Penn State Econometrics Qualifier 2017–2025 参考解答

Unofficial Student Reference Solutions

与系里公开原卷配套使用 · Not official solutions

Rui Zhou

Department of Economics, Penn State University

Public edition · August 2026

中文思路 · English derivations · Source and caveat notes included

说明、来源与勘误边界 · Read This First

非官方参考解答 · Unofficial reference solutions

本卷汇集 Penn State Econometrics Qualifier 2017–2025 的学生参考解答。**这些不是系里或教师发布的官方答案**；它们的用途是展示可得部分分、可在考场复现的推导路径。原始试卷版权归相应权利人所有，公开版资源页直接链接 Penn State Department of Economics 发布的原卷；请把本卷与对应年份原卷配合使用。

题面与年份结构依据系里公开档案核对。2021 年公开扫描件只有两页，缺失题面没有臆造或补写。解答中的课程章节指引保留为知识地图；若某处与课程讲义、教材或官方勘误冲突，应以后者为准。

推荐流程：先限时作答并写下 setup、所用定理与关键等式，再核对本卷。每题的中文部分解释识别方法和易错点，英文部分给出较正式、可复现的推导。欢迎报告可验证的错误；发布页将维护版本日期和说明。

目录

第一部分

真题精解 · Worked Solutions to Past Qualifiers

Chapter 1　2017 资格考精解（Worked Solutions）

本章给出 2017 年 PSU 经济学计量资格考的完整精解。上半场（Q1–Q6）由 Guggenberger 命题，覆盖离散概率（赌徒破产）、Neyman–Pearson 与 UMP 检验、OLS 渐近正态与稳健方差、期望的测度论定义与换元、分块正态条件分布、MLE 不变性；下半场（Q1–Q5）由 Aradillas-Lopez 命题，是一条主线：条件矩约束下的 GMM 渐近方差 → 最优工具（Chamberlain 效率界）→ 线性异方差特例 → 可行两步 GMM → 方差误设下的稳健性。每题先给中文思路，再给英文正式推导与装框结论；能引用前面章节方法的地方直接引用，不重复造轮子。

1.1　501 上半场（Guggenberger）：概率、检验、渐近、测度、正态、MLE

1.1.1　Q1（15 分）：大胆策略 vs. 谨慎策略（赌徒破产）

归属：ECON 501（原卷标签）　**重要度**：中　**再考判断**：中低
核心考点：赌徒破产概率、递推与策略比较
历史复现：2017；离散概率递推也见 2022 Q1
掌握方式：理解递推边界；不必死背最终分式

原题 (2017, Part 501, Q1, 15 pts)

You are in Las Vegas with \$2 and you need \$4 to pay off a debt. You consider two strategies: (a) bet all you have on Black; (b) bet \$1 at a time on Black until you either go bankrupt or end up with \$4. You can assume that the casino pays out twice the wager for a winning roll. Assume the probability of Black is p, where $0 < p < 1$. Compare the two strategies and determine which has the larger probability of succeeding (as a function of p).

Solution.

思路：两种策略各算成功概率再比。大胆策略 (a)：一把押满 \$2，赢（概率 p）就到 \$4，输就破产，所以成功概率正好是 p。谨慎策略 (b)：每次押 \$1 输赢各使财富 ± 1，这是 $\{0, 1, 2, 3, 4\}$ 上的随机游走，从 2 出发，吸收壁在 0 和 4——标准的

赌徒破产问题。算出 (b) 的成功概率后与 p 比较，关键在 p 与 1/2 的大小。

Strategy (a) —bold play. A single bet of the whole \$2 on Black wins with probability p (reaching \$4) and loses with probability $1-p$ (bankruptcy). Hence

$$\pi_a = p.$$

Strategy (b) —timid play. Let S_t be the fortune, a random walk on the state space $\{0,1,2,3,4\}$ with 0 and 4 absorbing, $S_0=2$, taking steps $+1$ w.p. p and -1 w.p. $q:=1-p$ each round. Let $h(s)=\mathrm{P}(\text{reach 4 before } 0 \mid S_0=s)$. Then h solves the gambler's-ruin recursion

$$h(s) = p\,h(s+1) + q\,h(s-1), \qquad h(0)=0,\ h(4)=1.$$

With $r:=q/p$, the standard solution (for $p\neq \frac{1}{2}$, so $r\neq 1$) is

$$h(s) = \frac{1-r^{\,s}}{1-r^{\,4}}.$$

Therefore, starting at $s=2$,

$$\pi_b = h(2) = \frac{1-r^2}{1-r^4} = \frac{1-r^2}{(1-r^2)(1+r^2)} = \frac{1}{1+r^2}.$$

Substituting $r=q/p=(1-p)/p$ and clearing denominators,

$$\boxed{\pi_b = \frac{1}{1+\big((1-p)/p\big)^2} = \frac{p^2}{p^2+(1-p)^2}}$$

For $p=\frac{1}{2}$ the recursion gives $h(s)=s/4$, so $\pi_b=h(2)=\frac{1}{2}=\pi_a$; the boxed formula also evaluates to $\frac{1}{2}$ at $p=\frac{1}{2}$, so it holds for all $p\in(0,1)$ by continuity.

Comparison. Compare $\pi_a=p$ with $\pi_b=\dfrac{p^2}{p^2+(1-p)^2}$. Form

$$\pi_a-\pi_b = p-\frac{p^2}{p^2+(1-p)^2} = \frac{p\big[p^2+(1-p)^2\big]-p^2}{p^2+(1-p)^2} = \frac{p(1-p)\big[(1-p)-p\big]}{p^2+(1-p)^2} = \frac{p(1-p)(1-2p)}{p^2+(1-p)^2}.$$

The denominator is positive and $p(1-p)>0$ on $(0,1)$, so the sign of $\pi_a-\pi_b$ equals the sign of $(1-2p)$. Hence

$$\boxed{\pi_a>\pi_b \text{ if } p<\tfrac{1}{2}, \qquad \pi_a=\pi_b \text{ if } p=\tfrac{1}{2}, \qquad \pi_a<\pi_b \text{ if } p>\tfrac{1}{2}.}$$

直觉（也是经典结论）：在不利赌局（$p<1/2$，正是真实赌场）下，**大胆策略更优**——尽快把钱押出去、少暴露在劣势赔率下；在有利赌局（$p>1/2$）下，谨慎策略更优——多玩几把让大数定律替你工作；公平赌局两者打平。

1.1.2 Q2（25 分）：Bernoulli 的 NP 引理、似然、UMP 检验

归属：ECON 501（原卷标签） **重要度**：高（必会） **再考判断**：高
核心考点：Neyman–Pearson、Bernoulli 似然、MLR 与 UMP
历史复现：2017、2019、2023、2025；五检验见 2018/2025
掌握方式：背 NP/MLR 定理骨架，并能现场化简似然比

原题（2017, Part 501, Q2, 25 pts）

Suppose $X_1, \dots, X_n$ are i.i.d. Bern(p) random variables.
(a) State the lemma by Neyman and Pearson.
(b) What is the likelihood function in this case?
(c) Find a UMP test for $H_0 : p = p_0$ versus $H_1 : p = p_1$, where $p_0 \neq p_1$.
(d) Show that the test from part (c) is UMP for $H_0 : p = p_0$ versus $H_1 : p > p_0$.
(e) Is the test from part (c) UMP for $H_0 : p \leq p_0$ versus $H_1 : p > p_0$? Why, why not?

