高级计量经济学

Lecture 8:
Large Sample Tests, Confidence Intervals, HCCME

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

大样本检验

精确检验的条件

我们把精确检验（t 检验和 F 检验）所需的条件总结如下：

- $\boldsymbol{X}$ 与 $\boldsymbol{u}$ 独立
- $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

如果以上条件不能被满足，就无法获得 t 统计量或 F 统计量的精确分布。

在大样本（$n \to \infty$）下，我们可以获得检验统计量的渐进分布。

中心极限定理

Central Limit Theorem (CLT)

大数法则是一个重要的渐进分析工具，另一个就是中心极限定理：

$(x_1, x_2, \ldots, x_n)$ 为 IID 随机样本，其分布的均值和方差分别为 μ 和 σ^2。当 $n \to \infty$ 时，

$$z \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{x_t - \mu}{\sigma}$$

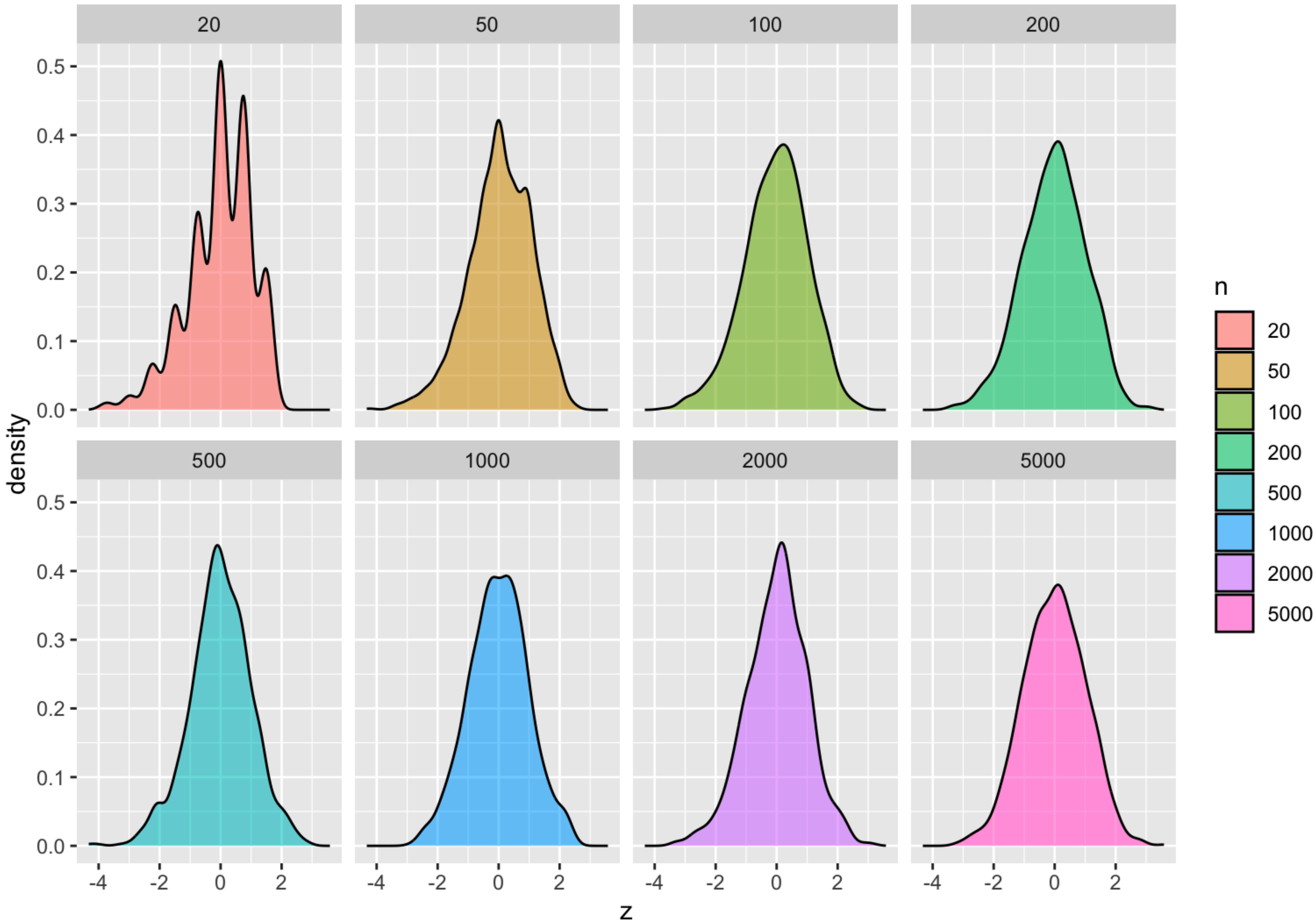
的服从 $N(0,1)$ 渐进分布。（Linderberg-Lévy CLT）

CLT 和 LLN 的最大区别在于，LLN 中的乘数是 $1/n$，而 CLT 中的乘数是 $1/\sqrt{n}$。

- $E[z] = 0$ 是显而易见的。$\mathrm{Var}[z] = n\mathrm{Var}\left[\frac{x_t - \mu}{\sqrt{n\sigma^2}}\right] = \frac{n}{n} = 1$。
- 当 $\boldsymbol{x}_t$ 是随机向量时，CLT 可以写成

$$\operatorname*{plim}_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{x}_t = \boldsymbol{x}_0 \sim N\Big(\boldsymbol{\mu},\ \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathrm{Var}[\boldsymbol{x}_t]\Big)$$

$x_t \sim \mathrm{Bernoulli}(p=0.9)$, 针对每个 n 生成 1000 组随机样本



大样本下的单一约束检验

我们假设回归模型满足 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$，$\boldsymbol{u} \sim \mathrm{IID}\,(\mathbf{0}, \sigma_0^2\boldsymbol{I})$，其中误差项满足 $E[u_t \mid \boldsymbol{X}_t] = 0$，$E[u_t^2 \mid \boldsymbol{X}_t] = \sigma_0^2$。同时假设 $\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{S}_{X^\top X}$，$\boldsymbol{S}_{X^\top X}$ 是有限非随机正定矩阵。在这个假设下，OLS 估计量 $\hat{\boldsymbol{\beta}}$ 满足一致性。

针对 $H_0 : \beta_2 = 0$，已知 t 统计量可以写成

