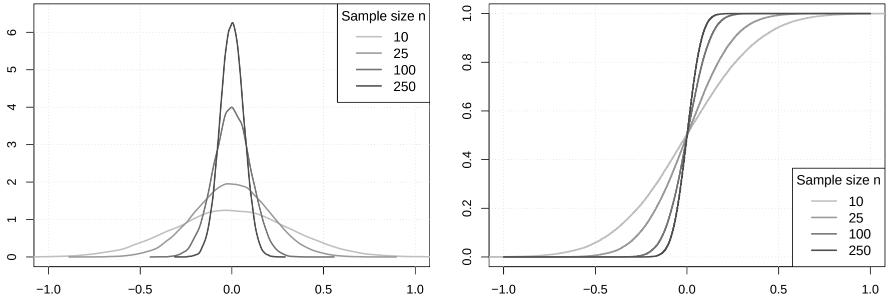
$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \xrightarrow{p} 0 \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu,$$

such that the sample average converges in probability to the expectation.

The basic flavour

Theorem (5.10 (Khinchin's WLLN))

Let $\{X_n\}$ *be a sequence of iid random variables with finite expectations* $\mathrm{E}(X_n) = \mu \;\forall n$. *Then* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$.



$X_n \sim iid\,\mathcal{N}(0,1)$. Left: pdf of $\bar{X}_n$; Right: cdf of $\bar{X}_n$

Gamma distributions

Example

Let $\{X_n\}$ be a sequence of iid random variables, with $X_n \sim$ Gamma(α, β) such that $\mathrm{E}(X_n) = \alpha\beta$. Khinchin's WLLN implies that

$$\bar{X}_n \xrightarrow{p} \mathrm{E}(X_n) = \alpha\beta.$$

Hence, for large enough n, the outcome of the random variable $\bar{X}_n$ can be taken as a close approximation of $\alpha\beta$.

This is *the* property of a *consistent estimator* for $\alpha\beta$ as we shall discuss in the course *Advanced Statistics II*.

Relax some, strengthen other assumptions...

Theorem (5.11)

Let $\{X_n\}$ be a sequence of random variables with finite variances, and let $\{\mu_n\}$ be the corresponding sequence of their expectations, Then

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0 \qquad \textit{iff} \qquad \mathrm{E}\left[\frac{(\bar{X}_n - \bar{\mu}_n)^2}{1 + (\bar{X}_n - \bar{\mu}_n)^2}\right] \to 0.$$

Theorem (5.12)

Let $\{X_n\}$ be a sequence of random variables with respective expectations given by $\{\mu_n\}$. If

$$\mathrm{Var}(\bar{X}_n) \to 0, \quad \textit{then} \quad \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

Variances

Example

Let $\{X_n\}$ be a sequence random variables, with

$$\mathrm{E}(X_i) = \mu_i = \frac{1}{2^i}\,, \qquad \mathrm{Var}(X_i) = 4, \quad \text{and} \quad \mathrm{Cov}(X_i, X_j) = 0 \quad \forall i \neq j.$$

The mean and variance for the average $\bar{X}_n$ are

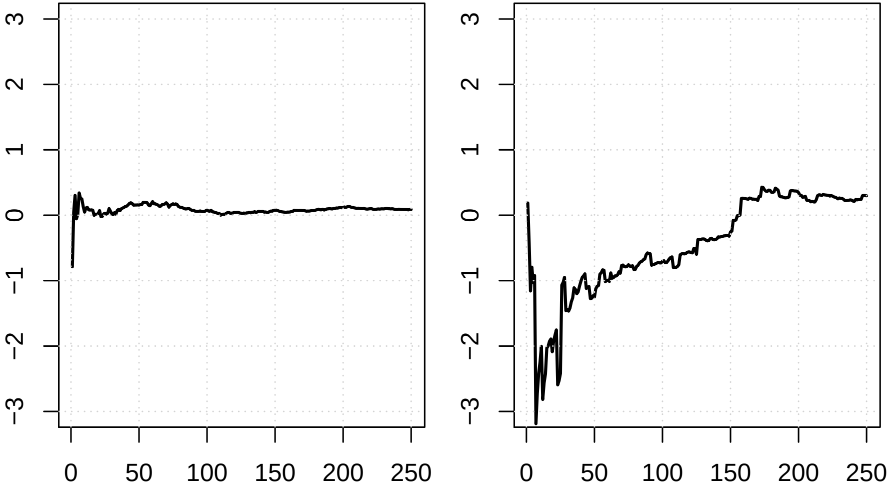
$$\mathrm{E}(\bar{X}_n) \;=\; \frac{1}{n}\sum_{i=1}^{n} \mu_i \;=\; \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2^i} \;=\; \frac{1-(\frac{1}{2})^n}{n}, \qquad \mathrm{Var}(\bar{X}_n) \;=\; \frac{4}{n}.$$