Solution.

这是一道串联题：NP 引理（简单对简单的最优功效检验）→ 写出 Bernoulli 似然 → 用 NP 引理造检验，发现拒绝域只依赖充分统计量 $T = \sum X_i$ → 因为似然比单调（MLR），单边复合备择仍 UMP（Karlin–Rubin）→ 复合零假设下 size 在 p_0 处取到，故仍 UMP。检验理论见 6。

(a) Neyman–Pearson Lemma. For a *simple* null $H_0 : \theta = \theta_0$ versus a *simple* alternative $H_1 : \theta = \theta_1$ with densities f_0, f_1 (w.r.t. a common dominating measure), fix a size $\alpha \in (0, 1)$. Consider the likelihood-ratio test that rejects when $f_1(x) > k\, f_0(x)$, accepts when $f_1(x) < k\, f_0(x)$, and randomizes on the boundary $f_1(x) = k\, f_0(x)$, with $k \geq 0$ and the randomization probability γ chosen so that

$$\mathbb{E}_{\theta_0}(\phi(X)) = \mathrm{P}(\text{reject} \,|\, \theta_0) = \alpha.$$

Then **(i) Sufficiency:** any such ϕ is most powerful (MP) of size α; **(ii) Existence:** a test of this form attaining size exactly α exists; **(iii) Necessity:** any MP size-α test must be of this likelihood-ratio form a.e. (except possibly on $\{f_1 = kf_0\}$).

(b) Likelihood. With $t = \sum_{i=1}^n x_i$ and each $x_i \in \{0, 1\}$,

$$\boxed{L(p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\,t}(1-p)^{\,n-t}, \qquad t = \sum_{i=1}^n x_i.}$$

(c) UMP test for $H_0 : p = p_0$ vs. $H_1 : p = p_1$. The likelihood ratio is

$$\frac{L(p_1)}{L(p_0)} = \left(\frac{p_1}{p_0}\right)^t \left(\frac{1-p_1}{1-p_0}\right)^{n-t} = \left(\frac{1-p_1}{1-p_0}\right)^n \left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]^t.$$

The ratio is monotone in t through the factor λ^t with $\lambda := \dfrac{p_1(1-p_0)}{p_0(1-p_1)}$.

Case $p_1 > p_0$: then $\lambda > 1$, so $L(p_1)/L(p_0)$ is increasing in t; the NP rejection region $\{L_1/L_0 > k\}$ becomes $\{T > c\}$. The MP test is

$$\phi(T) = \begin{cases} 1, & T > c, \\ \gamma, & T = c, \\ 0, & T < c, \end{cases} \qquad \mathrm{P}_{p_0}(T > c) + \gamma\, \mathrm{P}_{p_0}(T = c) = \alpha,$$

where under H_0, $T = \sum_{i=1}^n X_i \sim \mathrm{Bin}(n, p_0)$. (The randomization γ achieves exact size because T is discrete.)

Case $p_1 < p_0$: then $\lambda < 1$, the ratio decreases in t, and the rejection region is $\{T < c'\}$ (with the analogous boundary randomization).

(d) UMP for $H_0 : p = p_0$ vs. $H_1 : p > p_0$. 要点：当 $p_1 > p_0$ 时，(c) 给出的检验 $\{T > c\}$ 的形式**完全不依赖 p_1 的具体取值**——拒绝域、c、γ 都只由 p_0 和 α 定。 For every fixed alternative $p_1 > p_0$, part (c) yields the *same* test ϕ, since the rejection region $\{T > c\}$ and the constants (c, γ) depend only on p_0 and α, not on p_1. By the NP lemma this single ϕ is MP against each such p_1 simultaneously. A test that is MP against every point of the alternative $\{p > p_0\}$ is, by definition, UMP for $H_1 : p > p_0$. □

(e) UMP for $H_0 : p \le p_0$ vs. $H_1 : p > p_0$? —Yes. 思路：零假设从一点 $\{p_0\}$ 扩成区间 $\{p \le p_0\}$。要证 (i) 这个检验的 size（在整个零假设上的最大拒绝概率）仍是 α；(ii) 它仍对每个备择点最 powerful。两点都靠 Bernoulli 族的单调似然比（MLR）。

The Bernoulli family has *monotone likelihood ratio* (MLR) in $T = \sum_{i=1}^n X_i$: for $p < p'$, $L(p')/L(p) = \text{const} \cdot \lambda^T$ with $\lambda = \frac{p'(1-p)}{p(1-p')} > 1$, increasing in T. Write the power function $\beta(p) := \mathbb{E}_p(\phi(T)) = \mathrm{P}_p(T > c) + \gamma\, \mathrm{P}_p(T = c)$.

Step 1 (monotone power). For any test ϕ that is nondecreasing in T, MLR implies $\beta(\cdot)$ is nondecreasing in p. (Standard MLR fact: $\mathbb{E}_p(g(T))$ is nondecreasing in p for nondecreasing g.) Hence

$$\sup_{p \le p_0} \beta(p) = \beta(p_0) = \alpha,$$

so ϕ has size α for the composite null $\{p \le p_0\}$.

Step 2 (UMP). By part (d), ϕ is UMP among size-α tests for $H_0' : p = p_0$ vs. $H_1 : p > p_0$. Any size-α test for the larger null $\{p \le p_0\}$ is in particular a size-$\le \alpha$ test for the single null $\{p_0\}$; thus its power at any $p_1 > p_0$ cannot exceed $\beta(p_1)$. Since ϕ itself has size exactly α over $\{p \le p_0\}$ (Step 1) and attains the maximal power $\beta(p_1)$ at every $p_1 > p_0$, it is UMP for $H_0 : p \le p_0$ vs. $H_1 : p > p_0$.

$$\boxed{\text{Yes: } \{T > c\} \text{ is UMP for } H_0 : p \le p_0 \text{ vs. } H_1 : p > p_0.}$$

Remark (Karlin–Rubin).

