Econometrics Notes

Jiahui Shui

January 15, 2025

Contents

Conventions

$\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.
$\mathbb{N}$ denotes the set $\{1, 2, 3, ...\}$ of natural numbers (excluding 0).

1 Probability and Statistics Review

What is statistics?

1.1 Probability Space

1.2 Conditional Probability and Independence

1.3 Random Variable

1.4 Convergence

Definition 1.1 (Convergence in Probability). Let $\{X_n\}, X$ be random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We say X_n converges to X in probability if, $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| \geq \varepsilon) \to 0$ as $n \to \infty$. Or equivalently,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1, \quad \forall \varepsilon > 0 \tag{1}$$

We denote it as $X_n \xrightarrow{p} X$.

Definition 1.2 (Almost Surely Convergence). Let $\{X_n\}, X$ be random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We say X_n converges to X almost surely, if

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1 \tag{2}$$

This can be written as $X_n \xrightarrow{a.s.} X$

It is easy to say that $X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{p} X$. But the opposite direction is not true. Counterexample: consider $((0,1], \mathcal{B}_{(0,1]}, \lambda)$, where λ is Lebesgue measure. Let

$$\xi_n = \mathbf{1}_{(n/2^k - 1, (n+1)/2^k - 1)}, \quad 2^k \leq n < 2^{k+1}$$

Then $\mathbb{P}(|\xi_n| > \varepsilon) \leq 1/2^k \to 0$ but $\lim \xi_n(\omega)$ does not exist for any $\omega \in (0,1]$.

Definition 1.3 (Bounded in Probability). $\{X_n\}$ is said to be bounded in probability if $\forall \varepsilon > 0$, $\exists M > 0$ such that

$$\inf_n \mathbb{P}(|X_n| \leq M) \geq 1 - \varepsilon \tag{3}$$

or equivalently, $\inf_n \mathbb{P}(|X_n| > M) < \varepsilon$

If $X_n \xrightarrow{p} 0$, the we denote it as $X_n = o_p(1)$.

If X_n is bounded in probability, then we denote it as $X_n = O_p(1)$. Moreover:

$$X_n = o_p(a_n) \Leftrightarrow \frac{X_n}{a_n} = o_p(1)$$

and

$$X_n = O_p(a_n) \Leftrightarrow \frac{X_n}{a_n} = O_p(1)$$

Exercise. Prove each of the followings:

(i) $o_p(1) + o_p(1) = o_p(1)$

(ii) $o_p(1) + O_p(1) = O_p(1)$

(iii) $o_p(1)O_p(1) = o_p(1)$

(iv) $(1 + o_p(1))^{-1} = O_p(1)$

(v) $o_p(a_n) = a_n o_p(1)$

(vi) $O_p(a_n) = a_n O_p(1)$

(vii) $o_p(O_p(1)) = o_p(1)$

Remark. (iii) is an implication of Slutsky theorem. Since $o_p(1)O_p(1) \xrightarrow{d} 0$, then it must converge to 0 in probability by proposition 1.6

A powerful theorem to prove convergence in probability:

Theorem 1.4 (Continuous Mapping Theorem). Suppose that a measurable function g is (a.s.) continuous, then

$$X_n \xrightarrow{p} X_\infty \Rightarrow g(X_n) \xrightarrow{p} g(X_\infty) \tag{4}$$

Moreover, it applies to $a.s.$ convergence and convergence in distribution.

Definition 1.5 (Convergence in Distribution). We say $X_n \xrightarrow{d} X$ if the distribution $P_n := \mathbb{P}\{X_n \in \cdot\}$ converges to $P := \mathbb{P}\{X \in \cdot\}$. Or, equivalently

$$F_{X_n}(x) \to F(x), \quad \text{for any point } x \text{ that } F(x) \text{ is continuous} \tag{5}$$

Another important result is that if for any $t \in \mathbb{R}$, the characteristic function $\phi_{X_n}(t) \to \phi_X(t)$, then $X_n \xrightarrow{d} X$. We can prove that

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X \tag{6}$$

Another important result for convergence in distribution is:

Proposition 1.6. If $X_n \xrightarrow{d} c$ where c is a constant, then $X_n \xrightarrow{p} c$.

Proof. Let $X = c$, then $F_X(x) = \mathbf{1}_{\{x \geq c\}}$. Hence, $\forall \varepsilon > 0$

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(|X_n - c| < \varepsilon) &= \lim_{n\to\infty} \mathbb{P}(c - \varepsilon < X_n < c + \varepsilon) \\ &= \lim_{n\to\infty} (F_{X_n}(c+\varepsilon) - F_{X_n}(c-\varepsilon)) \\ &= 1 - 0 = 1\end{aligned} \tag{7}$$

Then we know that $X_n \xrightarrow{p} c$. □

Lemma 1.7 (Marginal Convergence and Joint Convergence).

