高级计量经济学

Lecture 5: Statistical Properties of OLS (1)

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼1510
E-mail huangjp@szu.edu.cn
Website https://huangjp.com

条件概率

联合概率分布

Joint Probability Distribution

- 随机变量 X 和 Y 的联合概率分布和联合密度函数

$$F(x, y) = \Pr(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(\xi, \zeta)\, d\xi\, d\zeta$$

- X 和 Y 的边际分布和边际密度函数

$$F_X(x) = F(x, \infty) = \int_{-\infty}^{x} f_X(\xi)\, d\xi, \quad f_X(x) = \int_{-\infty}^{\infty} f(x, \zeta)\, d\zeta$$

$$F_Y(y) = F(\infty, y) = \int_{-\infty}^{y} f_Y(\zeta)\, d\zeta, \quad f_Y(y) = \int_{-\infty}^{\infty} f(\xi, y)\, d\xi$$

- X 和 Y 独立（$X \perp\!\!\!\perp Y$）

$$F(x, y) = F_X(x)\, F_Y(y)$$ 若密度函数存在，则 $f(x, y) = f_X(x) f_Y(y)$

期望值、方差与协方差

Expectation, Variance and Covariance

- 边际分布的期望值与方差

$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\, dx\, dy = \mu_X$$

$$\begin{aligned}\mathrm{Var}[X] &= E[(X - E[X])^2] \\ &= \int_{-\infty}^{\infty} (x - E[X])^2\, f_X(x)\, dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - E[X])^2\, f(x,y)\, dx\, dy = E[X^2] - \mu_X^2 = \sigma_X^2\end{aligned}$$

- 协方差

$$\begin{aligned}\mathrm{Cov}[X,Y] &= E[(X - E[X])(Y - E[Y])] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - E[X])(y - E[Y])\, f(x,y)\, dx\, dy = E[XY] - \mu_X\mu_Y = \sigma_{XY}\end{aligned}$$

- Pearson 相关系数（线性相关系数）

$$r[X,Y] = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

X 与 Y 线性不相关 $\Leftrightarrow \rho_{XY} = 0$
$\Leftrightarrow \mathrm{Cov}[X,Y] = 0$
$\Leftrightarrow E[XY] = E[X]E[Y]$

独立性与相关性

$X \perp\!\!\!\perp Y \;\Rightarrow\; \rho_{XY} = 0$

证明：

$$\begin{aligned}
\sigma_{XY} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}(x-\mu_X)f_X(x)\,dx\int_{-\infty}^{\infty}(y-\mu_Y)f_Y(y)\,dy \\
&= 0
\end{aligned}
\qquad \Rightarrow\quad \rho_{XY} = 0$$

$X \perp\!\!\!\perp Y \;\not\Leftarrow\; \rho_{XY} = 0$

证明：令 $f(-1,1) = f(0,0) = f(1,1) = \frac{1}{3}$。则 $\mu_X = 0,\ \mu_Y = \frac{2}{3},\ \sigma_{XY} = 0$。但是 $f_X(1)f_Y(1) = \frac{1}{3}\frac{2}{3} = \frac{2}{9} \neq \frac{1}{3} = f(1,1)$。

$\rho_{XY} = 0,\ (X,Y) \sim N \;\Rightarrow\; X \perp\!\!\!\perp Y$

多变量分布
Multivariate Distribution

$\boldsymbol{x} = (X_1, X_2, \ldots, X_n)^\top$ 为随机向量，其联合密度函数为 $f(x_1, x_2, \ldots, x_n)$

- 期望值

$$E[\boldsymbol{x}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_n} \end{bmatrix} = \boldsymbol{\mu}$$

- 方差-协方差矩阵

$$\mathrm{Var}[\boldsymbol{x}] = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix} = E[\boldsymbol{x}\boldsymbol{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