This is exactly the Karlin–Rubin theorem: under MLR in T, the test rejecting for large T is UMP for one-sided hypotheses. See 6.

1.1.3 Q3（20 分）：OLS 渐近正态 + 稳健（White）方差的一致估计

归属：ECON 501（原卷标签；亦为 510 地基） **重要度**：高（必会） **再考判断**：高
核心考点：OLS 渐近正态与异方差稳健三明治方差
历史复现：2017、2024；极值估计模板贯穿当前 510
掌握方式：逐行掌握 LLN–CLT–Slutsky 和残差替换证明

原题 (2017, Part 501, Q3, 20 pts)

Suppose (Y_i, X_i) are i.i.d. with $Y_i = X_i'\beta + U_i$ for $X_i, \beta \in \mathbb{R}^k$, $EX_iU_i = 0$, EX_iX_i' is positive definite, and $E\|U_iX_i\|^2 < \infty$. Show that the least squares estimator $\widehat{\beta}_n$ of β satisfies $n^{1/2}(\widehat{\beta}_n - \beta) \to_d Z \sim N(0, V)$, where

$$V = (EX_iX_i')^{-1}\, EU_i^2X_iX_i'\, (EX_iX_i')^{-1}.$$

Find a consistent estimator $\widehat{V}_n$ of V and show that it is consistent. State any additional assumptions needed to prove asymptotic normality of $\widehat{\beta}_n$ and consistency of $\widehat{V}_n$.

Solution.

标准 OLS 三明治推导，用 4 的 WLLN + CLT + Slutsky 套路。记 $Q := \mathbb{E}(X_iX_i')$，$\Omega := \mathbb{E}\big(U_i^2X_iX_i'\big)$，则 $V = Q^{-1}\Omega Q^{-1}$（White 稳健方差）。方差估计用残差 $\widehat{U}_i$ 的插件，证一致性时要把 $\widehat{U}_i$ 与 U_i 的差控制住，这一步需要额外的矩条件。

Asymptotic normality. The OLS estimator solves the normal equations, giving the algebraic identity

$$\sqrt{n}(\widehat{\beta}_n - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_iU_i\right).$$

Let $Q := \mathbb{E}(X_iX_i')$ (PD by assumption) and $\Omega := \mathbb{E}\big(U_i^2X_iX_i'\big)$.

(i) Denominator. By the WLLN (i.i.d., and $\mathbb{E}\big(\|X_iX_i'\|\big) < \infty$ —implied below), $\frac{1}{n}\sum_{i=1}^{n} X_iX_i' \xrightarrow{p} Q$, so by continuity of matrix inversion $\big(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\big)^{-1} \xrightarrow{p} Q^{-1}$.

(ii) Numerator. The summands $g_i := X_iU_i$ are i.i.d. with mean $\mathbb{E}(X_iU_i) = 0$ and, by the stated $\mathbb{E}\Big(\|U_iX_i\|^2\Big) < \infty$, finite variance $\mathbb{E}(g_ig_i') = \mathbb{E}\big(U_i^2X_iX_i'\big) = \Omega < \infty$. By the multivariate Lindeberg–Lévy CLT,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_iU_i \xrightarrow{d} N(0, \Omega).$$

(iii) Slutsky. Combining (i)–(ii),

$$\sqrt{n}(\widehat{\beta}_n - \beta) \xrightarrow{d} Q^{-1}N(0,\Omega) = N\big(0,\ Q^{-1}\Omega Q^{-1}\big) = N(0,V), \qquad V = Q^{-1}\Omega Q^{-1}.$$

这正是 4 的「$\widehat{\theta} - \theta_0 = (\text{Hessian})^{-1}(\text{score}) + \text{CLT} + \text{Slutsky}$」三明治模板的最干净特例。

Consistent variance estimator. Let $\widehat{U}_i := Y_i - X_i'\widehat{\beta}_n$ be the OLS residual. Define the plug-in (White / Eicker–Huber) estimator

$$\boxed{\widehat{V}_n = \widehat{Q}_n^{-1}\,\widehat{\Omega}_n\,\widehat{Q}_n^{-1}, \qquad \widehat{Q}_n = \frac{1}{n}\sum_{i=1}^n X_iX_i', \quad \widehat{\Omega}_n = \frac{1}{n}\sum_{i=1}^n \widehat{U}_i^{\,2} X_iX_i'.}$$

Consistency proof. We already have $\widehat{Q}_n \xrightarrow{p} Q$ PD, so it suffices that $\widehat{\Omega}_n \xrightarrow{p} \Omega$; then $\widehat{V}_n = \widehat{Q}_n^{-1}\widehat{\Omega}_n\widehat{Q}_n^{-1} \xrightarrow{p} Q^{-1}\Omega Q^{-1} = V$ by the CMT. Write $\widehat{U}_i = U_i - X_i'(\widehat{\beta}_n - \beta)$, so $\widehat{U}_i^2 = U_i^2 - 2U_iX_i'(\widehat{\beta}_n - \beta) + \big(X_i'(\widehat{\beta}_n - \beta)\big)^2$. Hence

$$\widehat{\Omega}_n = \underbrace{\frac{1}{n}\sum_{i=1}^n U_i^2 X_iX_i'}_{=:A_n} - 2\underbrace{\frac{1}{n}\sum_{i=1}^n U_i\,\big(X_i'(\widehat{\beta}_n - \beta)\big)X_iX_i'}_{=:B_n} + \underbrace{\frac{1}{n}\sum_{i=1}^n \big(X_i'(\widehat{\beta}_n - \beta)\big)^2 X_iX_i'}_{=:C_n}.$$

Term A_n: by the WLLN, $A_n \xrightarrow{p} \mathbb{E}\big(U_i^2X_iX_i'\big) = \Omega$ (needs $\mathbb{E}\big(U_i^2\,\|X_iX_i'\|\big) < \infty$). *Terms B_n, C_n:* since $\widehat{\beta}_n - \beta = o_p(1)$, each is bounded by $o_p(1)$ times a sample average that converges in probability to a finite moment of $\|X_i\|$ and U_i; e.g.

$$\|B_n\| \le \left\|\widehat{\beta}_n - \beta\right\| \cdot \frac{1}{n}\sum_{i=1}^n |U_i|\ \|X_i\|^3, \qquad \|C_n\| \le \left\|\widehat{\beta}_n - \beta\right\|^2 \cdot \frac{1}{n}\sum_{i=1}^n \|X_i\|^4.$$

If $\mathbb{E}\Big(|U_i|\,\|X_i\|^3\Big) < \infty$ and $\mathbb{E}\Big(\|X_i\|^4\Big) < \infty$, the sample averages are $O_p(1)$ by the WLLN, so $B_n, C_n = o_p(1)$. Therefore $\widehat{\Omega}_n \xrightarrow{p} \Omega$ and $\widehat{V}_n \xrightarrow{p} V$. □

Additional assumptions needed.