$$t_{\beta_2} = \sqrt{\frac{n-k}{\boldsymbol{y}^\top\boldsymbol{M}_X\boldsymbol{y}}} \cdot \frac{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{y}/\sqrt{n}}{\sqrt{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{x}_2/n}}$$

根据 LLN，第一项依概率收敛于 $\frac{1}{\sigma_0}$，同时在 H_0 成立时 $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{u}$，因此

$$t_{\beta_2} \xrightarrow{p} \frac{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{u}/\sqrt{n}}{\sigma_0\sqrt{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{x}_2/n}}$$

右侧概率极限的分子符合 CLT 的形式，且其期望值为零，方差等于分母，因此可得 $t_{\beta_2} \overset{a}{\sim} N(0,1)$（渐进分布是标准正态分布）。

（详见 pp. 152-154）

大样本下的多重约束检验

已知多重约束检验的 F 统计量是

$$F_{\beta_2} = \frac{\boldsymbol{\varepsilon}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{\varepsilon} \big/ r}{\boldsymbol{\varepsilon}^\top \boldsymbol{M}_X \boldsymbol{\varepsilon} \big/ (n-k)}, \qquad \boldsymbol{\varepsilon} = \boldsymbol{u}/\sigma_0$$

可以将其改写成

$$F_{\beta_2} = \frac{n^{-1/2} \boldsymbol{\varepsilon}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (n^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} n^{-1/2} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{\varepsilon} \big/ r}{\boldsymbol{\varepsilon}^\top \boldsymbol{M}_X \boldsymbol{\varepsilon} \big/ (n-k)}$$

当 $n \to \infty$ 时，可得 $rF_{\beta_2} \sim \chi^2(r)$，或 $F_{\beta_2} \sim F(r, \infty)$。

（尝试推导此结论）

最大似然估计
Maximum Likelihood Estimation

假设随机向量 $\boldsymbol{x}$ 服从参数为 $\boldsymbol{\theta}$ 的分布。在 $\boldsymbol{\theta}$ 已知的情况下，$\boldsymbol{x}$ 的密度函数可以写成 $f(\boldsymbol{x} \mid \boldsymbol{\theta})$。

