9. Asymptotics

Spring 2021

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Last time: introducing estimators, looking at finite-sample properties.
- Now: can we say more as sample size grows?

1/ Asymptotics

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
    - $\overline{X}_n$ is **unbiased**, $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i] = \mu$
    - Sampling variance is $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$ where $\sigma^2 = \mathbb{V}[X_i]$
    - None of these rely on a **specific distribution** for X_i!
- Assuming $X_i \sim \mathcal{N}(\mu, \sigma^2)$, we know the exact distribution of $\overline{X}_n$.
    - What if the data isn't normal? What is the sampling distribution of $\overline{X}_n$?
- **Asymptotics**: approximate the sampling distribution of $\overline{X}_n$ as n gets big.

Sequence of sample means

- What can we say about the sample mean n gets large?
- Need to think about sequences of sample means with increasing n:

$$\begin{aligned}
\overline{X}_1 &= X_1 \\
\overline{X}_2 &= (1/2) \cdot (X_1 + X_2) \\
\overline{X}_3 &= (1/3) \cdot (X_1 + X_2 + X_3) \\
\overline{X}_4 &= (1/4) \cdot (X_1 + X_2 + X_3 + X_4) \\
\overline{X}_5 &= (1/5) \cdot (X_1 + X_2 + X_3 + X_4 + X_5) \\
&\vdots \\
\overline{X}_n &= (1/n) \cdot (X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_n)
\end{aligned}$$

- Note: this is a sequence of random variables!

Asymptotics and Limits

- Asymptotic analysis is about making **approximations** to finite sample properties.
- Useful to know some properties of deterministic sequences:

Definition

A sequence $\{a_n : n = 1, 2, \ldots\}$ has the **limit** a written $a_n \to a$ as $n \to \infty$ if for all $\delta > 0$ there is some $n_\delta < \infty$ such that for all $n \geq n_\delta$, $|a_n - a| \leq \delta$.

- a_n gets closer and closer to a as n gets larger (a_n **converges** to a)
- $\{a_n : n = 1, 2, \ldots\}$ is **bounded** if there is $b < \infty$ such that $|a_n| < b$ for all n.

Convergence in Probability

Definition

A sequence of random variables, $\{Z_n : n = 1, 2, ...\}$, is said to **converge in probability** to a value b if for every $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - b| > \varepsilon) \to 0,$$

as $n \to \infty$. We write this $Z_n \xrightarrow{p} b$.

- Basically: probability that Z_n lies outside any (teeny, tiny) interval around b approaches 0 as $n \to \infty$
- Economists writes $\text{plim}(Z_n) = b$ if $Z_n \xrightarrow{p} b$.
- An estimator is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$.
    - Distribution of $\hat{\theta}_n$ collapses on θ as $n \to \infty$.
    - Inconsistent estimator are bad bad bad: more data gives worse answers!

Chebyshev Inequality

- How can we show convergence in probability? Can verify if we know specific distribution of $\hat{\theta}$.
- But can we say anything for arbitrary distributions?

Chebyshev Inequality

Suppose that X is r.v. for which $\mathbb{V}[X] < \infty$. Then, for every real number $\delta > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta) \leq \frac{\mathbb{V}[X]}{\delta^2}.$$

- Variance places limits on how far an observation can be from its mean.

Proof of Chebyshev

- Let $Z = X - \mathbb{E}[X]$ with density $f_Z(x)$. Probability is just integral over the region:

$$\mathbb{P}\left(|Z| \geq \delta\right) = \int_{|x| \geq \delta} f_Z(x) dx$$

- Note that where $|x| \geq \delta$, we have $1 \leq x^2/\delta^2$, so

$$\mathbb{P}\left(|Z| \geq \delta\right) \leq \int_{|x| \geq \delta} \frac{x^2}{\delta^2} f_Z(x) dx \leq \int_{-\infty}^{\infty} \frac{x^2}{\delta^2} f_Z(x) dx = \frac{\mathbb{E}[Z^2]}{\delta^2} = \frac{\mathbb{V}[X]}{\delta^2}$$

Markov Inequality

Markov Inequality

For any r.v. X and any $\delta > 0$,

$$\mathbb{P}(|X| \geq \delta) \leq \frac{\mathbb{E}[|X|]}{\delta}.$$

- The expectation limits how much probability can be in the tail.
- Proof similar to Chebyshev.