条件概率

Conditional Probability

Y 关于 X 的条件概率可以表达为

$$\Pr(Y \le y \mid X = x) = \frac{\Pr(Y \le y, X = x)}{\Pr(X = x)}$$

然而，因为存在 $\Pr(X = x) = 0$ 的情况，以上定义无效。我们只能用密度函数定义条件概率分布的密度，即

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)}, \quad f(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

当 X 和 Y 独立时

$$f(y \mid x) = f_Y(y), \quad f(x \mid y) = f_X(x)$$

条件期望与条件方差

- 条件期望

$$E[Y \mid X] = \int_{-\infty}^{\infty} y f(y \mid x)\, dy = \mu_{Y|X}$$

- 条件方差

$$\begin{aligned}\mathrm{Var}[Y \mid X] &= \int_{-\infty}^{\infty} (y - E[y \mid x])^2 f(y \mid x)\, dy \\ &= E[Y^2 \mid X] - \mu_{Y|X}^2\end{aligned}$$

条件方差函数也被称为 skedastic function

迭代期望法则
Law of Iterated Expectation (LIE)

$$E[Y] = E_X\big[E[Y \mid X]\big]$$

证明：

$$
\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(y \mid x) f_X(x)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y f(y \mid x)\, dy\right) f_X(x)\, dx \\
&= E_X\big[E[Y \mid X]\big]
\end{aligned}
$$

OLS 假设

线性模型与 OLS 假设

independently and identically distributed
在给出 $\boldsymbol{X}$ 的条件下，u_t 服从独立同分布，
其均值为 0，方差为 σ^2，任意 u_s 与 u_t 的协方差为 0。

完整的线性回归模型可以写成

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{u} \mid \boldsymbol{X} \sim \mathrm{IID}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

这个简洁的表述中包含以下假设：

1. 线性 linearity：$\boldsymbol{y}$ 是 $\boldsymbol{\beta}$ 的线性函数
2. 解释变量 $\boldsymbol{X}$ 可以是随机的，此时 $(y_t, x_{t1}, \ldots, x_{tk})$ 为同一个联合分布的独立样本
3. 外生性 exogeneity：$E[u_t \mid \boldsymbol{X}] = 0$ $\Rightarrow E[\boldsymbol{y} \mid \boldsymbol{X}] = \boldsymbol{X}\boldsymbol{\beta}$
4. 同方差性 homoskedasticity：$\mathrm{Var}[u_t \mid \boldsymbol{X}] = \sigma^2$
   无自相关 non-autocorrelation：$\mathrm{Cov}[u_s, u_t \mid \boldsymbol{X}] = 0$

$\Rightarrow$ 在外生性成立时，
$\mathrm{Var}[u_t \mid \boldsymbol{X}] = \sigma^2 \Leftrightarrow E[u_t^2 \mid \boldsymbol{X}] = \sigma^2$
$\mathrm{Cov}[u_s, u_t \mid \boldsymbol{X}] = 0 \Leftrightarrow E[u_s u_t \mid \boldsymbol{X}] = 0$

右侧两个条件可以简洁地写成 $E[\boldsymbol{u}\boldsymbol{u}^\top | \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$

为了保证 OLS 估计量 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ 存在，我们还需要假设：

5. $\boldsymbol{X}$ 列满秩：$\mathrm{rank}(\boldsymbol{X}) = k$

外生性条件

The Exogeneity Condition

$E[u_t \mid \boldsymbol{X}] = 0$ 被称为严格外生性（strict exogeneity），在横截面数据中容易被满足，但在时间序列数据中往往过强。一个较弱的条件是前定性（predeterminedness）$E[u_t \mid \boldsymbol{X}_t] = 0$。