- For asymptotic normality of $\widehat{\beta}_n$: the listed conditions suffice provided the statement "$Q = \mathbb{E}(X_iX_i')$ is positive definite" is read in the standard way—Q exists as a finite matrix and is PD. Equivalently, state explicitly that $\mathbb{E}\Big(\|X_i\|^2\Big) < \infty$. This moment gives the WLLN for $\widehat{Q}_n$; it is *not* implied by $\mathbb{E}\Big(\|U_iX_i\|^2\Big) < \infty$ alone. The latter instead gives $\Omega < \infty$ for the CLT.
- For consistency of $\widehat{V}_n$: the higher-moment conditions $\mathbb{E}\Big(U_i^2\,\|X_i\|^2\Big) < \infty$ (for A_n), and enough additional moments to kill B_n, C_n, e.g. $\mathbb{E}\Big(\|X_i\|^4\Big) < \infty$ and $\mathbb{E}\Big(|U_i|\,\|X_i\|^3\Big) < \infty$ (both hold automatically if, say, $\mathbb{E}\Big(\|X_i\|^4\Big) < \infty$ and $\mathbb{E}\big(U_i^4\big) < \infty$ by Cauchy–Schwarz).

Remark.

> The estimator is heteroskedasticity-robust precisely because $\widehat{\Omega}_n$ uses $\widehat{U}_i^2 X_i X_i'$ rather than a scalar s^2 times $\widehat{Q}_n$. Under conditional homoskedasticity $\mathbb{E}\big(U_i^2 \mid X_i\big) = \sigma^2$, $\Omega = \sigma^2 Q$ and V collapses to $\sigma^2 Q^{-1}$.

1.1.4 Q4（15 分）：期望的测度论定义与换元（image-measure）

归属：ECON 501（原卷标签） **重要度**：高（必会） **再考判断**：高
核心考点：Lebesgue 期望定义与推前测度换元
历史复现：2017、2018、2019、2024
掌握方式：背 indicators–simple–nonnegative–integrable 四步机器

原题 (2017, Part 501, Q4, 15 pts)

Let X be a random variable on the probability space $(\Omega, \mathcal{F}, P)$ and $g(X)$ a scalar function of X. (a) State the definition of $Eg(X)$. [We assume $E|g(X)| < \infty$.] (b) Show that $Eg(X) = \int g\, dP_X$.

Solution.

(a) 期望就是 Lebesgue 积分 $\int_\Omega g(X(\omega))\, dP(\omega)$，要按「简单函数 → 非负可测 → 可积」三步给出 Lebesgue 积分的构造。(b) 是测度的换元（image measure / 推前测度）定理，证明用标准的「measure-theoretic machine」：从示性函数开始，逐级推到一般可积 g。理论见 1。

(a) Definition of $Eg(X)$. Let $Y := g(X) : \Omega \to \mathbb{R}$, assumed $\mathcal{F}$-measurable with $\mathbb{E}(|Y|) < \infty$. Then $Eg(X)$ is the Lebesgue integral of Y against P, built in three stages:

1. *Simple functions.* If $Y = \sum_{j=1}^m a_j \mathbb{1}_{A_j}$ with $a_j \ge 0$, $A_j \in \mathcal{F}$, then $\int_\Omega Y\, dP := \sum_{j=1}^m a_j\, P(A_j)$.
2. *Nonnegative measurable.* If $Y \ge 0$, set $\int_\Omega Y\, dP := \sup\Big\{\int_\Omega S\, dP : 0 \le S \le Y,\ S \text{ simple}\Big\}$.
3. *General integrable.* Write $Y = Y^+ - Y^-$ with $Y^\pm = \max(\pm Y, 0) \ge 0$. Since $\mathbb{E}(|Y|) < \infty$, both $\int Y^\pm dP < \infty$, and

$$Eg(X) := \int_\Omega g(X(\omega))\, dP(\omega) = \int_\Omega Y^+\, dP - \int_\Omega Y^-\, dP.$$

(b) Change of variables: $Eg(X) = \int g\, dP_X$. Let P_X be the law (pushforward) of X on the target space $(\mathbb{R}, \mathcal{B})$, defined by $P_X(B) := P\big(X^{-1}(B)\big) = \mathrm{P}(X \in B)$ for $B \in \mathcal{B}$. We prove $\int_\Omega g(X)\, dP = \int_\mathbb{R} g\, dP_X$ by the standard machine.

Proof for Theorem

Step 1 (indicators). Let $g = \mathbb{1}_B$, $B \in \mathcal{B}$. Then $g(X) = \mathbb{1}_{\{X\in B\}} = \mathbb{1}_{X^{-1}(B)}$, so

$$\int_\Omega \mathbb{1}_B(X)\,dP = P\big(X^{-1}(B)\big) = P_X(B) = \int_{\mathbb{R}} \mathbb{1}_B\,dP_X.$$

Step 2 (nonnegative simple). For $g = \sum_{j=1}^m a_j \mathbb{1}_{B_j}$ with $a_j \ge 0$, linearity of both integrals and Step 1 give $\int_\Omega g(X)\,dP = \sum_j a_j P_X(B_j) = \int_{\mathbb{R}} g\,dP_X$. *Step 3 (nonnegative measurable).* For $g \ge 0$ Borel, pick simple $g_m \uparrow g$ pointwise (e.g. dyadic staircase). Then $g_m(X) \uparrow g(X)$ on Ω, and the Monotone Convergence Theorem applied on both sides yields

$$\int_\Omega g(X)\,dP = \lim_m \int_\Omega g_m(X)\,dP = \lim_m \int_{\mathbb{R}} g_m\,dP_X = \int_{\mathbb{R}} g\,dP_X.$$