Since $\mathrm{Var}(\bar{X}_n) \to 0$, it follows by Theorem 5.12, that

$$\bar{X}_n - \bar{\mu}_n \;=\; \bar{X}_n - \frac{1-(\frac{1}{2})^n}{n} \quad \xrightarrow{p} \quad 0.$$

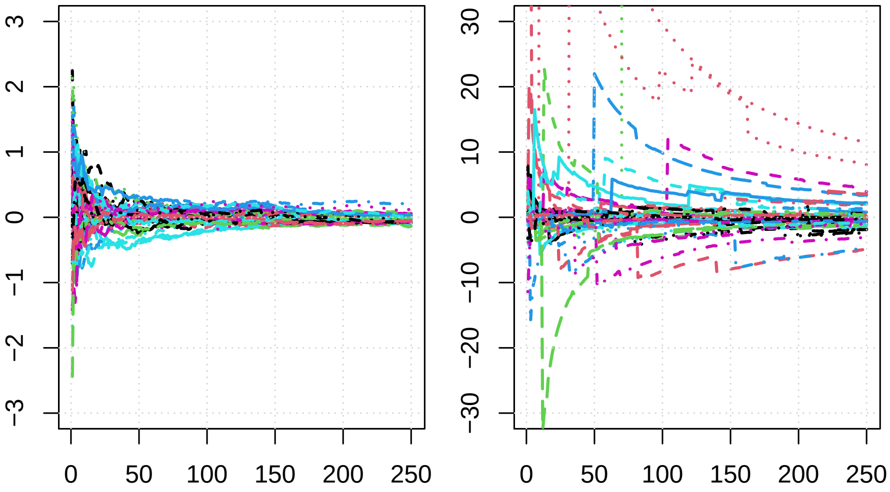
Also note that $\bar{\mu}_n \;=\; \frac{1-(\frac{1}{2})^n}{n} \to 0$, such that $\bar{X}_n \xrightarrow{p} 0$.

Finite vs. infinite mean: single outcome paths



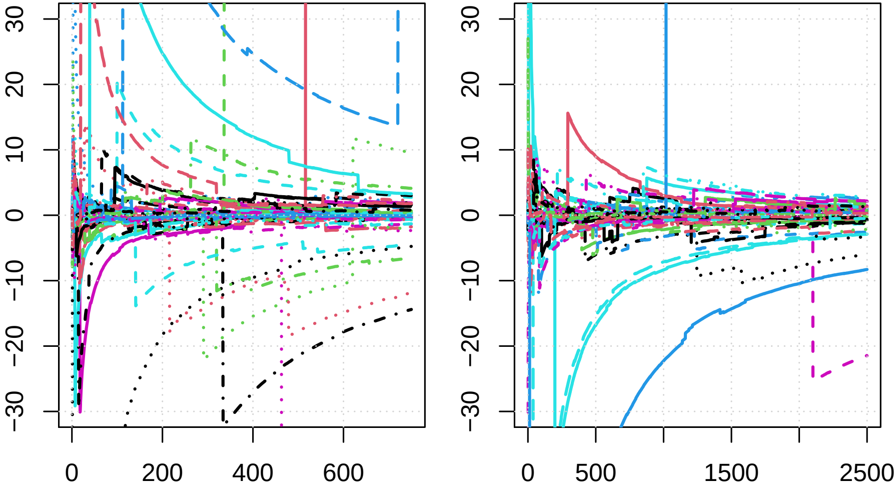
Left: path of $\bar{X}_n$ for $X_n \sim iid\,\mathcal{N}(0,1)$; Right: path of $\bar{X}_n$ for $X_n \sim iid\,t(1)$

Finite vs. infinite mean: the bigger picture



50 paths for $\mathcal{N}(0,1)$ and $t(1)$ distributed summands.

Finite vs. infinite mean: the scary picture



No apparent convergence for larger n either.

Outline

Terminology

Central limit theorems (CLTs) are concerned with the conditions under which sequences of random variables **converge in distribution** to known families of distribution.

