Theoretical Statistics, STOR 655

Tools for Extending Weak Convergence

Andrew Nobel

January 2023

Extending Weak Convergence

Of interest: Extending existing weak convergence results, in particular the multivariate CLT, to more complex settings. Key tools:

- Continuous mapping theorem
- Slutsky's theorem
- The Delta method

Continuous Mapping Theorem

Extending Weak Convergence: Continuous Mapping Theorem

Theorem (CMT): Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be a function with continuity set

$$C(g) = \{x \in \mathbb{R}^d \text{ s.t. } g \text{ is continuous at } x\}.$$

If $X_n \Rightarrow X$ in $\mathbb{R}^d$ and $\mathbb{P}(X \in C(g)) = 1$, then $g(X_n) \Rightarrow g(X)$ in $\mathbb{R}^k$

Note: If g is continuous then $C(g) = \mathbb{R}^d$ and weak convergence extends from X_n to $g(X_n)$ without further assumptions

First Examples of Continuous Mapping Theorem

If $X_1, X_2, \ldots$ are random variables with $X_n \Rightarrow X \sim \mathcal{N}(0,1)$ then

- $X_n^2 \Rightarrow X^2 \sim \chi_1^2$ distribution
- $e^{X_n} \Rightarrow e^X \sim$ Log-normal distribution
- $X_n^{-1} \Rightarrow X^{-1} \sim$ Inverse normal distribution

Slutsky’s Theorem

The Gluing Lemma

Lemma: Let $X_1, X_2, \ldots \in \mathbb{R}^d$ and $Y_1, Y_2, \ldots \in \mathbb{R}^k$. If $X_n \Rightarrow X$ and $Y_n \Rightarrow v$, a constant vector, then

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \Rightarrow \begin{bmatrix} X \\ v \end{bmatrix}$$

Note: Result does not hold if weak limit of Y_n is not constant

Example: Let $Z \sim \mathcal{N}(0,1)$. Define $X_n = Z$ and $Y_n = (-1)^n Z$. Then it is easy to see that $X_n \Rightarrow Z$ and $Y_n \Rightarrow Z$, but

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \not\Rightarrow \begin{bmatrix} Z \\ Z \end{bmatrix}$$

Slutsky's Theorem

Idea: Apply Gluing Lemma and CMT to different continuous functions $g(x, y)$

Theorem: Suppose that $X_n \Rightarrow X$ in $\mathbb{R}^d$

1. If $Y_n \Rightarrow v$ then $X_n + Y_n \Rightarrow X + v$
2. If $Y_n = o_p(1)$ then $X_n + Y_n \Rightarrow X$
3. If $Z_n \Rightarrow c$ then $X_n Z_n \Rightarrow X c$
4. If $Z_n \neq 0$ and $Z_n \Rightarrow c \neq 0$ then $X_n Z_n^{-1} \Rightarrow c^{-1} X$

The Delta Method

The Delta Method

Motivation: Suppose that $X_1, X_2, \ldots \in \mathbb{R}$ satisfy the usual CLT

$$n^{1/2}(X_n - \mu) \Rightarrow \mathcal{N}(0,1)$$

Let $\phi : \mathbb{R} \to \mathbb{R}$ be differentiable at μ. Expanding ϕ in a Taylor series around μ, an informal argument using the CMT and Slutsky yields

$$n^{1/2}(\phi(X_n) - \phi(\mu)) \Rightarrow \mathcal{N}(0, \phi'(\mu)^2)$$

Upshot: Asymptotic normality of X_n yields asymptotic normality of $\phi(X_n)$

Task: Make this rigorous

A Connection Between Standard and Stochastic Order

Fact: Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be random vectors with $X_n = o_p(1)$ and let $\phi : \mathbb{R}^d \to \mathbb{R}$ satisfy $\phi(0) = 0$. Fix $q > 0$.

1. If $\phi(x) = o(||x||^q)$ as $x \to 0$ then $\phi(X_n) = o_p(||X_n||^q)$
2. If $\phi(x) = O(||x||^q)$ as $x \to 0$ then $\phi(X_n) = O_p(||X_n||^q)$

The Delta Method, Formal Statement

Theorem: Suppose $T_1, T_2, \ldots, V \in \mathbb{R}^d$ and $\theta \in \mathbb{R}^d$ are such that

$$r_n(T_n - \theta) \Rightarrow V \text{ where } r_n \to \infty$$

If $\phi : \mathbb{R}^d \to \mathbb{R}^k$ is differentiable at θ with derivative matrix $\dot{\phi}(\theta) \in \mathbb{R}^{k \times d}$ then

$$r_n(\phi(T_n) - \phi(\theta)) \Rightarrow \dot{\phi}(\theta) V$$

Cor: If $n^{1/2}(T_n - \theta) \Rightarrow \mathcal{N}_d(\mu, \Sigma)$ and $\phi : \mathbb{R}^d \to \mathbb{R}^k$ is differentiable at θ then

$$n^{1/2}(\phi(T_n) - \phi(\theta)) \Rightarrow \mathcal{N}_k(\dot{\phi}(\theta)\,\mu, \dot{\phi}(\theta)\,\Sigma\,\dot{\phi}(\theta)^t)$$