例如正态分布 $N(\mu, \sigma^2)$ 有两个参数，其密度函数为

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

而在参数估计中，$\boldsymbol{x}$ 的观测值是已知的，但 $\boldsymbol{\theta}$ 是未知的。此时，我们可以将上述密度函数理解为 $\boldsymbol{\theta}$ 的函数，即

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) \equiv f(\boldsymbol{x} \mid \boldsymbol{\theta})$$

称为似然函数（likelihood function）。据此我们可以获得令似然函数最大的参数估计量

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \; L(\boldsymbol{\theta} \mid \boldsymbol{x})$$

在古典正态模型中，
$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$，详见pp.402-404

称为最大似然估计量（maximum likelihood estimator, MLE）。

我们也可以用对数似然函数 $\ell(\boldsymbol{\theta} \mid \boldsymbol{x}) = \log L(\boldsymbol{\theta} \mid \boldsymbol{x})$ 替代似然函数。

在一定条件下，MLE 是一致估计量，其渐进分布是正态分布，并且是渐进最优估计量。

大样本下的三种经典检验方法

这里简单介绍三种大样本下基于最大似然估计的检验方法。设 $H_0 : \boldsymbol{r}(\boldsymbol{\theta}) = \mathbf{0}$ 中包含 r 个非线性假设。
$\boldsymbol{g}(\boldsymbol{\theta}) = \left[\partial \ell(\boldsymbol{\theta}) \big/ \partial \theta_i\right]$ 被称为 score 函数，$\boldsymbol{I}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\boldsymbol{g}(\boldsymbol{\theta})\, \boldsymbol{g}^\top(\boldsymbol{\theta})\right]$ 被称为 information matrix。

- Likelihood ratio test（似然比检验）

$$\mathrm{LR} \equiv 2[\ell(\hat{\boldsymbol{\theta}}_\mathrm{R} - \ell(\hat{\boldsymbol{\theta}}_\mathrm{U})] = 2 \log\left[L(\hat{\boldsymbol{\theta}}_\mathrm{R}) \big/ L(\hat{\boldsymbol{\theta}}_\mathrm{U})\right] \;\overset{a}{\sim}\; \chi^2(r)$$

原理：在 H_0 下，有约束模型和无约束模型的最大似然函数值应该差不太多

- Wald test（沃尔德检验）

$$W \equiv \boldsymbol{r}^\top(\hat{\boldsymbol{\theta}}_\mathrm{U})\left(\boldsymbol{R}(\hat{\boldsymbol{\theta}}_\mathrm{U})\,\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_\mathrm{U})\,\boldsymbol{R}^\top(\hat{\boldsymbol{\theta}}_\mathrm{U})\right)^{-1}\boldsymbol{r}(\hat{\boldsymbol{\theta}}_\mathrm{U}) \;\overset{a}{\sim}\; \chi^2(r), \quad \boldsymbol{R}(\boldsymbol{\theta}) = \left[\partial r_i(\boldsymbol{\theta})/\partial \theta_j\right]$$