Definition

Let $\{X_n\}$ be a sequence of random variables, and let $S_n = \sum_{i=1}^{n} X_i$, $n = 1, 2, ..$ Here we focus on the convergence in distribution of sequences of random variables of the following form

$$b_n^{-1}(S_n - a_n) \xrightarrow{d} Y \sim \mathcal{N}(0, \Sigma),$$

where $\{a_n\}$ and $\{b_n\}$ are sequences of appropriately chosen real constants.

A statement of conditions on $\{X_n\}$, $\{a_n\}$, and $\{b_n\}$ that ensure the convergence in distribution result constitutes a particular CLT. (May have other distributions than the normal as limits, actually.)

Why bother?

If a CLT applies, then

$$S_n \overset{asy}{\sim} \mathcal{N}\left(a_n, b_n^2\Sigma\right).$$

In fact:

- As we shall see in *Advanced Statistics II*, many procedures for parameter estimation and hypothesis testing are specified as functions of sums of random variables such as $S_n = \sum_{i=1}^n X_i$.
- CLTs are then often useful for establishing the asymptotic distributions for those procedures.

The basic flavour

Similar to the case of the WLLN, there are a variety of conditions that can be placed on the variables in the sum $S_n = \sum_{i=1}^{n} X_i$.

Theorem (5.13 (Lindeberg-Lévy))

Let $\{X_n\}$ *be a sequence of iid random variables with* $\mathrm{E}(X_i) = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)\ \forall i$. *Then*

$$Y_n = \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n} X_i - n\mu\right) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

Obtaining asymptotic approximations

For the variable $S_n = \sum_{i=1}^n X_i$, for example, we obtain

$$S_n = \sqrt{n}\sigma Y_n + n\mu \Big|_{\text{def.}} \overset{a}{\sim} \sqrt{n}\sigma Y + n\mu \qquad \Rightarrow \qquad S_n \overset{a}{\sim} \mathcal{N}(n\mu, n\sigma^2).$$

For the average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ we have

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}} Y_n + \mu \Big|_{\text{def.}} \overset{a}{\sim} \frac{\sigma}{\sqrt{n}} Y + \mu \qquad \Rightarrow \qquad \bar{X}_n \overset{a}{\sim} \mathcal{N}(\mu, \tfrac{\sigma^2}{n}).$$

Don't forget that, actually, $|S_n| \xrightarrow{p} \infty$ and $\bar{X}_n \xrightarrow{p} \mu$.

Average durations

Example

Let X_n be a sequence of independent exponentially distributed durations,

$$f(x) = \lambda \exp(-\lambda x)\mathbb{I}(x \geq 0).$$

We are interested in the average duration $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and its distribution!

1. The sum $S_n = \sum_{i=1}^{n} X_i$ is Gamma-distributed with parameters $\alpha = n$ and $\beta = 1/\lambda$,
$$f_S(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}.$$
2. So $\bar{X}_n = \frac{1}{n} S$ is Gamma-distributed with parameters $\alpha = n$ and $\beta = 1/(n\lambda)$,
$$f_{\bar{X}_n}(x) = \frac{(n\lambda)^n}{\Gamma(n)} x^{n-1} e^{-n\lambda x}.$$

... and their approximate distribution

Example (cont'd)

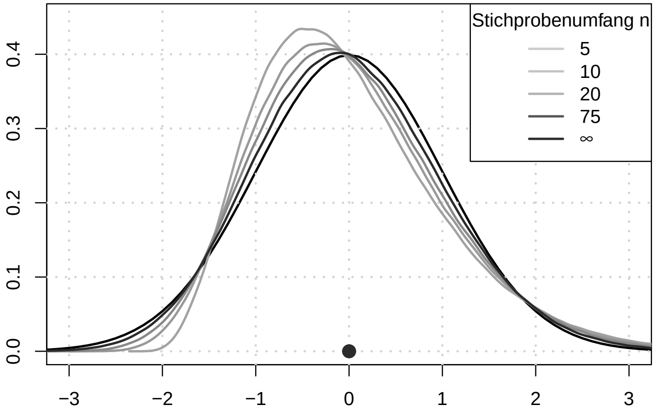
- Since $X_n \sim iid$ with $\mathrm{E}(X_n) = \frac{1}{\lambda}$ & $\mathrm{Var}(X_n) = \frac{1}{\lambda^2} < \infty$,

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\lambda}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right),$$

- ... leadit to the approximation

$$\bar{X}_n \overset{approx}{\sim} \mathcal{N}\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right).$$

How good is the normal approximation?



