

Econometrics Notes

Jiahui Shui

January 18, 2025

Contents

1	Probability and Statistics Review	2
1.1	Probability Space	2
1.2	Conditional Probability and Independence	2
1.3	Random Variable	2
1.4	Convergence	2
1.5	Law of Large Numbers	4
1.6	Central Limit Theorem	4
1.7	Normal Distribution	4
1.8	Hypothesis Testing	4
1.9	Wald's Test	5
2	Causal Inference and Potential Outcomes	5
2.1	Potential Outcomes Model	5
2.2	Treatments	5

Conventions

\mathbb{F} denotes either \mathbb{R} or \mathbb{C} .
 \mathbb{N} denotes the set $\{1, 2, 3, \dots\}$ 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, $\forall \varepsilon > 0, \mathbb{P}(|X_n - X| \geq \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$. Or equivalently,

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

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

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1 \quad (2)$$

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 $\mathbb{P}(|\xi_n| > \varepsilon) \leq 1/2^k \rightarrow 0$ 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

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

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

If $X_n \xrightarrow{p} 0$, then 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

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

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

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

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

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X \quad (6)$$

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, $\forall \varepsilon > 0$

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

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 $X_n \sim \text{Gamma}(\alpha_n, \beta_n)$. $\alpha_n \rightarrow \alpha, \beta_n \rightarrow \beta$ 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 \mathbf{c} , for any sequence $0 < a_n \rightarrow \infty$, we have

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

If $\nabla g(\mathbf{c}) = 0$, then we have similar expansion: [To be completed](#)

Example. Suppose that $\{X_i\}$ i.i.d with mean μ and variance σ^2 . Consider the following estimator:

$$\hat{\theta} = \bar{x}^2 \quad (9)$$

Let $\theta := \mu^2$. (1) Find the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ provided $\mu \neq 0$. (2) If $\mu = 0$, find the convergence rate of $\hat{\theta}$ and its limit distribution.

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

$$\mathbb{P}(|X| > M_0) < \varepsilon \quad (10)$$

Then, since $\mathbb{P}(|X_n| > M_0) \rightarrow \mathbb{P}(|X| > M_0)$, 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

$$\mathbb{P}(|X_i| > M_1) < \varepsilon, \quad \forall i = 1, \dots, n_0 - 1 \quad (11)$$

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

A natural question is: will bounded in probability imply convergence in distribution? The answer is **No**. Consider $X_n = 2 + 1/n$ for even n and $X_n = 1 + 1/(n+1)$ for odd n . Then the sequence (X_{2k}) converges in distribution to $Y = 2$. And (X_{2k-1}) converges in distribution to $W = 1$. Since $Y \neq W$ then the sequence does not converge in distribution. Since all X_n lie in the interval $[1, 5/2]$, then we can easily show that $X_n = O_p(1)$.

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

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu \quad (12)$$

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

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{a.s.} \mu \quad (13)$$

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

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2) \quad (14)$$

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

1.7 Normal Distribution

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

$$M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[e^{\mathbf{t}'\mathbf{X}}] = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^d \quad (15)$$

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

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad \mathbf{x} \in \mathbb{R}^d \quad (16)$$

- If $X_1, \dots, X_n \sim N(0, 1)$ i.i.d, then $X_1^2 + \dots + X_n^2 \sim \chi_n^2$.
- 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 $\Theta_1 = \Theta \setminus \Theta_0$. 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 $\theta \in \Theta_0$. $\mathbb{P}_{\theta}(\mathbf{X} \in R)$ for $\theta \in \Theta_0$
- **Type II Error:** Accept H_0 when $\theta \in \Theta_1$. $\mathbb{P}_{\theta}(\mathbf{X} \in R^c)$ for $\theta \in \Theta_1$

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

$$\beta(\theta) = \mathbb{P}_{\theta}(\mathbf{X} \in R) \quad (17)$$

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

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

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

is a *level α test* if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

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

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

Definition 1.16. Let \mathcal{C} be a class of tests. A test in class \mathcal{C} with power function $\beta(\theta)$ is a *uniformly most powerful* (UMP) class \mathcal{C} test if $\beta(\theta) \geq g(\theta)$ 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

$$R = \{x : f(x|\theta_1) > k f(x|\theta_0)\}$$

for some $k \geq 0$ 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 $0 \leq p(x) \leq 1$ 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 $0 \leq \alpha \leq 1$,

$$\mathbb{P}_{\theta}(p(X) \leq \alpha) \leq \alpha \quad (18)$$

If $p(X)$ is a valid *p-value*, then it is easy to construct a level α test based on this statistics. We reject 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

$$p(x) = \sup_{\theta \in \Theta_0} \mathbb{P}_{\theta}(T(X) \geq T(x)) \quad (19)$$

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*:

$$T = \frac{\hat{\beta} - c}{\text{se}(\hat{\beta})} = \frac{\hat{\beta} - c}{\hat{\sigma}/\sqrt{n}} \quad (20)$$

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

$$\frac{\hat{\beta} - c}{\hat{\sigma}/\sqrt{n}} = \frac{1}{\hat{\sigma}} \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, 1) \quad (21)$$

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:

$$\text{CI}_{1-\alpha} = [\hat{\beta} - z_{\alpha/2} \text{se}(\hat{\beta}), \hat{\beta} + z_{\alpha/2} \text{se}(\hat{\beta})] \quad (22)$$

Exercise. To be completed.

1.9 Wald's Test

2 Causal Inference and Potential Outcomes

2.1 Potential Outcomes Model

We denote x_i the treatment status, i.e. $x_i = 1$ implies the individual is treated and $x_i = 0$ indicates the individual is not treated. Also, the out come $y_i(1)$ is "the outcome of the i^{th} individual had she received the treatment. Therefore, for each i , we have three random variables, $(x_i, y_i(1), y_i(0))$.

Definition 2.1. Individual treatment effect (ITE):

$$\tau_i = y_i(1) - y_i(0)$$

Average treatment effect (ATE):

$$\tau_{\text{ATE}} = \mathbb{E}[\tau_i] = \mathbb{E}[y_i(1) - y_i(0)]$$

Treatment effect on the treated (ATT):

$$\tau_{\text{ATT}} = \mathbb{E}[\tau_i | x_i = 1] = \mathbb{E}[y_i(1) - y_i(0) | x_i = 1]$$

Treatment effect on the untreated (ATU):

$$\tau_{\text{ATU}} = \mathbb{E}[\tau_i | x_i = 0] = \mathbb{E}[y_i(1) - y_i(0) | x_i = 0]$$

We can always write y_i as

$$y_i = y_i(1) \mathbb{1}_{\{x_i=1\}} + y_i(0) \mathbb{1}_{\{x_i=0\}} = x_i y_i(1) + (1 - x_i) y_i(0) \quad (23)$$

In words, we will never be able to observe the two potential outcomes simultaneously for any individual.

2.2 Treatments