由严格外生性可以导出：

1. $E[u_t] = E_{\boldsymbol{X}}\left[E[u_t \mid \boldsymbol{X}]\right] = 0$
2. $E[\boldsymbol{u}\boldsymbol{u}^\top] = E_{\boldsymbol{X}}\left[E[\boldsymbol{u}\boldsymbol{u}^\top \mid \boldsymbol{X}]\right] = E[\sigma^2\boldsymbol{I}] = \sigma^2\boldsymbol{I}$

此处假设同方差性和无自相关

因此，$\boldsymbol{X}$ 为随机变量时的结论也适用于 $\boldsymbol{X}$ 为固定变量时。

（条件）均值独立

(Conditional) Mean Independence

均值独立性

当 $E[Y \mid X] = E[Y]$ 时，称 Y 均值独立于 X。

由定义可知，均值独立性不是对称的概念， Y 均值独立于 X 不等于 X 均值独立于 Y。

定理：

X 与 Y 独立 $\underset{\nLeftarrow}{\Rightarrow}$ X 与 Y 均值独立 $\underset{\nLeftarrow}{\Rightarrow}$ X 与 Y 线性不相关

$X \perp\!\!\!\perp Y$

$E[X \mid Y] = E[X]$,
$E[Y \mid X] = E[Y]$

$\mathrm{Cov}[X, Y] = 0 \;\Leftrightarrow\; E[XY] = E[X]E[Y]$

详细信息可参考 https://www.econometrics.blog/post/why-econometrics-is-confusing-part-ii-the-independence-zoo/

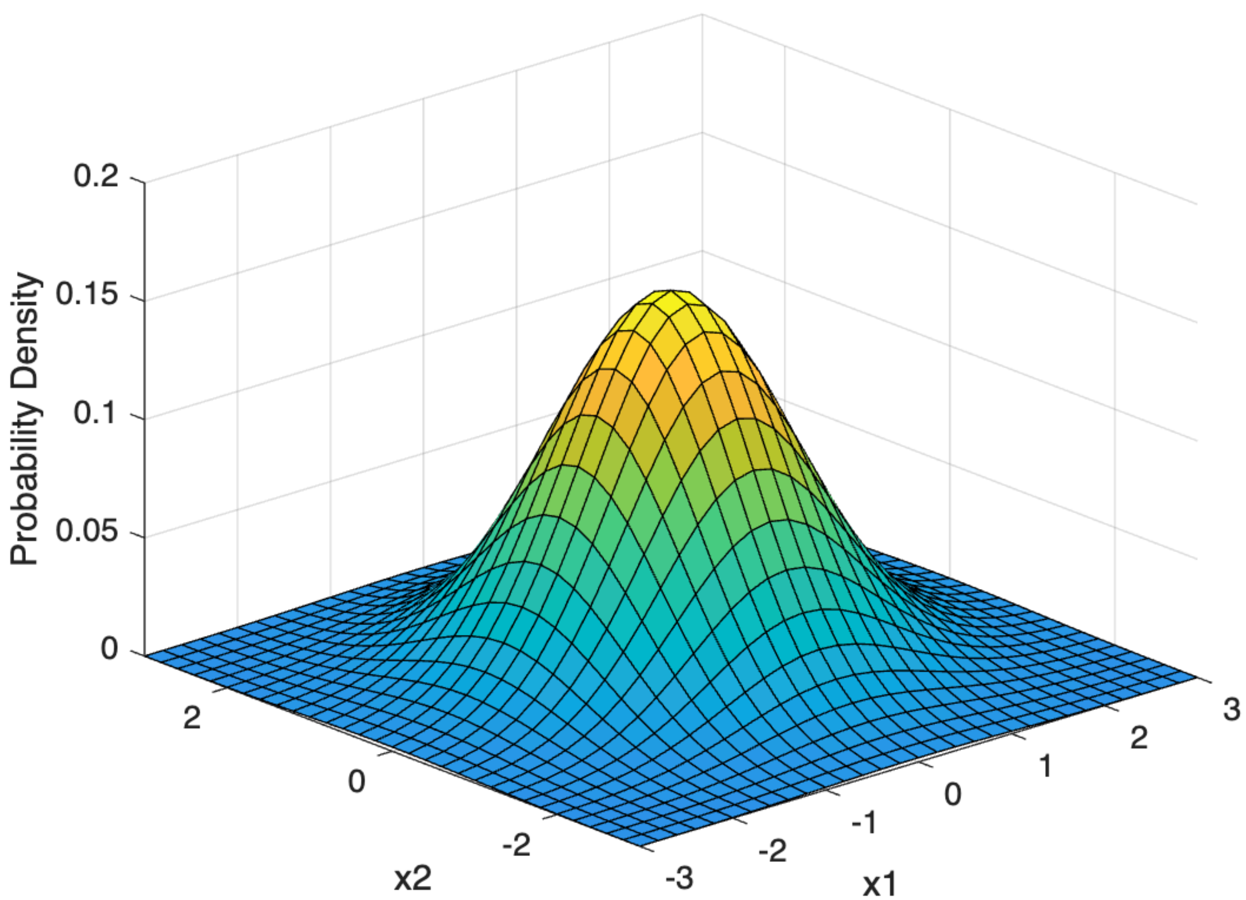
若严格外生性 $E[u_t \mid \boldsymbol{X}] = 0$ 成立，则误差项 $\boldsymbol{u}$ 均值独立于解释变量 $\boldsymbol{X}$。