原理：在 H_0 下，无约束模型的估计量应该渐进地满足约束条件

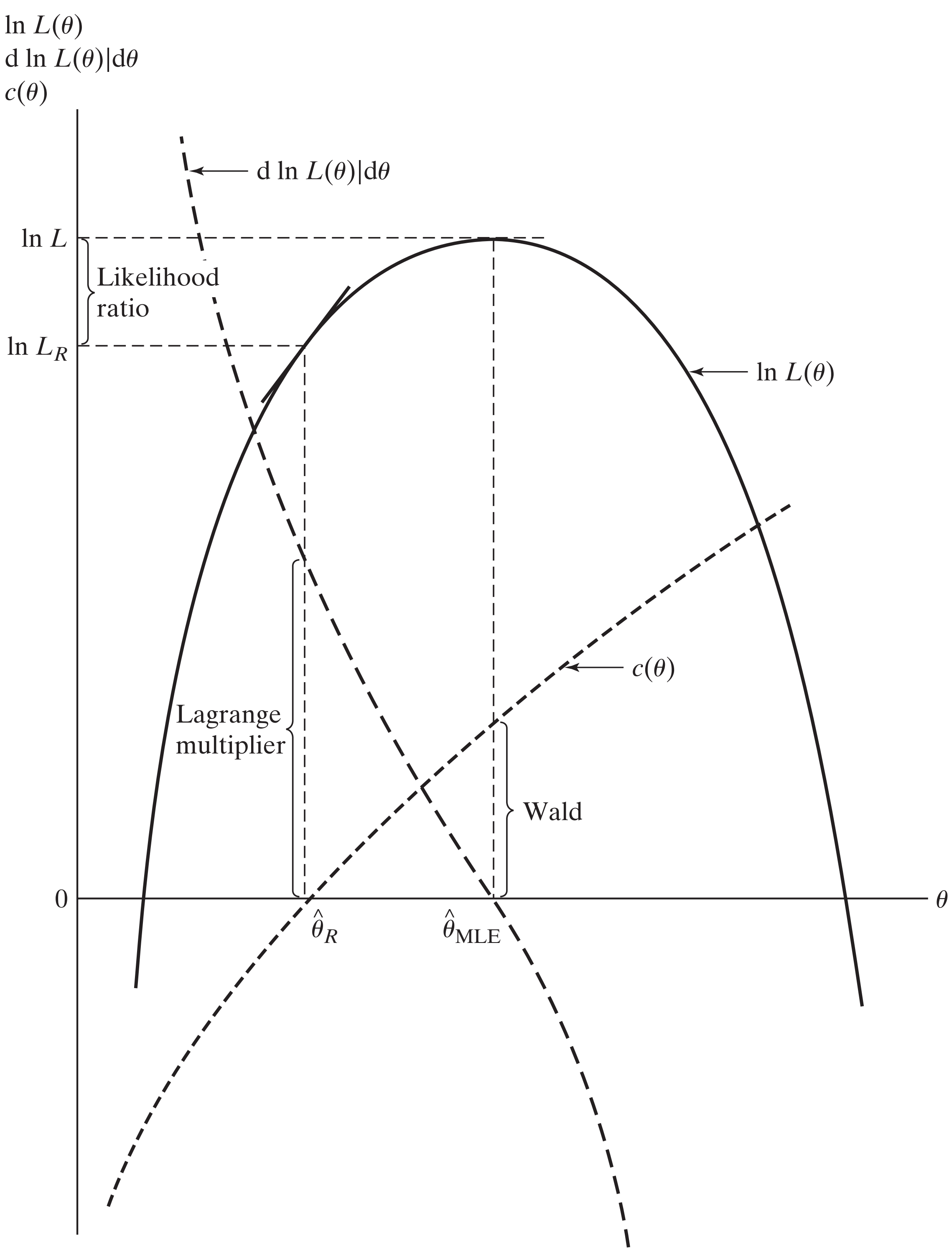
- Lagrange multiplier test（拉格朗日乘数检验）或 Rao’s score test

$$\mathrm{LM} \equiv \boldsymbol{g}^\top(\hat{\boldsymbol{\theta}}_\mathrm{R})\,\boldsymbol{I}(\hat{\boldsymbol{\theta}}_\mathrm{R})^{-1}\,\boldsymbol{g}(\hat{\boldsymbol{\theta}}_\mathrm{R}) \;\overset{a}{\sim}\; \chi^2(r)$$

原理：在 H_0 下，有约束模型的估计量应该很接近真实值，但在 H_1 下应该差异较大

LR, W, and LM Tests

FIGURE 14.2 Three Bases for Hypothesis Tests.



Greene (2020), p.593.