Delta Method, First Examples

Let $X_1, X_2, \ldots \in \mathbb{R}$ be iid with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. By CLT

$$n^{1/2}(\overline{X}_n - \mu) \Rightarrow \mathcal{N}(0, \sigma^2)$$

A. CLT for $(\overline{X}_n)^2$: delta method gives $n^{1/2}((\overline{X}_n)^2 - \mu^2) \Rightarrow \mathcal{N}(0, 4\sigma^2\mu^2)$

A'. CLT for $(\overline{X}_n)^2$: when $\mu = 0$ direct analysis gives $n(\overline{X}_n)^2 \Rightarrow \sigma^2\chi_1^2$

B. CLT for $(X_n)^{-1}$: if $\mu \neq 0$ we have $n^{1/2}\left((\overline{X}_n)^{-1} - \mu^{-1}\right) \Rightarrow \mathcal{N}(0, \sigma^2/\mu^4)$

Skewness and Kurtosis

Definition: Let $X \in \mathbb{R}$ be a random variable

- For $j \geq 1$ let $\alpha_j = \mathbb{E}X^j$ (moments of X)
- For $j \geq 1$ let $\mu_j = \mathbb{E}(X - \mathbb{E}X)^j$ (central moments of X)
- The *skewness* of X is $\lambda(X) = \mu_3/\mu_2^{3/2} = \mu_3/\sigma^3$
- The *kurtosis* of X is $\kappa(X) = \mu_4/\mu_2^2 - 3$

Note: For $Z \sim \mathcal{N}(0,1)$ it is easy to see that $\lambda(Z) = \kappa(Z) = 0$

- $\lambda(Z) \neq 0$ for asymmetric distributions
- $\kappa(Z) > 0$ for distributions with heavier tails than the normal

Sample Variance, Consistency

Let $X_1, X_2, \ldots X \in \mathbb{R}$ be iid with $\mathbb{E}X = \mu$, $\mathrm{Var}(X) = \sigma^2$, and $\mathbb{E}X^4 < \infty$

Goal: Inference about $\mathrm{Var}(X)$. Natural starting point is sample variance

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

where $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $\overline{X}_n^2 = n^{-1}\sum_{i=1}^n X_i^2$.

Consistency: By LLN $\overline{X}_n \to \mathbb{E}X$ and $\overline{X}_n^2 \to \mathbb{E}X^2$ wp1, so

$$S_n^2 \to \mathbb{E}X^2 - (\mathbb{E}X)^2 = \sigma^2 \text{ wp1}$$

Sample Variance, Asymptotic Normality

Fact: If $X_1, X_2, \ldots \in \mathbb{R}$ are iid with $\mathbb{E}X_i^4 < \infty$ then we have

$$n^{1/2}(S_n^2 - \sigma^2) \Rightarrow \mathcal{N}(0, \mu_4 - \mu_2^2)$$

Corollary

1. $n^{1/2}(S_n^2/\sigma^2 - 1) \Rightarrow \mathcal{N}(0, \kappa + 2)$ where $\kappa = \mu_4/\mu_2^2 - 3$ is kurtosis of X
2. If $X_i \sim \mathcal{N}(0, \sigma^2)$ then $\mu_4 = 3\sigma^4$ so $n^{1/2}(S_n^2 - \sigma^2) \Rightarrow \mathcal{N}(0, 2\sigma^4)$

Confidence Interval for $\mathrm{Var}(X)$

Goal: Test $\mathrm{H}_0 : \sigma^2 \leq 1$ vs. $\mathrm{H}_1 : \sigma^2 > 1$ based on $X_1, \ldots, X_n$ iid

Normal theory approach

- If $X_i \sim \mathcal{N}(\mu, \sigma^2)$ then $nS_n^2/\sigma^2 \sim \chi^2_{n-1}$
- In worst (boundary) case $\sigma^2 = 1$ we have $nS_n^2 \sim \chi^2_{n-1}$
- Reject H_0 if $n\,S_n^2 > \chi^2_{n-1,\alpha}$ = upper $(1-\alpha)$ percentile of χ^2_{n-1} distribution

Confidence Interval for $\mathrm{Var}(X)$

Question: What is level of this test if $\mathrm{Var}(X_i) = 1$ but X_i not normal?