Consistency

- We can use Chebyshev to determine consistency when variance is finite.
- **Theorem**: For any sequence of r.v.s, Z_n with $\mathbb{V}[Z_n] \to 0$, then $Z_n - \mathbb{E}[Z_n] \xrightarrow{p} 0$.
- If $\text{bias}[\hat{\theta}_n] \to 0$ and $\mathbb{V}[\hat{\theta}_n] \to 0$ as $n \to \infty$, then $\hat{\theta}_n$ is consistent.
- Example: sample mean.
  - $\overline{X}_n$ is unbiased for μ with $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\rightsquigarrow \overline{X}_n$ consistent since $\mathbb{V}[\overline{X}_n] \to 0$
- NB: Unbiasedness does not imply consistency, nor vice versa.

Law of large numbers

Weak Law of Large Numbers

Let $X_1, \dots, X_n$ be a an i.i.d. draws from a distribution with mean $\mathbb{E}[X_i] < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then, $\overline{X}_n \xrightarrow{p} \mathbb{E}[X_i]$.

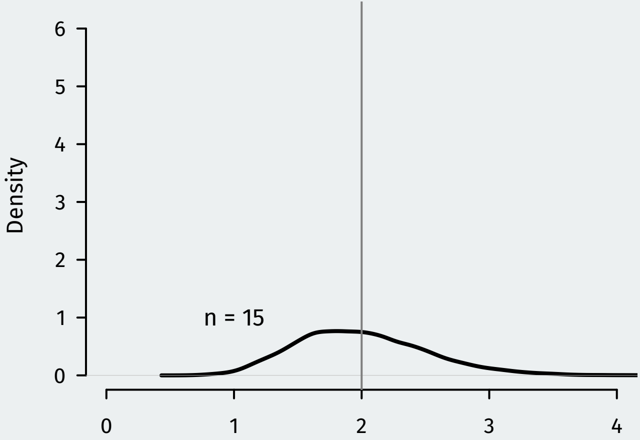
- Note: we don't assume finite variance, only finite expectation.
- Intuition: The probability of $\overline{X}_n$ being "far away" from μ goes to 0 as n gets big.
- Implies many sample means converge:
  - If $\mathbb{E}[X_i^2] < \infty$, then $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} \mathbb{E}[X_i^2]$

LLN by simulation in R

- Draw different sample sizes from Exponential distribution with rate 0.5
- $\rightsquigarrow \mathbb{E}[X_i] = 2$

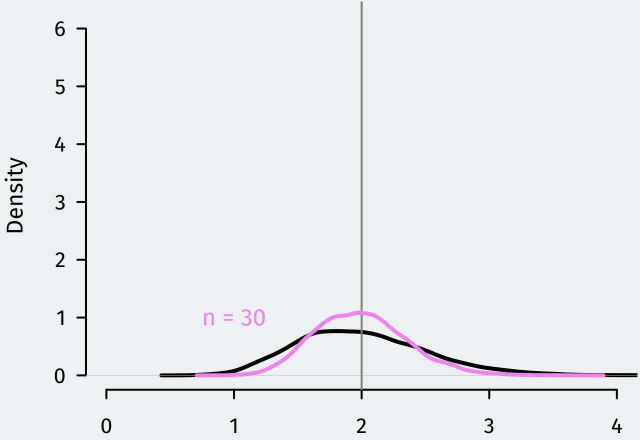
```r
nsims <- 10000
holder <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
  s5 <- rexp(n = 5, rate = 0.5)
  s15 <- rexp(n = 15, rate = 0.5)
  s30 <- rexp(n = 30, rate = 0.5)
  s100 <- rexp(n = 100, rate = 0.5)
  s1000 <- rexp(n = 1000, rate = 0.5)
  s10000 <- rexp(n = 10000, rate = 0.5)

  holder[i,1] <- mean(s5)
  holder[i,2] <- mean(s15)
  holder[i,3] <- mean(s30)
  holder[i,4] <- mean(s100)
  holder[i,5] <- mean(s1000)
  holder[i,6] <- mean(s10000)
}
```