Step 4 (general integrable). Decompose $g = g^+ - g^-$. By Step 3 applied to g^+, g^- (each finite since $\int |g|\;dP_X = \int |g(X)|\;dP = \mathbb{E}(|g(X)|) < \infty$),

$$Eg(X) = \int_\Omega g^+(X)\,dP - \int_\Omega g^-(X)\,dP = \int_{\mathbb{R}} g^+\,dP_X - \int_{\mathbb{R}} g^-\,dP_X = \int_{\mathbb{R}} g\,dP_X.$$

$$\boxed{Eg(X) = \int_\Omega g(X(\omega))\,dP(\omega) = \int_{\mathbb{R}} g\,dP_X.}$$

这就是「无意识统计学家定律」(LOTUS) 的测度论版本：算 $\mathbb{E}(g(X))$ 不必先求 $g(X)$ 的分布，直接拿 X 的分布 P_X（连续时 $dP_X = f_X\,dx$，离散时为点质量）去积分即可。

1.1.5　Q5（15 分）：分块正态的条件分布

归属：ECON 501（原卷标签）　**重要度**：高（必会）　**再考判断**：高
核心考点：分块多元正态的条件分布
历史复现：2017、2018、2020、2021、2024、2025
掌握方式：公式要背；推导靠构造零协方差残差

原题 (2017, Part 501, Q5, 15 pts)

Assume $X = (X_1', X_2')' \sim N(\mu, \Sigma)$, where $\mu = (\mu_1', \mu_2')'$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, and $X_i \in \mathbb{R}^{p_i}$ for $i = 1, 2$ for some integers $p_i > 0$. Derive the conditional distribution of X_1 given $X_2 = x_2$. (In your derivation, you can use the facts that linear transformations of X are again normally distributed and that two jointly normal random vectors are independent iff they have zero covariance.)

Solution.

标准技巧（见 3）：构造一个 X_1 与 X_2 的线性组合 $W = X_1 - BX_2$，选 B 使 W 与 X_2 不相关。因为 (W, X_2) 联合正态，零相关 $\Rightarrow$ 独立，于是 $W \mid X_2 = W$（无条件），把条件信息全装进 BX_2 这一项。假设 Σ_{22} 可逆。

Proof for Theorem

Define, with $B := \Sigma_{12}\Sigma_{22}^{-1}$ (Σ_{22} assumed invertible),

$$W := X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 = X_1 - BX_2.$$

(1) (W, X_2) *is jointly normal.* $(W', X_2')' = \begin{pmatrix} I_{p_1} & -B \\ 0 & I_{p_2} \end{pmatrix} X$ is a linear transform of the normal vector X, hence jointly normal.

(2) W *and* X_2 *are uncorrelated.*

$$\mathrm{Cov}(W, X_2) = \mathrm{Cov}(X_1 - BX_2, X_2) = \Sigma_{12} - B\,\Sigma_{22} = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = \Sigma_{12} - \Sigma_{12} = 0.$$

By joint normality, zero covariance $\Rightarrow W \perp X_2$.

(3) Moments of W*.*

$$\mathbb{E}(W) = \mu_1 - B\mu_2 = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,$$

$$\mathrm{Var}(W) = \mathrm{Var}(X_1 - BX_2) = \Sigma_{11} - B\Sigma_{21} - \Sigma_{12}B' + B\Sigma_{22}B' = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

using $B\Sigma_{21} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $\Sigma_{12}B' = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $B\Sigma_{22}B' = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (and $\Sigma_{21} = \Sigma_{12}'$).

(4) Condition. Since $W \perp X_2$, conditioning on $X_2 = x_2$ does not change the law of W: $W \mid X_2 = x_2 \sim N\big(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big)$. But given $X_2 = x_2$, $X_1 = W + Bx_2$ is an affine (deterministic given x_2) shift of W, so it is normal with mean shifted by Bx_2 and the same covariance. ■

$$\boxed{X_1 \mid X_2 = x_2 \ \sim\ N\Big(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2),\ \ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Big).}$$

条件均值 = 无条件均值 + 回归项 $\Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$（正是 X_1 对 X_2 的最优线性预测）；条件方差 = Schur 补 $\Sigma_{11\cdot 2}$，比无条件方差 Σ_{11} 小（知道 X_2 后不确定性下降），且**不依赖** x_2——正态特有的同方差性。

1.1.6 Q6（10 分）：MLE 的不变性

归属：ECON 501（原卷标签） **重要度**：中高 **再考判断**：中高
核心考点：MLE 不变性及多对一参数映射
历史复现：2017；MLE/条件 MLE 在 2018、2019、2025 再现
掌握方式：理解 induced/profile likelihood，避免只背一一映射版本

原题 (2017, Part 501, Q6, 10 pts)

State and prove the invariance property of the maximum likelihood estimator.

Solution.

命题：若 $\widehat{\theta}$ 是 θ 的 MLE，$\tau = g(\theta)$ 是任意（不必一一对应的）变换，则 $g(\widehat{\theta})$ 是 τ 的 MLE。证明关键是对非单射的 g 定义**诱导似然**（profile likelihood）$L^*(\tau) = \sup_{\theta: g(\theta)=\tau} L(\theta)$。见 5。

Theorem 1.1: MLE 不变性 · Invariance of the MLE

Let $\widehat{\theta}$ maximize the likelihood $L(\theta)$ over $\theta \in \Theta$, and let $g : \Theta \to \Lambda$ be any function (not necessarily one-to-one) with $\tau = g(\theta)$. Then $g(\widehat{\theta})$ is an MLE of τ.

Proof for Theorem

For a general (possibly many-to-one) g, define the *induced / profile likelihood* of $\tau \in \Lambda$ by maximizing the original likelihood over the preimage:

$$L^*(\tau) := \sup_{\{\theta\in\Theta:\, g(\theta)=\tau\}} L(\theta).$$

(When g is one-to-one the supremum is over the single point $\theta = g^{-1}(\tau)$ and $L^*(\tau) = L(g^{-1}(\tau))$.) By definition the MLE of τ is $\widehat{\tau} = \arg\max_\tau L^*(\tau)$. We show $\widehat{\tau} = g(\widehat{\theta})$.