若 $\boldsymbol{X} \perp\!\!\!\perp \mu_t$，则 $E[u_t \mid \boldsymbol{X}] = E[u_t]$。此时可以直接假设 $E[u_t] = 0$。

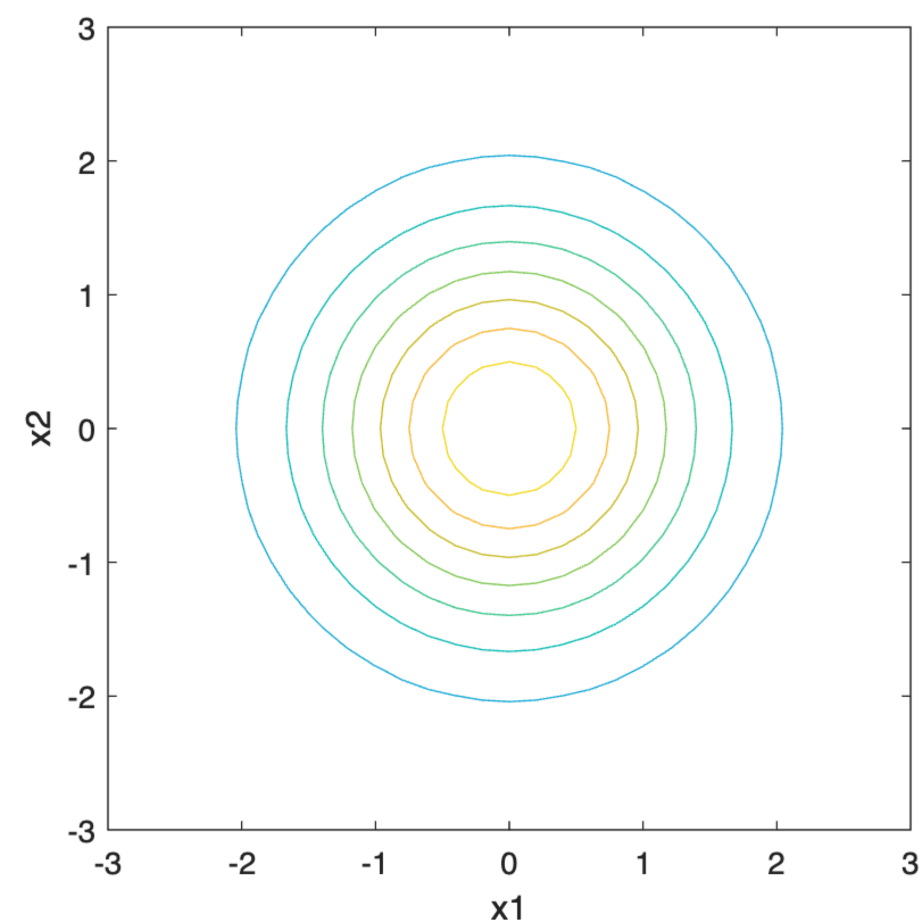
同方差无自相关性条件

The Homoskedasticity & Non-autocorrelation Condition

$\mathrm{Var}[\boldsymbol{u} \mid \boldsymbol{X}] = \sigma^2 \boldsymbol{I}$ 常被称为 spherical error，或称误差项服从 spherical distribution（球面分布）。



pdf of bivariate $N(\mathbf{0}, \boldsymbol{I})$



contour plot

2D → 3D: circle → sphere

$\hat{\beta}$ 的非偏性

非偏性

Unbiasedness

非偏性

若参数 θ 的真实值是 θ_0，$\hat{\theta}$ 是 θ 的一个估计量。则称 $E[\hat{\theta}] - \theta_0$ 为估计量 $\hat{\theta}$ 的偏差（bias）或偏误。当 $E[\hat{\theta}] - \theta_0 = 0$，即

$$E[\hat{\theta}] = \theta_0$$

时，$\hat{\theta}$ 是非偏的。

假设随机样本 $(x_1, x_2, \ldots, x_n)$ 为 IID，且总体均值为 $\mu = \mu_0$。μ 的估计量可以是

$$\hat{\mu}_{\mathrm{m}} = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad \hat{\mu}_1 = x_1$$

由此可推出 $E[\hat{\mu}_{\mathrm{m}}] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n} n\mu_0 = \mu_0$，$E[\hat{\mu}_1] = E[x_1] = \mu_0$。因此 $\hat{\mu}_{\mathrm{m}}$ 和 $\hat{\mu}_1$ 都是 μ 的非偏估计量。

$\hat{\boldsymbol{\beta}}$ 的非偏性

假设 $\boldsymbol{\beta}$ 的真实值是 $\boldsymbol{\beta}_0$。$\boldsymbol{\beta}$ 的 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 为