LLN in action



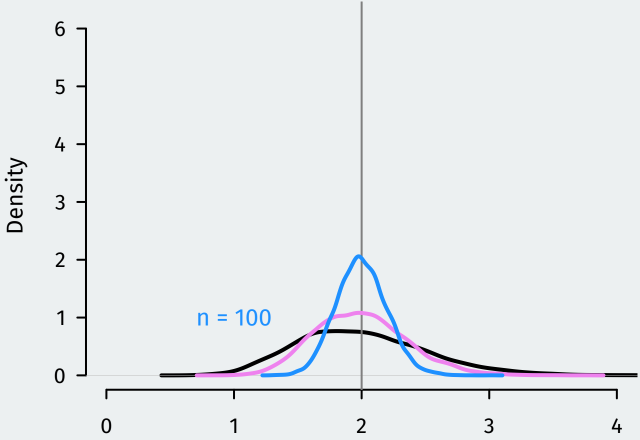
- Distribution of $\overline{X}_{15}$

LLN in action



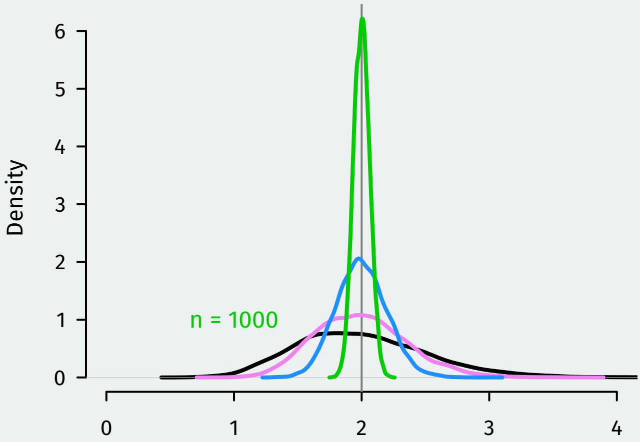
- Distribution of $\overline{X}_{30}$

LLN in action



- Distribution of $\overline{X}_{100}$

LLN in action



- Distribution of $\overline{X}_{1000}$

Properties of convergence in probability

1. **Continuous mapping theorem**: if $X_n \xrightarrow{p} c$, then $g(X_n) \xrightarrow{p} g(c)$ for any continuous function g.
2. if $X_n \xrightarrow{p} a$ and $Z_n \xrightarrow{p} b$, then
    - $X_n + Z_n \xrightarrow{p} a + b$
    - $X_n Z_n \xrightarrow{p} ab$
    - $X_n / Z_n \xrightarrow{p} a/b$ if $b > 0$

- Thus, by LLN and CMT:
    - $\left(\overline{X}_n\right)^2 \xrightarrow{p} \mu^2$
    - $\log(\overline{X}_n) \xrightarrow{p} \log(\mu)$

Unbiased versus consistent

- **Unbiased, not consistent**: "first observation" estimator, $\hat{\theta}_n^f = X_1$.
    - Unbiased because $\mathbb{E}[\hat{\theta}_n^f] = \mathbb{E}[X_1] = \mu$
    - Not consistent: $\hat{\theta}_n^f$ is constant in n so its distribution never collapses.
    - Said differently: the variance of $\hat{\theta}_n^f$ never shrinks.
- **Consistent, but biased**: sample mean with n replaced by $n-1$:

$$\frac{n}{n-1}\overline{X}_n = \frac{1}{n-1}\sum_{i=1}^{n} X_i$$

- Bias: $\mathbb{E}[\frac{n}{n-1}\overline{X}_n] - \mu = \frac{1}{n-1}\mu$
- Consistent because bias and se $\to 0$ as $n \to \infty$.