- If $W_n \sim \chi^2_n$ then $(W_n - n)/(2n)^{1/2} \Rightarrow \mathcal{N}(0,1)$
- This implies $(\chi^2_{n,\alpha} - n)/n^{1/2} \to \sqrt{2} z_\alpha$ and therefore

$$P_{\sigma^2=1}\left(n\, S_n^2 > \chi^2_{n-1,\alpha}\right) \to 1 - \Phi\left(\frac{\sqrt{2} z_\alpha}{\sqrt{\kappa+2}}\right)$$

Upshot

- Asymptotic level of test is α iff kurtosis κ of X is zero. As $\kappa \to \infty$ level converges to $1 - \Phi(0) = 1/2$, which is anti-conservative
- When $\mathbb{E}|X_i|^4$ finite, formulate test based on limiting distribution $n^{1/2}(S_n^2/\sigma^2 - 1) \Rightarrow \mathcal{N}(0, \kappa+2)$ plus estimate of κ

Variance Stabilizing Transformations

Variance Stabilizing Transformations

Setting

- Family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of distributions on $\mathbb{R}^d$ with $\Theta \subseteq \mathbb{R}$
- Observations $X_1, X_2, \ldots$ iid with $X_i \sim P_\theta$
- Goal: Confidence interval for θ using estimates $T_n = T_n(X_1^n)$

Common situation

- If $X_i \sim P_\theta$ then $n^{1/2}(T_n - \theta) \Rightarrow \mathcal{N}(0, s^2(\theta))$, for some function $s^2()$
- Natural CI for θ is $T_n \pm z_{\alpha/2}\, s(\theta)/\sqrt{n}$
- Problem: CI depends on unknown parameter θ

Variance Stabilizing Transformations

Recall: Estimates $T_n = T_n(X_1^n)$ such that $n^{1/2}(T_n - \theta) \Rightarrow \mathcal{N}(0, s^2(\theta))$

Idea: Variance stabilizing transformation

- Find invertible map $\phi : \Theta \to \Theta$ s.t. $n^{1/2}(\phi(T_n) - \phi(\theta)) \Rightarrow \mathcal{N}(0, \sigma^2)$, that is, asymptotic variance σ^2 is *independent of* θ
- Obtain CI $\phi(T_n) \pm z_{\alpha/2}\sigma/\sqrt{n}$ for $\phi(\theta)$ from CLT
- Obtain CI for θ by applying ϕ^{-1} to CI for $\phi(\theta)$

$$\left[\phi^{-1}\left(\phi(T_n) - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right), \phi^{-1}\left(\phi(T_n) + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right)\right]$$

Variance Stabilizing Transformations

Suppose: $\Theta = (a, b)$ and asymptotic variance $s^2(\theta) > 0$ for all $\theta \in \Theta$

- Define $\phi(\theta) := \int_a^\theta \frac{1}{s(u)} du$. Note $\phi'(\theta) = \frac{1}{s(\theta)} > 0$ for all $\theta \in (a, b)$
- Definition ensures ϕ is strictly increasing and $\phi'(\theta)s(\theta) = 1$ for all θ
- The delta method gives

$$n^{1/2}(\phi(T_n) - \phi(\theta)) \Rightarrow \mathcal{N}(0, \phi'(\theta)^2 s^2(\theta)) = \mathcal{N}(0, 1)$$

- Form CI $\phi(T_n) \pm z_{\alpha/2}/\sqrt{n}$ for $\phi(\theta)$
- CI for θ is given by

$$\left[\phi^{-1}\left(\phi(T_n) - \frac{z_{\alpha/2}}{\sqrt{n}}\right), \phi^{-1}\left(\phi(T_n) + \frac{z_{\alpha/2}}{\sqrt{n}}\right)\right]$$

Example: Sample Correlation Coefficient

Setting: Bivariate observations

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}, \begin{bmatrix} X_2 \\ Y_2 \end{bmatrix}, \ldots, \begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{R}^2 \text{ iid with } \mathbb{E}X^4, \mathbb{E}Y^4 < \infty$$

Notation: $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, and $\sigma_{X,Y} = \mathrm{Cov}(X, Y)$

Goal: Estimate the population correlation coefficient $\rho := \sigma_{X,Y} / \sigma_X \sigma_Y$

- Recall that $-1 \leq \rho \leq 1$, with equality iff $Y = aX + b$

Sample Correlation Coefficient, Consistency

Definition: The sample correlation coefficient is $\hat{r}_n := S_{X,Y}/S_X S_Y$, where

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \qquad S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$$

are the sample variances of X and Y, and the sample covariance

$$S_{X,Y} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$$

Consistency: The law of large numbers ensures that wp1

$$S_X^2 \to \sigma_X^2, \quad S_Y^2 \to \sigma_Y^2, \quad S_{X,Y} \to \sigma_{X,Y}$$

and therefore $\hat{r}_n \to \rho$ wp1 as well

Sample Correlation Coefficient, Asymptotic Normality

Outline: Begin with CLT for average of sample vectors

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i, X_i^2, Y_i, Y_i^2, X_iY_i\right)^t$$

Then apply the delta method to establish a CLT for $\hat{r}_n$

Fact: If $(X_i, Y_i)^t$ are jointly normal then $n^{1/2}(\hat{r}_n - \rho) \Rightarrow \mathcal{N}(0, (1-\rho^2)^2)$

In this case variance stabilizing transformation takes the form

$$\phi(u) \;=\; \int_0^u \frac{1}{1-u^2}\,du \;=\; \frac{1}{2}\log\frac{1+u}{1-u} \;=\; \text{tanh}^{-1}(u)$$

and we have $n^{1/2}(\phi(\hat{r}_n) - \phi(\rho)) \Rightarrow \mathcal{N}(0,1)$