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}\\ &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u})\\ &= \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}\\ \Rightarrow \quad E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta}_0 + E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}]\end{aligned}$$

根据迭代期望法则，右侧第二项可以写成

$$\begin{aligned}E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}] &= E_X\big[E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u} \mid \boldsymbol{X}]\big]\\ &= E_X\big[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top E[\boldsymbol{u} \mid \boldsymbol{X}]\big]\end{aligned}$$

因此，当外生性条件 $E[\boldsymbol{u} \mid \boldsymbol{X}] = \boldsymbol{0}$ 成立时，

$$E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}] = E_X[\boldsymbol{0}] = \boldsymbol{0}$$

$$\Rightarrow \quad E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}_0 \qquad (\hat{\boldsymbol{\beta}} \text{ 是 } \boldsymbol{\beta} \text{ 的非偏估计量})$$

时间序列数据往往不满足外生性，而只满足前定性。此时 $\hat{\boldsymbol{\beta}}$ 有偏。详见 D&M (2021) pp.90-92

$\hat{\beta}$ 的一致性

一致性

Consistency

概率极限 (probability limit)

$\boldsymbol{x}^n$ 为长度为 n 的随机向量，$\boldsymbol{a}(\boldsymbol{x}^n)$ 是 $\boldsymbol{x}^n$ 的向量值函数。当 $n \to \infty$ 时 $\boldsymbol{a}(\boldsymbol{x}^n)$ 依概率收敛 (converge in probability) 于 $\boldsymbol{a}_0$ 的定义是：对任意的 $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr(\|\boldsymbol{a}(\boldsymbol{x}^n) - \boldsymbol{a}_0\| < \varepsilon) = 1$$