Lower bound. Since $\widehat{\theta} \in \{\theta : g(\theta) = g(\widehat{\theta})\}$,

$$L^*\big(g(\widehat{\theta})\big) = \sup_{\theta:\, g(\theta)=g(\widehat{\theta})} L(\theta) \;\ge\; L(\widehat{\theta}).$$

Upper bound. For every $\tau \in \Lambda$, each θ in the preimage $\{g(\theta) = \tau\}$ satisfies $L(\theta) \le L(\widehat{\theta})$ (as $\widehat{\theta}$ maximizes L over *all* of Θ); taking the sup over the preimage,

$$L^*(\tau) = \sup_{\theta:\, g(\theta)=\tau} L(\theta) \;\le\; L(\widehat{\theta}) \qquad \text{for all } \tau.$$

Combining, $L^*(\tau) \le L(\widehat{\theta}) \le L^*(g(\widehat{\theta}))$ for all τ, i.e. $g(\widehat{\theta})$ maximizes L^*. Hence

$$\boxed{\widehat{\tau} = \arg\max_\tau L^*(\tau) = g(\widehat{\theta}), \quad \text{the MLE of } \tau = g(\theta) \text{ is } g(\widehat{\theta}).}$$

∎

直觉：不变性让我们「先在最方便的参数化下求 MLE，再代入想要的函数」。诱导似然的定义保证了即使 g 把多个 θ 映到同一个 τ，「在该 τ 上能达到的最高似然」仍由 $\widehat{\theta}$ 实现。

1.2　510 下半场（Aradillas-Lopez）：条件矩 GMM 与最优工具

下半场共用一套设定：$Z = (Y,X)$，$\rho(Z,\theta) = Y - f(X,\theta)$，条件矩约束 $\mathbb{E}(\rho(Z,\theta_0) \mid X) = 0$（w.p.1）。考虑用其蕴含的无条件矩 $\mathbb{E}(A(X)\rho(Z,\theta_0)) = 0$ 构造的 GMM 类，$A(X)$ 是 $k\times 1$ 工具向量（恰好识别，k 个矩条件配 k 个参数）。这五问是一条逻辑链。本节用 8 的极值估计/GMM 渐近模板。记号约定：$D(X) := \mathbb{E}(\nabla_\theta \rho(Z,\theta_0) \mid X) = -\mathbb{E}(\nabla_\theta f(X,\theta_0) \mid X)$（$k\times 1$），$\sigma^2(X) := \mathbb{E}\big(\rho(Z,\theta_0)^2 \mid X\big) = \mathrm{Var}(Y \mid X)$。

1.2.1　Q1（20 分）：GMM 类的渐近方差通式

归属：ECON 510（原卷标签，旧考法）　**重要度**：中高　**再考判断**：中
核心考点：条件矩 GMM 的一般渐近方差
历史复现：2017–2022 最优工具一脉；2024 三明治模板
掌握方式：背 $G^{-1}\Omega G^{-\prime}$ 并能从矩条件推导

原题（2017, Part 510, Q1, 20 pts）

[Setup] Let Y be a real-valued random variable, X a vector of variables, $Z \equiv (Y,X)$, and $\rho(Z,\theta) = Y - f(X,\theta)$ (1). θ is a k-element parameter, f has known functional form and is as smooth in θ as needed, $\Theta \subset \mathbb{R}^k$ compact. We observe a random sample $\{Z_i\}_{i=1}^n$, and at $\theta_0 \in \Theta$: $E[\rho(Z,\theta_0) \mid X] = 0$ w.p.1 (2). Consider the class of GMM estimators based on $E[A(X)\rho(Z,\theta_0)] = 0$ (3), $A(X)$ a $k\times 1$ vector of functions of X.
Q1. Describe the generic form for the asymptotic variance of the class of GMM estimators described by Equation (3).

Solution.

恰好识别（k 个矩、k 个参数），权重矩阵无关，渐近方差是「三明治」$G^{-1}\Omega G^{-1\prime}$。用 8 的标准展开：对样本矩 $\bar{m}_n(\widehat{\theta}) = 0$ 在 θ_0 处一阶展开即得。

The estimator $\widehat{\theta}_n$ solves the sample analog of (3):

$$\frac{1}{n}\sum_{i=1}^{n} A(X_i)\,\rho(Z_i,\widehat{\theta}_n) = 0.$$

A first-order (mean-value) expansion of $\rho(Z_i,\widehat{\theta}_n)$ around θ_0 gives

$$\frac{1}{n}\sum_{i=1}^{n} A(X_i)\rho(Z_i,\theta_0) + \left[\frac{1}{n}\sum_{i=1}^{n} A(X_i)\nabla_\theta\rho(Z_i,\bar{\theta})'\right](\widehat{\theta}_n - \theta_0) = 0,$$

so

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\left[\tfrac{1}{n}\sum_{i=1}^{n} A(X_i)\nabla_\theta\rho(Z_i,\bar{\theta})'\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A(X_i)\rho(Z_i,\theta_0).$$

Define the $k \times k$ Jacobian and the $k \times k$ variance of the moments:

$$G := \mathbb{E}\big(A(X)\,\nabla_\theta \rho(Z,\theta_0)'\big) = -\,\mathbb{E}\big(A(X)\,\nabla_\theta f(X,\theta_0)'\big)\,, \qquad \Omega := \mathbb{E}\big(A(X)A(X)'\,\rho(Z,\theta_0)^2\big)\,.$$

(Here $\nabla_\theta \rho = -\nabla_\theta f$ since $\rho = Y - f$.) By the WLLN the bracket $\xrightarrow{p} G$; by the CLT (the moments $A(X)\rho(Z,\theta_0)$ are i.i.d., mean zero by (3), variance Ω),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A(X_i)\rho(Z_i,\theta_0) \xrightarrow{d} N(0,\Omega).$$

Slutsky then yields $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0,V_A)$ with

$$\boxed{V_A = G^{-1}\,\Omega\,(G^{-1})' = \Big[\mathbb{E}\big(A(X)\nabla_\theta \rho(Z,\theta_0)'\big)\Big]^{-1}\mathbb{E}\big(A(X)A(X)'\rho(Z,\theta_0)^2\big)\,\Big[\mathbb{E}\big(\nabla_\theta \rho(Z,\theta_0)A(X)'\big)\Big]^{-1}.}$$

Var 内的 $\rho(Z,\theta_0)^2$ 期望可先对 X 取条件：$\Omega = \mathbb{E}\big(A(X)A(X)'\sigma^2(X)\big)$，$G = \mathbb{E}(A(X)D(X)')$，其中 $D(X) = \mathbb{E}(\nabla_\theta \rho \mid X)$、$\sigma^2(X) = \mathbb{E}\big(\rho^2 \mid X\big)$。这两个条件量是 Q2 求最优工具的核心。

1.2.2 Q2（40 分）：最优（高效）工具 $\bar{A}(X)$ ——Chamberlain 效率界

归属：ECON 510（原卷标签，旧考法） **重要度**：中高 **再考判断**：中
核心考点：Chamberlain 最优工具与效率界
历史复现：2017、2018、2020、2021、2022
掌握方式：理解 $D(X)/\sigma^2(X)$ 的投影逻辑，记公式

原题 (2017, Part 510, Q2, 40 pts)

Using your answer to part (1), characterize the efficient (optimal) choice of instruments $\overline{A}(X)$. That is, the instruments $\overline{A}(X)$ that achieve the smallest possible variance within the class of GMM estimators described by Equation (3).