Multivariate LLN

- Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$ be a random vectors of length k.
- Random (iid) sample of n of these k vectors, $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
- Vector sample mean:

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = \begin{pmatrix} \overline{X}_{n,1} \\ \overline{X}_{n,2} \\ \vdots \\ \overline{X}_{n,k} \end{pmatrix}$$

- **Vector WLLN**: if $\mathbb{E}[\|\mathbf{X}\|] < \infty$, then as $n \to \infty$, $\overline{\mathbf{X}}_n \xrightarrow{p} \mathbb{E}[\mathbf{X}]$.
    - Converge in probability of a vector is just convergence of each element.
    - $\mathbb{E}[\|\mathbf{X}\|] < \infty$ is equivalent to $\mathbb{E}[|X_{ij}|] < \infty$ for each $j = 1, \ldots, k$

2/ Central Limit Theorem

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!
- Can we say more?
  - Can we approximate $\Pr(a < \overline{X}_n < b)$?
  - What family of distributions (Binomial, Uniform, Gamma, etc)?
- Again, need to analyze when n is large.

Convergence in Distribution

Definition

Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $G_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $G_W(u)$ if

$$\lim_{n \to \infty} G_n(u) = G_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W
    - Also known as the **asymptotic distribution** or **large-sample distribution**
- We use c.d.f.s here to avoid messy details with discrete vs continuous.
- If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$

Central Limit Theorem

Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

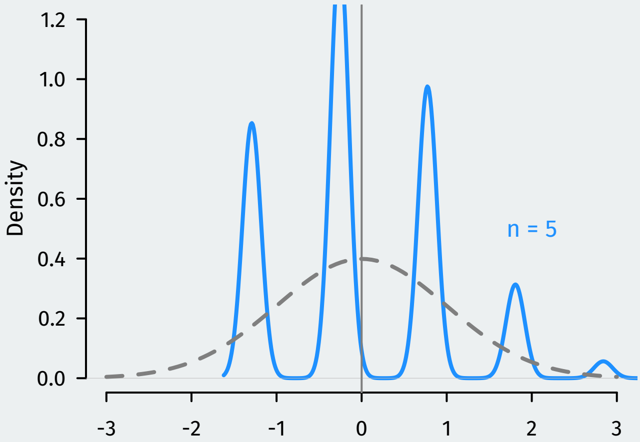
$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- Distribution free! No specific assumptions about the distribution of X_i except finite variance.
- Implies that $\overline{X}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$,
  - $\overset{a}{\sim}$ is "approximately distributed as".
- $\rightsquigarrow$ easy approximations to probability statements about $\overline{X}_n$ when n is big!

CLT by simulation in R

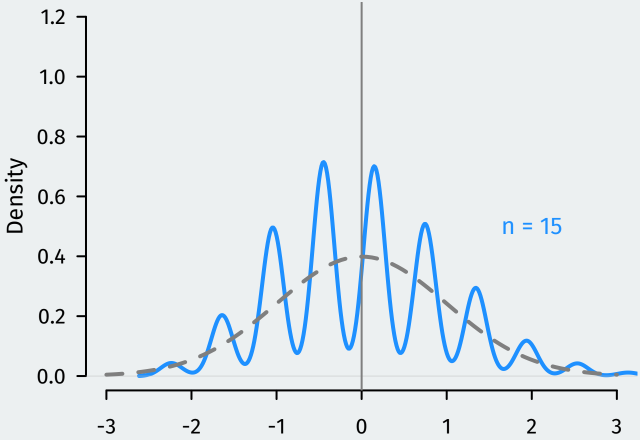
```r
set.seed(02138)
nsims <- 10000
holder2 <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
  s5 <- rbinom(n = 5, size = 1, prob = 0.25)
  s15 <- rbinom(n = 15, size = 1, prob = 0.25)
  s30 <- rbinom(n = 30, size = 1, prob = 0.25)
  s100 <- rbinom(n = 100, size = 1, prob = 0.25)
  s1000 <- rbinom(n = 1000, size = 1, prob = 0.25)
  s10000 <- rbinom(n = 10000, size = 1, prob = 0.25)

  holder2[i,1] <- mean(s5)
  holder2[i,2] <- mean(s15)
  holder2[i,3] <- mean(s30)
  holder2[i,4] <- mean(s100)
  holder2[i,5] <- mean(s1000)
  holder2[i,6] <- mean(s10000)
}
```