Standard normal and exact distribution of $\sqrt{n\lambda^2}(\bar{X}_n - 1/\lambda)$, $\lambda = 1$, $n \in \{5, 10, 20, 75\}$.

The iid assumption

The CLT of Lindberg-Levy requires that the **random variables are iid**.

- However, in many applications the assumption that the variables are iid is violated since we have variables which are correlated and/or have different distributions.
- Fortunately, there are various other CLTs, which do not need the iid condition. Instead, they place alternative conditions on the stochastic behavior of the random variables in the sequence $\{X_n\}$.

In Time Series Analysis, we allow for serial dependence. Below, we allow for heterogeneity.

Lindeberg's CLT

Theorem (5.14 (Lindeberg's CLT))

Let $\{X_n\}$ *be a sequence of independent random variables with* $\mathrm{E}(X_i) = \mu_i$ *and* $\mathrm{Var}\,(X_i) = \sigma_i^2 < \infty \, \forall \, i$. *Define* $b_n^2 = \sum_{i=1}^n \sigma_i^2$, $\bar{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \sigma_i^2$, $\bar{\mu}_n = n^{-1} \sum_{i=1}^n \mu_i$, *and let* f_i *be the pdf of* X_i. *If* $\forall \, \varepsilon > 0$,

(continuous case:)
$$\lim_{n\to\infty} \frac{1}{b_n^2} \sum_{i=1}^n \int_{(x_i-\mu_i)^2 \geq \varepsilon b_n^2} (x_i - \mu_i)^2 f_i(x_i) \, \mathrm{d}x_i = 0,$$

(discrete case:)
$$\lim_{n\to\infty} \frac{1}{b_n^2} \sum_{i=1}^n \sum_{\substack{(x_i-\mu_i)^2 \geq \varepsilon b_n^2 \\ f_i(x_i) > 0}} (x_i - \mu_i)^2 f_i(x_i) = 0,$$

then

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} = \frac{n^{1/2}\left(\bar{X}_n - \bar{\mu}_n\right)}{\bar{\sigma}_n} \xrightarrow{d} \mathcal{N}(0,1).$$

Bounded random variables satisfy this (see Theorem 5.15 in lecture notes).

Multivariate Central Limit Theorems

The CLTs presented so far are applicable to sequences of **random scalars**.

In order to discuss CLTs for a sequence of **random vectors** a result of Cramér and Wold, termed the **Cramér-Wold device** is very useful.

The Cramér-Wold device allows to reduce the question of convergence in distribution for multivariate random vectors to the question of convergence in distribution for random scalars.

Thus it facilitates the use of CLTs for random scalars in order to obtain multivariate CLTs.

The CW device

Theorem (5.16 (Cramér-Wold Device))

The sequence of $(k \times 1)$-dim. random vectors $\{\boldsymbol{X}_n\}$ converges in distribution to the $(k \times 1)$-dim. random vector $\boldsymbol{X}$ iff

$$\boldsymbol{\ell}' \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{\ell}' \boldsymbol{X} \qquad \forall \ \boldsymbol{\ell} \in \mathbb{R}^k.$$

This really means "any linear combination"...

Corollary (5.2)

$$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \textit{iff} \qquad \boldsymbol{\ell}' \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{\ell}' \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\ell}' \boldsymbol{\mu} \ , \ \boldsymbol{\ell}' \boldsymbol{\Sigma} \boldsymbol{\ell}).$$

... allows for multivariate CLTs

Theorem (5.17 (Multivariate Lindeberg-Lévy))

Let $\{\boldsymbol{X}_n\}$ *be a sequence of iid* $(k \times 1)$ *random vectors with* $\mathrm{E}(\boldsymbol{X}_i) = \boldsymbol{\mu}$ *and* $\mathrm{Cov}(\boldsymbol{X}_i) = \boldsymbol{\Sigma}\ \forall\, i$*, where* $\boldsymbol{\Sigma}$ *is a* $(k \times k)$ *positive definite matrix. Then*

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i - \boldsymbol{\mu}\Big) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right).$$

It follows from the multivariate Lindeberg-Lévy CLT that $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$.

Outline

A transformation

Take the exponential durations again, where we are interested in $\frac{1}{\bar{X}_n}$:

- Since $\bar{X}_n \xrightarrow{p} \frac{1}{\lambda}$, the CMT implies $\frac{1}{\bar{X}_n} \xrightarrow{p} \lambda$
- ... and we may *estimate* an unknown λ by means of $\frac{1}{\bar{X}_n}$.