$\boldsymbol{a}_0$ 可以是随机或非随机向量

此时我们称 $\boldsymbol{a}_0$ 为 $\boldsymbol{a}(\boldsymbol{x}^n)$ 的概率极限，并记为 $\operatorname*{plim}_{n\to\infty} \boldsymbol{a}(\boldsymbol{x}^n) = \boldsymbol{a}_0$ 。

估计量的一致性

若参数 θ 的真实值是 θ_0，$\hat{\theta}^n$ 是基于含有 n 个观测值样本的估计量。则当

$$\operatorname*{plim}_{n\to\infty} \hat{\theta}^n = \theta_0$$

时，$\hat{\theta}^n$ 是一致的。

弱大数法则（弱大数定律）

Weak Law of Large Numbers

$(x_1, x_2, \ldots, x_n)$ 为独立随机样本，每个 x_i 的总体均值为 $E[x_i]=\mu$，总体方差为 $\mathrm{Var}[x_i]=\sigma^2$。此时，样本均值 $\bar{x}$ 是 μ 的一致估计量。

总体方差有限的条件可替换为样本服从同分布

证明：

$$E[\bar{x}] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

$$\begin{aligned}\mathrm{Var}[\bar{x}] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] \\ &= \sum_{i=1}^{n}\mathrm{Var}\left[\frac{1}{n}x_i\right] \qquad \leftarrow \text{独立性} \\ &= \sum_{i=1}^{n}\frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}\end{aligned}$$

由 Chebyshev 不等式 $\Pr(|X-E[X]| \geq k) \leq \dfrac{\mathrm{Var}[X]}{k^2}$ 可得

$$\Pr(|\bar{x}-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \Rightarrow \quad \lim_{n\to\infty}\Pr(|\bar{x}-\mu| \geq \varepsilon) = 0$$

$$\Leftrightarrow \quad \lim_{n\to\infty}\Pr(|\bar{x}-\mu| < \varepsilon) = 1$$

非偏性与一致性

$(x_1, x_2, \ldots, x_n)$ 为独立随机样本，且每个 x_i 的总体均值为 $E[x_i] = \mu$。

已知在适当的条件下（例如总体方差有限，或服从同分布），样本均值 $\bar{x}$ 是 μ 的非偏统计量和一致估计量（大数法则）。

考虑下面三个 μ 的估计量：

$$\hat{\mu}_1 = \frac{1}{n+1}\sum_{i=1}^{n} x_i \qquad = \frac{n}{n+1}\bar{x} \xrightarrow{p} \mu$$

$$\hat{\mu}_2 = \frac{1.01}{n}\sum_{i=1}^{n} x_i \qquad = 1.01\bar{x} \xrightarrow{p} 1.01\mu$$

$$\hat{\mu}_3 = 0.01x_1 + \frac{0.99}{n-1}\sum_{i=2}^{n} x_i \qquad = 0.01x_1 + 0.99\bar{x} \xrightarrow{p} 0.01x_1 + 0.99\mu$$

其中仅 $\hat{\mu}_1$ 满足一致性，仅 $\hat{\mu}_3$ 满足非偏性。

$\hat{\boldsymbol{\beta}}$ 的一致性

已知 $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}$

我们的目标是把右侧第二项的概率极限分解为

$$\left(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\left(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u}\right)$$

$\boldsymbol{X}^\top \boldsymbol{X}$ 是 $k \times k$ 矩阵，其 (i, j) 要素是

$$\boldsymbol{x}_i^\top \boldsymbol{x}_j = \sum_{t=1}^{n} x_{ti}x_{tj}$$

一般情况下不存在 $\operatorname*{plim}_{n\to\infty} \boldsymbol{x}_i^\top \boldsymbol{x}_j$，但有可能存在 $\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{x}_i^\top \boldsymbol{x}_j$。