CLT in action



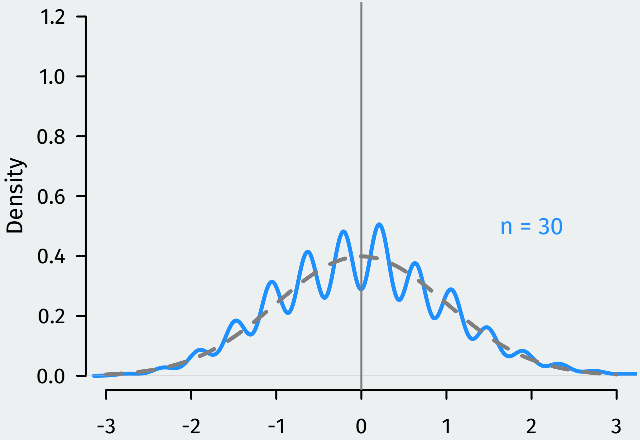
- Distribution of $\frac{\overline{X}_5-\mu}{\sigma/\sqrt{5}}$

CLT in action



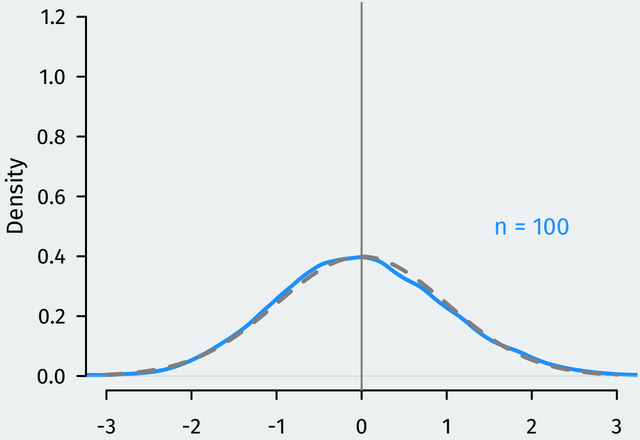
- Distribution of $\frac{\overline{X}_{15}-\mu}{\sigma/\sqrt{15}}$

CLT in action



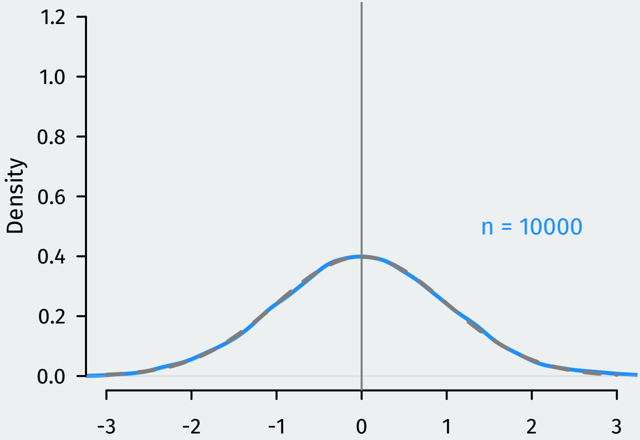
- Distribution of $\frac{\overline{X}_{30} - \mu}{\sigma / \sqrt{30}}$

CLT in action



- Distribution of $\frac{\overline{X}_{100}-\mu}{\sigma/\sqrt{100}}$

CLT in action



- Distribution of $\frac{\overline{X}_{10000}-\mu}{\sigma/\sqrt{10000}}$

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \qquad \Longrightarrow \qquad g(Z_n) \xrightarrow{d} g(Z).$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number, c
- Slutsky's Theorem gives the following result:
    1. $X_n Y_n$ converges in distribution to cX
    2. $X_n + Y_n$ converges in distribution to $X + c$
    3. X_n / Y_n converges in distribution to X/c if $c \neq 0$
- Extremely useful when trying to figure out what the large-sample distribution of an estimator is.

Delta method

Delta method

If $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V).$$

- Why $h()$ continuously differentiable?
  - Near θ we can approximate $h()$ with a line where h' is the slope.
  - So $h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)\left(\hat{\theta}_n - \theta\right)$

Asymptotic normality

- An estimator is **asymptotically normal** if

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, \mathbb{V}[\hat{\theta}_n]\right)$$

  - Allows us to approximate the probability of $\hat{\theta}_n$ being far away from θ in large samples.