- Wald 检验和 LM 检验可以看作是基于距离函数的检验。
- 三个检验统计量都渐进地服从 $\chi^2(r)$。
- 在线性模型和线性约束下，三个统计量都可以用 *F* 统计量表达，且满足
$$W > LR > LM$$
- 三个检验都是大样本检验，在小样本中的分布未知。
- 在实践中，可以根据计算的简易程度选择统计量。如果有约束模型和无约束模型都很容易估计，就可以选择 LR；如果无约束模型下的统计量相对容易计算（例如线性模型下的非线性约束），就可以选择 W；反之（例如约束条件可以消除模型的非线性）可选择 LM。

置信区间

区间估计

Interval Estimation

真实参数值 θ 和估计量 $\hat{\theta}$ 之间的关系是

$$\theta = \hat{\theta} + \text{抽样误差}$$

$\hat{\theta}$ 是 θ 的点估计（point estimation），而 $\hat{\theta}$ 加上抽样分布可以给出区间估计（interval estimation）。

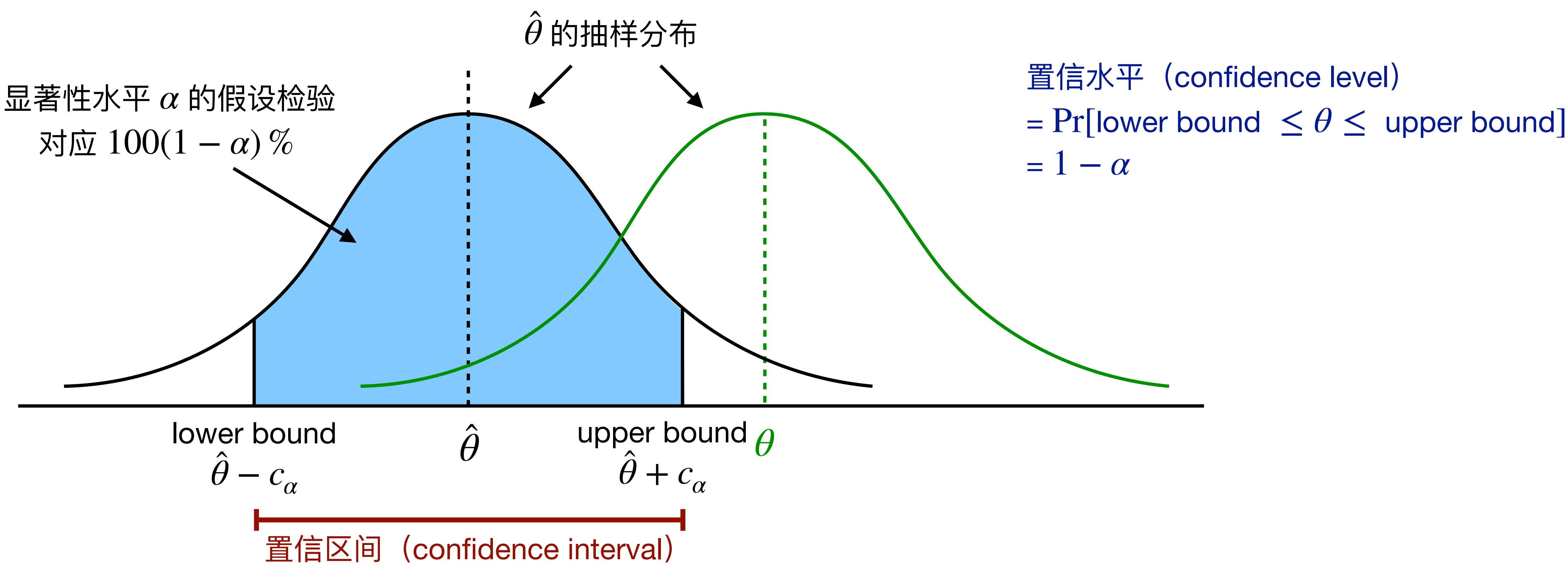
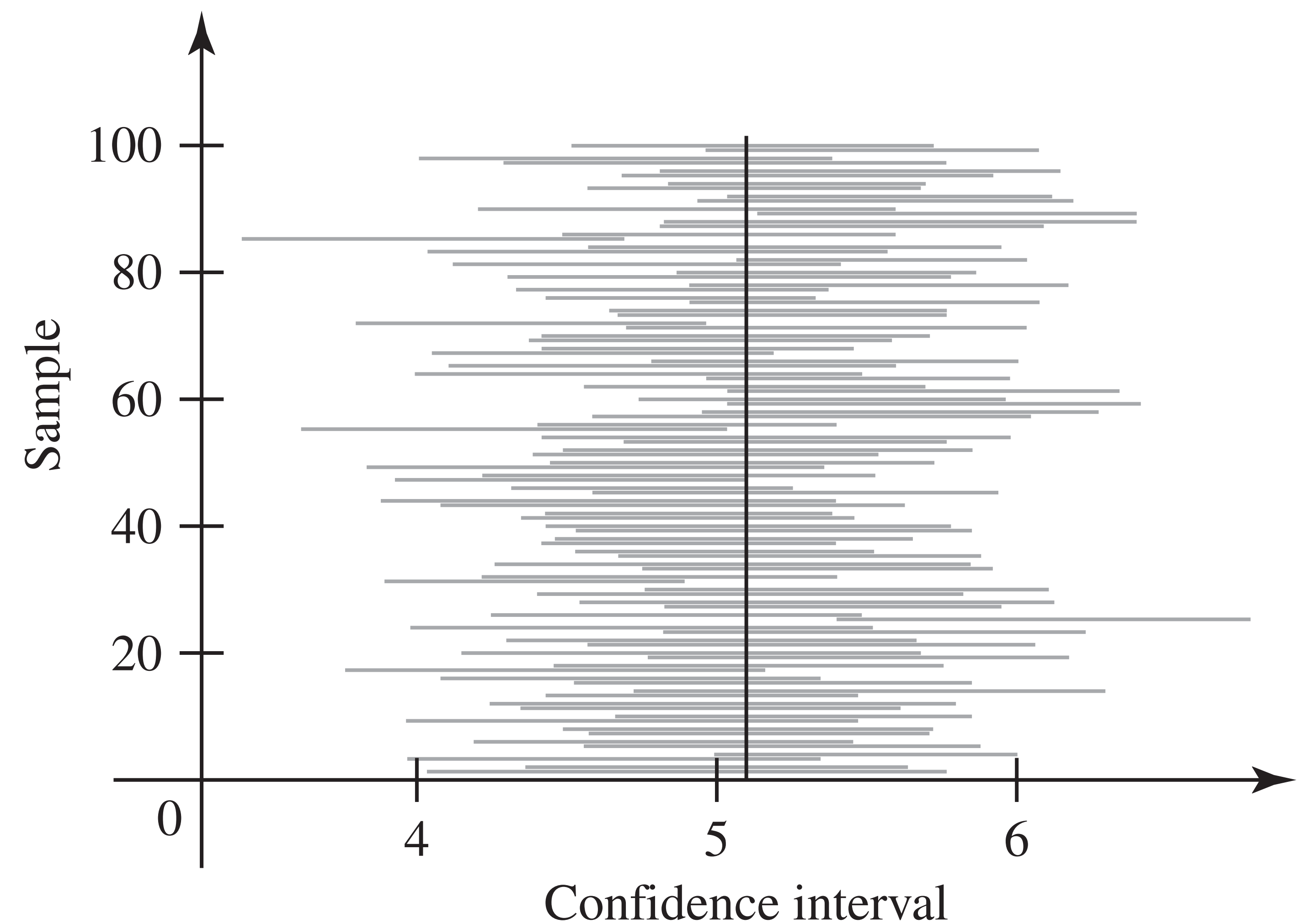