因为 $\boldsymbol{X}^\top \boldsymbol{X} = \sum_{i=t}^{n} \boldsymbol{X}_t^\top \boldsymbol{X}_t$，则根据大数法则，$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t^\top \boldsymbol{X}_t = E[\boldsymbol{X}_t^\top \boldsymbol{X}_t]$，因此

需要假设 $\boldsymbol{X}_t$ 是 IID 的随机样本

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X} = E[\boldsymbol{X}_t^\top \boldsymbol{X}_t] \equiv \boldsymbol{Q} \quad \Rightarrow \quad \operatorname*{plim}_{n\to\infty} \frac{1}{n}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \boldsymbol{Q}^{-1}$$

Theorem D.14 (Greene, 2020)
积的概率极限 = 概率极限之积，即

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{X}^n = \boldsymbol{A}, \operatorname*{plim}_{n\to\infty} \boldsymbol{Y}^n = \boldsymbol{B} \quad \Rightarrow \quad \operatorname*{plim}_{n\to\infty} \boldsymbol{X}^n\boldsymbol{Y}^n = \boldsymbol{AB}$$

Theorem D.14 (Greene, 2020)
逆矩阵的概率极限

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{W}^n = \boldsymbol{\Omega} \quad \Rightarrow \quad \operatorname*{plim}_{n\to\infty} (\boldsymbol{W}^n)^{-1} = \boldsymbol{\Omega}^{-1}$$

书中将 $\boldsymbol{Q}$ 记为 $\boldsymbol{S}_{X^\top X}$

$\hat{\beta}$ 的一致性

接下来我们讨论 $\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}$。

因为 $\boldsymbol{X}^\top \boldsymbol{u} = \sum_{i=t}^{n} \boldsymbol{X}_t^\top u_t$，则根据大数法则，$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{X}_t^\top u_t = E[\boldsymbol{X}_t^\top u_t]$，因此

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u} = E[\boldsymbol{X}_t^\top u_t]$$

如果假设前定性 $E[u_t \mid \boldsymbol{X}_t] = 0$，则根据 LIE，

$$\begin{aligned} E[\boldsymbol{X}_t^\top u_t] &= E_{\boldsymbol{X}_t}\left[E[\boldsymbol{X}_t^\top u_t \mid \boldsymbol{X}_t]\right] = E_{\boldsymbol{X}_t}\left[\boldsymbol{X}_t^\top \cdot E[u_t \mid \boldsymbol{X}_t]\right] \\ &= E_{\boldsymbol{X}_t}[\boldsymbol{X}_t^\top \cdot 0] = \boldsymbol{0} \end{aligned}$$

因此，$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u} = E[\boldsymbol{X}_t^\top u_t] = \boldsymbol{0}$。

当样本为 IID，且满足**前定性**条件时（不需要外生性条件），$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u}$ 右侧第二项的概率极限为 $\boldsymbol{0}$，即 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的一致估计量。

关于非偏性与一致性的一些讨论

- 非偏性适用于任何大小的样本，而一致性仅适用于大样本。
- 保证 $\hat{\boldsymbol{\beta}}$ 的非偏性需要假设外生性，而保证 $\hat{\boldsymbol{\beta}}$ 的一致性只需要假设前定性。虽然前定性条件弱于外生性条件，但这并不代表非偏性可以推导出一致性，它们是两个不同的概念。非偏性只和期望值有关，而一致性是分布上的性质。
- 解释变量中包含被解释变量的滞后项（lag）的时间序列模型不满足外生性，但满足前定性。因此其 OLS 估计量有偏但一致。
- 非偏性并不一定是最好的性质。有时我们需要在非偏但方差很大的估计量和略偏但方差很小的估计量中进行选择。
- 不满足一致性有两种情况：一种是概率极限不存在（不收敛），另一种是概率极限存在但不收敛于真实的参数值。