- If $X_n \xrightarrow{a.s.} X$, $Y_n \xrightarrow{a.s.} Y$, then $(X_n, Y_n) \xrightarrow{a.s.} (X, Y)$
- If $X_n \xrightarrow{p} X, Y_n \xrightarrow{p} Y$, then $(X_n, Y_n) \xrightarrow{p} (X, Y)$
- If $X_n \xrightarrow{d} X, Y_n \xrightarrow{d} Y$ and X_n, Y_n are independent for all n, X, Y are independent, then $(X_n, Y_n) \xrightarrow{d} (X, Y)$
- If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} c$, then $(X_n, Y_n) \xrightarrow{d} (X, c)$

Theorem 1.8 (Slutsky's Theorem). Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, where c is constant.

- $X_n + Y_n \xrightarrow{d} X + c$
- $X_n Y_n \xrightarrow{d} cX$
- $X_n / Y_n \xrightarrow{d} X/c$ if $c \neq 0$

Exercise. Let $\{X_n\}$ be independent with $X_n \sim \text{Gamma}(\alpha_n, \beta_n)$. $\alpha_n \to \alpha, \beta_n \to \beta$ for some positive real number α, β. Now, let $\hat{\beta}_n$ be a consistent estimator for β. Prove that $X_n/\hat{\beta}_n \xrightarrow{d} \text{Gamma}(\alpha, 1)$

Theorem 1.9 (Delta Method). First order expansion: Suppose that g is differentiable at $\boldsymbol{c}$, for any sequence $0 < a_n \to \infty$, we have

$$a_n(\boldsymbol{X}_n - \boldsymbol{c}) \xrightarrow{d} \boldsymbol{X} \Rightarrow a_n[g(\boldsymbol{X}_n) - g(\boldsymbol{c})] \xrightarrow{d} [\nabla g(\boldsymbol{c})]^\top \boldsymbol{X} \tag{8}$$

Theorem 1.10 (Prohorov's Theorem). If $X_n \xrightarrow{d} X$, then $X_n = O_p(1)$

Proof. $\forall \varepsilon > 0$, we can choose M_0 sufficiently large such that

$$\mathbb{P}(|X| > M_0) < \varepsilon) \tag{9}$$

Then, since $\mathbb{P}(|X_n| > M_0) \to \mathbb{P}(|X| > M_0)$, then we can choose n_0 such that for all $n \geq n_0$, $\mathbb{P}(|X_n| > M_0) < \varepsilon$. Now, we can select M_1 such that

$$\mathbb{P}(|X_i| > M_1) < \varepsilon, \quad \forall i = 1, \cdots, n_0 - 1 \tag{10}$$

Then let $M = \max(M_0, M_1)$ we have $\mathbb{P}(|X_n| > M) < \varepsilon$ for all n. □

1.5 Law of Large Numbers

Theorem 1.11 (WLLN, Khintchin). If $\{X_n\}$ are i.i.d with $\mathbb{E}[X_1] = \mu < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu \tag{11}$$

Theorem 1.12 (SLLN, Kolmogorov). If $\{X_n\}$ are i.i.d with $\mathbb{E}[X_1] = \mu < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu \tag{12}$$

1.6 Central Limit Theorem

Theorem 1.13 (Levy CLT). Suppose that $\{X_n\}$ i.i.d with mean μ and variance σ^2, then

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2) \tag{13}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1.7 Normal Distribution

Consider multivariate normal distribution: $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. The moment generating function is

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[e^{\boldsymbol{t'X}}] = e^{\boldsymbol{t'\mu} + \frac{1}{2}\boldsymbol{t'\Sigma t}}, \quad \boldsymbol{t} \in \mathbb{R}^d \tag{14}$$

For any $\boldsymbol{A} \in \mathbb{R}^{m\times d}$, we have $\boldsymbol{AX} + \boldsymbol{b} \sim N(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A'})$. The probability density function of $\boldsymbol{X}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{(x-\mu)'\Sigma^{-1}(x-\mu)}\right\}, \quad \boldsymbol{x} \in \mathbb{R}^d \tag{15}$$

- If $X_1, \cdots, X_n \sim N(0,1)$ i.i.d, then $X_1^2 + \cdots + X_n^2 \sim \chi_n^2$.
- If $X \sim N(0,1)$ and $Q \sim \chi_n^2$ are independent, then $\frac{X}{\sqrt{Q/n}} \sim t_n$
- If $Q_1 \sim \chi_m^2, Q_2 \sim \chi_n^2$ are independent, then $\frac{Q_1/m}{Q_2/n} \sim F_{m,n}$

1.8 Hypothesis Testing

Consider $H_0 : \theta \in \Theta$, this is called the null hypothesis. The alternative hypothesis is $H_1 : \theta \in \Theta_1$, where $\Theta_1 = \Theta \backslash \Theta_0$. We have to decide between H_0 and H_1. Let R denotes the reject region. A Test might have two types of mistake.