Figure 8.5 A sample of one hundred observed 95% confidence intervals based on samples of size 26 from the normal distribution with mean $\mu = 5.1$ and standard deviation $\sigma = 1.6$. In this figure, 94% of the intervals contain the value of μ.



DeGroot & Schervish (2012), *Probability and Statistics*, 4th Edition, Pearson. (p.478)

根据正态分布 $N(\mu = 5.1, \sigma^2 = 1.6^2)$ 随机生成 $n = 26$ 的样本，然后生成置信区间。图中包含了100个这样的置信区间，其中94个包含真实的分布均值 $\mu = 5.1$。

线性回归系数的置信域

Confidence Region of Linear Regression Coefficients

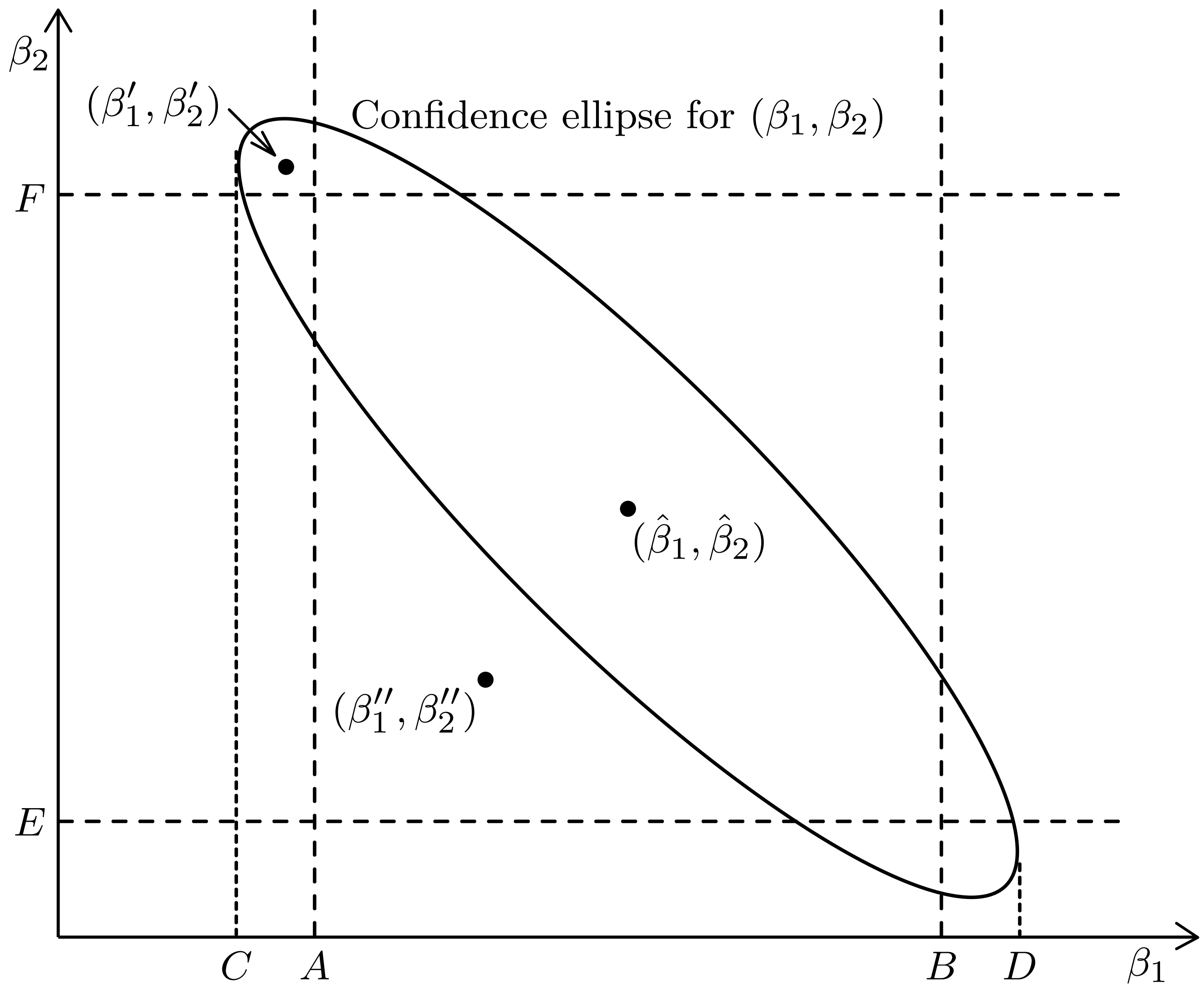


Figure 5.3 Confidence ellipses and confidence intervals

多变量估计量的置信域通常可以写成

$$\text{变量的二次函数} \leq C$$

的形式，其图形含义为椭圆或椭圆体。

注意图中置信域和一维置信区间的区别！

异方差稳健统计量

异方差性及其影响

Heteroskedasticity and its Consequences

异方差性：$\mathrm{Var}[\boldsymbol{u} \mid \boldsymbol{X}] = \boldsymbol{\Omega}$，$\boldsymbol{\Omega}$ 的非对角要素为零，对角要素 ω_t^2 不相同。