Solution.

目标：在 $A(\cdot)$ 这个函数类上极小化 $V_A = G^{-1}\Omega G^{-1\prime}$（按 Loewner 序，找让 V_A 最小的 A）。先用迭代期望把 G,Ω 写成对 X 条件量的形式，再用一个 Cauchy–Schwarz / 投影型不等式证明 V_A 有下界，且下界由 $\bar{A}(X) = D(X)/\sigma^2(X)$ 取到。这是 Chamberlain (1987) 的条件矩效率界。

Setup via conditioning. With $D(X) := \mathbb{E}(\nabla_\theta \rho(Z,\theta_0) \mid X)$ $(k \times 1)$ and $\sigma^2(X) := \mathbb{E}\big(\rho(Z,\theta_0)^2 \mid X\big) > 0$, the law of iterated expectations gives, since $A(X)$ is X-measurable,

$$G = \mathbb{E}\big(A(X)D(X)'\big)\,, \qquad \Omega = \mathbb{E}\big(A(X)A(X)'\sigma^2(X)\big)\,, \qquad V_A = G^{-1}\Omega G^{-1\prime}.$$

注意 G 里出现的是 $D(X)=\mathbb{E}(\nabla_\theta\rho \mid X)$ 而非 $\nabla_\theta\rho$ 本身——因为 $A(X)$ 可拉进条件期望。

Claim. The optimal instrument is

$$\boxed{\overline{A}(X)=\frac{D(X)}{\sigma^2(X)}=-\frac{\mathbb{E}(\nabla_\theta f(X,\theta_0)\mid X)}{\mathbb{E}(\rho(Z,\theta_0)^2\mid X)}}\,,$$

and it attains the efficiency bound

$$\boxed{V_{\overline{A}}=\left(\mathbb{E}\left(\tfrac{D(X)D(X)'}{\sigma^2(X)}\right)\right)^{-1}=\left(\mathbb{E}\big(\sigma^{-2}(X)\,\nabla_\theta f\,\nabla_\theta f'\big)\right)^{-1}.}$$

Proof for Theorem

Efficiency bound by a Cauchy–Schwarz argument. It suffices to show $V_A\succeq V_{\overline{A}}$ for every admissible A, i.e. $V_A-V_{\overline{A}}$ is positive semidefinite (PSD). Equivalently (since both are inverted), it is cleanest to work with the "information" form. Define the inner product on $k\times 1$ functions of X

$$\langle a,b\rangle:=\mathbb{E}\big(a(X)\,b(X)'\,\sigma^2(X)\big)\quad(k\times k),\qquad S:=\mathbb{E}\left(\tfrac{D(X)D(X)'}{\sigma^2(X)}\right).$$

In this inner product, using $\overline{A}=D/\sigma^2$ (so $\overline{A}\,\sigma^2=D$),

$$G=\langle A,\overline{A}\rangle,\qquad \Omega=\langle A,A\rangle,\qquad S=\langle\overline{A},\overline{A}\rangle.$$

By the matrix Cauchy–Schwarz inequality applied to $U:=A(X)\sigma(X)$ and $W:=\overline{A}(X)\sigma(X)=D(X)/\sigma(X)$,

$$\mathbb{E}\big(UU'\big)\ \succeq\ \mathbb{E}\big(UW'\big)\,\big(\mathbb{E}\big(WW'\big)\big)^{-1}\mathbb{E}\big(WU'\big)\,,$$

that is

$$\Omega=\mathbb{E}\big(UU'\big)\ \succeq\ \mathbb{E}\big(UW'\big)\ S^{-1}\,\mathbb{E}\big(WU'\big)=G\,S^{-1}\,G',$$

because $\mathbb{E}(UW')=\mathbb{E}(A(X)\sigma(X)\,(D(X)/\sigma(X))')=\mathbb{E}(A(X)D(X)')=G$ and $\mathbb{E}(WW')=\mathbb{E}\big(D(X)D(X)'/\sigma^2(X)\big)=S$. Pre- and post-multiplying by G^{-1} and $G^{-1\prime}$ (which preserves the PSD order),

$$V_A=G^{-1}\Omega G^{-1\prime}\ \succeq\ G^{-1}\big(GS^{-1}G'\big)G^{-1\prime}=S^{-1}=V_{\overline{A}}.$$

Attainment. For $A=\overline{A}=D/\sigma^2$: $G=\mathbb{E}\big(\overline{A}D'\big)=\mathbb{E}\big(DD'/\sigma^2\big)=S$ and $\Omega=\mathbb{E}\left(\overline{A}\,\overline{A}'\sigma^2\right)=\mathbb{E}\big(DD'/\sigma^2\big)=S$, hence $V_{\overline{A}}=S^{-1}S\,S^{-1}=S^{-1}$, achieving the bound with equality. ■

解读：最优工具 = （条件雅可比 $D(X)$）除以（条件方差 $\sigma^2(X)$）。分子 $D(X)$ 是「这条矩条件对 θ 有多敏感」——越敏感、信息越多，权重越大；分母 $\sigma^2(X)$

是「这条观测有多吵」——越吵、越不可信，权重越小。这正是 GLS 式的「按信噪比加权」，是条件矩模型的半参数效率界（Chamberlain 1987）。等号条件靠 Cauchy–Schwarz 的投影解释：把任意 A 在内积 $\langle\cdot,\cdot\rangle$ 下投影到 $\overline{A}$ 方向，残差只会增大 Ω 而不改 G。

1.2.3　Q3（20 分）：线性异方差特例下的最优工具

归属：ECON 510（原卷标签，旧考法）　**重要度**：中　**再考判断**：中
核心考点：线性异方差模型中的最优工具
历史复现：2017、2020、2021、2022
掌握方式：由通式现场代入，不单独死背

原题 (2017, Part 510, Q3, 20 pts)

Consider a special case of the model in (1)–(3) where $f(X,\theta) = X'\theta$, so $\rho(Z,\theta) = Y - X'\theta$. Suppose we assume $E[\rho(Z,\theta_0)^2 \mid X] = W(X)'\alpha_0$ for some known function $W(X)$ of X, with α_0 a vector of parameters. What do the optimal instruments $\overline{A}(X)$ look like in this case?

Solution.