Since $\bar{X}_n$ is Gamma-distributed, $1/\bar{X}_n$ follows a so-called inverse Gamma distribution which is not just as tractable.

Recall however that $\bar{X}_n \overset{asy}{\sim} \mathcal{N}\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$.

- We could use that by working out the distribution of $1/\mathcal{N}(\mu, \sigma^2)$. Again not very tractable.
- But if we linearized $1/x$ in a neighbourhood of $\frac{1}{\lambda}$...

Functions of asymptotically normally distributed RVs

Theorem (5.18 (The delta method))

Take $\{\boldsymbol{X}_n\} \in \mathbb{R}^k$ *where* $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

Let $g(\boldsymbol{x})$ *be a function that has first-order partial derivatives in a neighborhood of the point* $\boldsymbol{x} = \boldsymbol{\mu}$ *that are continuous at* $\boldsymbol{\mu}$*, and suppose the gradient vector of* $g(\boldsymbol{x})$ *evaluated at* $\boldsymbol{x} = \boldsymbol{\mu}$,

$$\boldsymbol{g}_{(1\times k)} = \left[\partial g(\boldsymbol{\mu})/\partial x_1 \ldots \partial g(\boldsymbol{\mu})/\partial x_k\right],$$

is not the zero vector.[a] *Then*

$$\sqrt{n}\left(g\left(\boldsymbol{X}_n\right) - g\left(\boldsymbol{\mu}\right)\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{g}\boldsymbol{\Sigma}\boldsymbol{g}')$$

and correspondingly $g\left(\boldsymbol{X}_n\right) \overset{asy}{\sim} \mathcal{N}\left(g(\boldsymbol{\mu}), n^{-1}\boldsymbol{g}\boldsymbol{\Sigma}\boldsymbol{g}'\right)$.

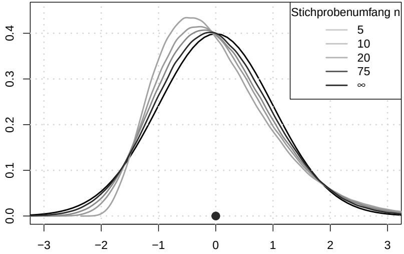
[a] This is the row version of the gradient.

Exponential durations take three

We have with $\mu = 1/\lambda$ that $\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right)$, and $g' = -1/x^2$.

So $\sqrt{n}\left(\frac{1}{\bar{X}_n} - \lambda\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda^4}{\lambda^2}\right)$ and

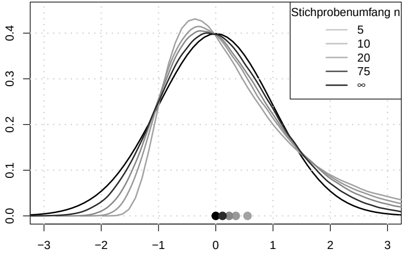
$$\frac{1}{\bar{X}_n} \overset{approx}{\sim} \mathcal{N}\left(\lambda, \frac{\lambda^2}{n}\right).$$



Right: exact distributions, $\lambda = 1$.

Top: $\sqrt{n\lambda^2}(\bar{X}_n - 1/\lambda)$

Bottom: $\sqrt{n/\lambda^2}\left(\frac{1}{\bar{X}_n} - \lambda\right)$



The normal only approximates!

Some final remarks

The delta method is based on a first-order Taylor approximation.

- We note that we may work analogously with a vector function $\boldsymbol{g}(\boldsymbol{x})$ (the limit then involves the Jacobian of $\boldsymbol{g}$).
- If $\boldsymbol{g}$ is smooth enough, a higher-order Taylor approximation can tell you what the approximation error looks like.
- A higher-order approximation also helps when $\partial g/\partial \boldsymbol{x} = \mathbf{0}$ at $\boldsymbol{x} = \boldsymbol{\mu}$.
- (By the way, such a stationary point of g simply means that $g\left(\boldsymbol{X}_n\right) - g\left(\boldsymbol{\mu}\right) = o_p(n^{-1/2})$.)
- And we can even work out approximations for moments of $g\left(\boldsymbol{X}_n\right)$.

If g is not smooth (or not continuous) at $\boldsymbol{\mu}$, then we need a case-by-case discussion, but these are fortunately rarely encountered.

Outline

Coming up

Inferential Statistics (*fka* Adv Stat II)