在外生性成立时，$\hat{\boldsymbol{\beta}}$ 的协方差矩阵可以写成

$$\begin{aligned}\mathrm{Var}[\hat{\boldsymbol{\beta}}] &= E\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top\right] \\ &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}\, \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\end{aligned}$$

最后一行的表达式被称为 sandwich covariance matrix。 详见 Lecture 6

异方差性的影响：

- OLS 估计量 $\hat{\boldsymbol{\beta}}$ 不再是最有效的，但还是一致的。
- $s^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ 不再是协方差矩阵的非偏估计量，因此影响假设检验的准确性。

异方差时的一致估计

Consistent Estimation under Heteroskedasticity

当 ω_t^2 未知时，我们通常需要对其进行估计。但是我们只有 n 个观测值，却需要估计 n 个 ω_t^2，因此无法直接得到 $\mathbf{\Omega}$ 的一致估计量。

但是我们可以估计 OLS 估计量的协方差矩阵。这里我们用 $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ 替代 $\hat{\boldsymbol{\beta}}$，则有

$$\begin{aligned}\operatorname{Var}\left[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] &= E\left[n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top\right] \\ &= (\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\frac{1}{n}\boldsymbol{X}^\top\mathbf{\Omega}\,\boldsymbol{X})(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X})^{-1}\end{aligned}$$

已知 $\operatorname*{plim}_{n\to\infty} \frac{1}{n}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = (\boldsymbol{S}_{\boldsymbol{X}^\top\boldsymbol{X}})^{-1}$，我们可以用 $\frac{1}{n}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ 作为该极限的一致估计量。

中间项的极限 $\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top\mathbf{\Omega}\,\boldsymbol{X}$ 是 $k \times k$ 的对称矩阵，因此只有 $\frac{1}{2}k(k+1)$ 个参数需要估计。在一定条件下，我们可以通过 $\mathbf{\Omega}$ 的某些非一致估计量 $\hat{\mathbf{\Omega}}$ 对该项进行一致估计，即 $\frac{1}{n}\boldsymbol{X}^\top\hat{\mathbf{\Omega}}\,\boldsymbol{X}$。（White, 1980）

在实际应用中，我们可以忽略 $1/n$ 而直接估计 $\hat{\boldsymbol{\beta}}$ 的协方差矩阵：

$$\widehat{\operatorname{Var}}_{\mathrm{h}}[\hat{\boldsymbol{\beta}}] = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\hat{\mathbf{\Omega}}\,\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$

这种估计量被称为 heteroskedasticity-consistent covariance matrix estimator (HCCME)，或 heteroskedasticity-robust estimator。

HCCMEs

HCCME 的关键是如何找到合适的估计量 $\hat{\boldsymbol{\Omega}}$。因为 $\boldsymbol{\Omega}$ 是对角矩阵，$\hat{\boldsymbol{\Omega}}$ 也是对角矩阵。下面通过定义 $\hat{\boldsymbol{\Omega}}$ 的第 t 对角要素介绍几种常用的 HCCME。

- $HC_0 : \hat{u}_t^2$
- $HC_1 : \dfrac{n}{n-k}\hat{u}_t^2$
- $HC_2 : \hat{u}_t^2 \big/ (1-h_t)$，$h_t$ 是 $\boldsymbol{P_X}$ 的第 t 对角要素
- $HC_3 : \hat{u}_t^2 \big/ (1-h_t)^2$

这四个 HCCME 都满足一致性，但在有限样本下表现都不够好。四个当中 HC_0 表现最差，HC_2 或 HC_3 表现最好。

需要注意的是，有些软件里的默认设定是使用 HC_0，在实践操作中需要人为指定。

课外阅读

- White, H. (1980).
  A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.
  Econometrica, 48:4, 817-838.
  http://www.jstor.org/stable/1912934

- MacKinnon, J. G. and White, H. (1985).
  Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties.
  Journal of Econometrics, 29:3, 305-325.
  https://doi.org/10.1016/0304-4076(85)90158-7

- MacKinnon, J. G. (2005).
  Thirty Years of Heteroskedasticity-Robust Inference.
  In: Chen, X. and Swanson, N. R. (eds.), *Recent Advances and Future Directions 437 in Causality, Prediction, and Specification Analysis*, 437-461, Springer.
  https://doi.org/10.1007/978-1-4614-1653-1_17