- Usually follows by some version of the CLT.
  - CLT: $\overline{X}_n$ is **asymptotically normal**

Variance estimation with plug-in estimators

- Setting: $X_1, \dots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$, by CLT:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

  where $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$.

- But we don't know V_θ?! Estimate it!

$$\widehat{V_\theta} = \frac{1}{n}\sum_{i=1}^{n}\left(g(X_i) - \hat{\theta}_n\right)^2$$

- We can show that $\widehat{V_\theta} \xrightarrow{p} V_\theta$ and so by Slutsky:

$$\frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\widehat{V_\theta}}} \xrightarrow{d} \frac{\mathcal{N}(0, V_\theta)}{\sqrt{V_\theta}} \sim \mathcal{N}(0, 1)$$

Multivariate CLT

- Convergence in distribution is the same vector $\mathbf{Z}_n$: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate Central Limit Theorem

If $\mathbf{X}_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$, then as $n \to \infty$,

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \overset{d}{\to} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i]$ and $\boldsymbol{\Sigma} = \mathbb{V}[\mathbf{X}_i] = \mathbb{E}\left[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})'\right]$.

- $\mathbb{E}\|\mathbf{X}_i\|^2 < \infty$ is equivalent to $\mathbb{E}[X_{i,j}^2] < \infty$ for all $j = 1, \ldots, k$.
  - Basically: multivariate CLT holds if each r.v. in the vector has finite variance.
- Very common for when we're estimating multiple parameters $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}_n$

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?
- Let $\mathbf{h}(\boldsymbol{\theta})$ map from $\mathbb{R}^k \to \mathbb{R}^m$ and be continuously differentiable.
  - Ex: $\mathbf{h}(\theta_1, \theta_2, \theta_3) = (\theta_2/\theta_1, \theta_3/\theta_1)$, from $\mathbb{R}^3 \to \mathbb{R}^2$
  - Like univariate case, we need the derivatives arranged in $m \times k$ Jacobian matrix:

$$\mathbf{H}(\boldsymbol{\theta}) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}\mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial h_1}{\partial \theta_1} & \frac{\partial h_1}{\partial \theta_2} & \cdots & \frac{\partial h_1}{\partial \theta_k} \\ \frac{\partial h_2}{\partial \theta_1} & \frac{\partial h_2}{\partial \theta_2} & \cdots & \frac{\partial h_2}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial \theta_1} & \frac{\partial h_m}{\partial \theta_2} & \cdots & \frac{\partial h_m}{\partial \theta_k} \end{pmatrix}$$

- Multivariate delta method: if $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma})$, then

$$\sqrt{n}\left(\mathbf{h}(\hat{\boldsymbol{\theta}}_n) - \mathbf{h}(\boldsymbol{\theta})\right) \xrightarrow{d} \mathcal{N}(0, \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Sigma}\mathbf{H}(\boldsymbol{\theta})')$$

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
  - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
  - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^\lambda)$
  - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
  - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^\lambda)$
- Stochastic order notation for random sequence, Z_n
  - If $Z_n \xrightarrow{p} 0$, we write $Z_n = o_p(1)$ ("little-oh-p-one").
  - For any consistent estimator, we have $\hat{\theta}_n = \theta + o_p(1)$
  - If $a_n^{-1} Z_n \xrightarrow{p} 0$, we write $Z_n = o_p(a_n)$

Bounded in probability

Definition

A random sequence Z_n is **bounded in probability**, written $Z_n = O_p(1)$ ("big-oh-p-one") for all $\delta > 0$ there exists a M_δ and n_δ, such that for $n \geq n_\delta$,

$$\mathbb{P}(|Z_n| > M_\delta) < \delta$$

- $Z_n = o_p(1)$ implies $Z_n = O_p(1)$ but not the reverse.
- If Z_n converges in distribution, it is $O_p(1)$, so if the CLT applies we have:

$$\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$$

  - If $a_n^{-1} Z_n = O_p(1)$, we write $Z_n = O_p(a_n)$, so we have: $\hat{\theta}_n = \theta + O_p(n^{-1/2})$.