- **Type I Error**: Reject H_0 when $\theta \in \Theta_0$. $\mathbb{P}_\theta(\boldsymbol{X} \in R)$ for $\theta \in \Theta_0$
- **Type II Error**: Accept H_0 when $\theta \in \Theta_1$. $\mathbb{P}_\theta(\boldsymbol{X} \in R^c)$ for $\theta \in \Theta_1$

In this notes, we will use φ to denote *power function* for a hypothesis test, i.e.

$$\varphi(\theta) = \mathbb{P}_\theta(\boldsymbol{X} \in R) \tag{16}$$

When $\theta \in \Theta_0$, then $\varphi(\theta) = \mathbb{P}(\text{Type I Error})$. If $\theta \in \Theta_1$, then $\varphi(\theta) = 1 - \mathbb{P}(\text{Type II Error})$

Definition 1.14. For $\alpha \in [0,1]$, a test with power function $\varphi(\theta)$ is a *size* α test if

$$\sup_{\theta\in\Theta_0} \varphi(\theta) = \alpha,$$

is a *level* α test if

$$\sup_{\theta\in\Theta_0} \varphi(\theta) \le \alpha$$

Definition 1.15. A test is *unbiased* if $\varphi(\theta') \ge \varphi(\theta'')$ for all $\theta' \in \Theta_1$ and $\theta'' \in \Theta_0$. A test is *consistent* if

$$\lim_{n\to\infty} \varphi(\theta) = 1, \quad \forall \theta \in \Theta_1$$

Definition 1.16. Let $\mathcal{C}$ be a class of tests. A test in class $\mathcal{C}$ with power function $\varphi(\theta)$ is a *uniformly most powerful* (UMP) class $\mathcal{C}$ test if $\varphi(\theta) \ge g(\theta)$ for every $\theta \in \Theta_1$ and every $g(\theta)$ that is a power function of a test in class $\mathcal{C}$. Generally we take $\mathcal{C}$ as the class of all level α test.

Theorem 1.17 (Neyman-Pearson Lemma). Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf or pmf corresponding to θ_i is $f(x|\theta_i), i = 0, 1$. Then

$$R = \{x : f(x|\theta_1) > kf(x|\theta)\}$$

for some $k \ge 0$ and $\alpha = \mathbb{P}_{\theta_0}(X \in R)$ is a UMP level α test.

Definition 1.18 (p-value). A p-value $p(X)$ is a test statistic satisfying $0 \le p(x) \le 1$ for every sample point x. Small values of $p(X)$ give evidence that H_1 is true. A p-value is valid if, for every $\theta \in \Theta_0$ and every $0 \le \alpha \le 1$,

$$\mathbb{P}_\theta(p(X) \le \alpha) \le \alpha \tag{17}$$

If $p(X)$ is a valid p-value, then it is easy to construct a level α test based on this statistics. We rejects H_0 if and only if $p(X) \leq \alpha$.

Theorem 1.19. Suppose that $T(X)$ is a test statistic such that large values of T give evidence that H_1 is true. For each sample point x, define

$$p(x) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T(X) \geq T(x)) \tag{18}$$

Then $p(X)$ is a valid p-value.

Now, suppose that β is a parameter, and $\hat{\beta}$ is an estimator of β. Moreover, we assume that $\hat{\beta}$ is consistent and asymptotically normal, i.e.

$$\hat{\beta} \xrightarrow{p} \beta, \quad \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2)$$

Also, suppose that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ is an estimator for asymptotical variance. Now we want to test the hypothesis: $H_0 : \beta = c$, where c is a constant. To this end, we can employ t-test:

$$T = \frac{\hat{\beta} - c}{\mathrm{se}(\hat{\beta})} = \frac{\hat{\beta} - c}{\hat{\sigma}/\sqrt{n}} \tag{19}$$

Then under the null, T is asymptotically normal since by Slutsky's theorem, we have

$$\frac{\hat{\beta} - c}{\hat{\sigma}/\sqrt{n}} = \frac{1}{\hat{\sigma}}\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, 1) \tag{20}$$

Then let the rejection region be

$$R := \{|T| \geq z_{\alpha/2}\}, \quad z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2)$$

where $\Phi(x)$ is standard normal cdf. Also, we can construct confidence interval by test inversion:

$$\mathrm{CI}_{1-\alpha} = \left[\hat{\beta} - z_{\alpha/2}\mathrm{se}(\hat{\beta}), \hat{\beta} + z_{\alpha/2}\mathrm{se}(\hat{\beta})\right] \tag{21}$$

1.9 Wald's Test