把 Q2 的通式代入线性情形。$\rho = Y - X'\theta \Rightarrow \nabla_\theta\rho = -X$，它已经是 X 的函数，故 $D(X) = \mathbb{E}(-X \mid X) = -X$。条件方差被参数化为 $\sigma^2(X) = W(X)'\alpha_0$。代入 $\overline{A} = D/\sigma^2$。

In the linear case $\nabla_\theta\rho(Z,\theta_0) = -X$, which is already X-measurable, so

$$D(X) = \mathbb{E}(\nabla_\theta\rho \mid X) = -X, \qquad \sigma^2(X) = \mathbb{E}\big(\rho(Z,\theta_0)^2 \mid X\big) = W(X)'\alpha_0.$$

Plugging into the optimal-instrument formula from 1.2.2:

$$\boxed{\overline{A}(X) = \frac{D(X)}{\sigma^2(X)} = \frac{-X}{W(X)'\alpha_0} \propto \frac{X}{W(X)'\alpha_0}.}$$

符号无所谓（矩条件 $\mathbb{E}(A(X)\rho) = 0$ 乘 -1 还是同一条），重点是结构：最优工具 = 回归元 X 按其条件方差 $W(X)'\alpha_0$ 的倒数加权——这正是 **GLS/加权最小二乘**。同方差（$\sigma^2(X) \equiv \sigma^2$ 常数）时退化为 $\overline{A} \propto X$，即 OLS 的矩条件 $\mathbb{E}(X(Y - X'\theta)) = 0$。

The resulting estimator solves $\frac{1}{n}\sum_{i=1}^n \dfrac{X_i(Y_i - X_i'\theta)}{W(X_i)'\alpha_0} = 0$, i.e. infeasible GLS with weights $1/(W(X_i)'\alpha_0)$, achieving the bound $V_{\overline{A}} = \big(\mathbb{E}(XX'/(W(X)'\alpha_0))\big)^{-1}$.

1.2.4　Q4（10 分）：可行高效（两步）GMM

归属：ECON 510（原卷标签，旧考法）　**重要度**：中　**再考判断**：中
核心考点：可行有效两步 GMM/FGLS
历史复现：2017；两步估计与残差替换在多年份出现
掌握方式：背 first step–estimate variance–plug in–second step 流程

原题 (2017, Part 510, Q4, 10 pts)

For the model described in part (3), describe a *feasible* efficient GMM estimator (i.e. a GMM estimator that uses a first-step estimator for the optimal instruments $\overline{A}(X)$).

Solution.

$\overline{A}(X) = X/(W(X)'\alpha_0)$ 依赖未知的 α_0，不可行。两步法：第一步用 OLS（或任意一致估计）拿残差，把残差平方回归到 $W(X)$ 上估 α_0，得到 $\widehat{\sigma}^2(X)$；第二步把估出的工具插回矩条件求解（即可行 GLS）。

Two-step feasible efficient GMM (FGLS).

1. *First step —consistent θ.* Run OLS of Y_i on X_i to get $\widetilde{\theta}$ (consistent for θ_0 under (2)). Form residuals $\widetilde{U}_i = Y_i - X_i'\widetilde{\theta}$.
2. *Estimate the skedastic function.* Since $\mathbb{E}\left(U_i^2 \mid X_i\right) = W(X_i)'\alpha_0$, regress $\widetilde{U}_i^2$ on $W(X_i)$ (OLS) to obtain $\widehat{\alpha}$, and set
$$\widehat{\sigma}^2(X_i) = W(X_i)'\widehat{\alpha}, \qquad \widehat{\overline{A}}(X_i) = \frac{X_i}{W(X_i)'\widehat{\alpha}}.$$
(Trim/floor $\widehat{\sigma}^2$ at a small positive constant to guard against nonpositive fitted variances.)
3. *Second step —plug-in GMM.* Solve the feasible sample moment condition
$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\overline{A}}(X_i)\big(Y_i - X_i'\theta\big) = 0 \iff \frac{1}{n}\sum_{i=1}^{n}\frac{X_i(Y_i - X_i'\theta)}{W(X_i)'\widehat{\alpha}} = 0,$$
whose closed form is the feasible GLS estimator
$$\boxed{\widehat{\theta}_{\text{FGLS}} = \Big(\sum_{i=1}^{n}\tfrac{X_iX_i'}{W(X_i)'\widehat{\alpha}}\Big)^{-1}\Big(\sum_{i=1}^{n}\tfrac{X_iY_i}{W(X_i)'\widehat{\alpha}}\Big).}$$

Because $\widehat{\alpha} \xrightarrow{p} \alpha_0$, $\widehat{\overline{A}}(\cdot) \to \overline{A}(\cdot)$. More precisely, the population moment $M(\theta,\alpha) = \mathbb{E}(X\{Y - X'\theta\}/(W(X)'\alpha))$ is orthogonal to first-step perturbations at the truth:

$$\partial_{\alpha'}M(\theta_0,\alpha)|_{\alpha_0} = -\mathbb{E}\left(\frac{X\,U\,W(X)'}{\{W(X)'\alpha_0\}^2}\right) = -\mathbb{E}\left(\frac{XW(X)'}{\{W(X)'\alpha_0\}^2}\mathbb{E}(U \mid X)\right) = 0.$$

Under the usual smoothness, stochastic-equicontinuity, and root-n first-step conditions, estimating α_0 therefore contributes only $o_p(n^{-1/2})$ to the second-step expansion. Hence $\widehat{\theta}_{\mathrm{FGLS}}$ is asymptotically equivalent to the infeasible optimal-instrument GMM of 1.2.3, attaining $V_{\overline{A}} = \left(\mathbb{E}(XX'/(W(X)'\alpha_0))\right)^{-1}$.

1.2.5 Q5（10 分）：方差误设下可行 GMM 仍一致吗？

归属：ECON 510（原卷标签，旧考法） **重要度**：中 **再考判断**：中
核心考点：方差模型错设下的一致性与效率区分
历史复现：2017；模型错设/识别逻辑在当前 510 仍常见
掌握方式：必须会写条件均值为零使任意工具矩仍为零

原题（2017, Part 510, Q5, 10 pts）

For the feasible efficient GMM estimator you proposed in part (4), suppose the assumption $E[\rho(Z,\theta_0)^2 \mid X] = W(X)'\alpha_0$ is incorrect, but the assumption $E[\rho(Z,\theta_0) \mid X] = 0$ is still correct. Will your feasible estimator still be consistent? Prove your claim.

Solution.

结论：仍一致。关键直觉——一致性只靠**条件均值约束** $\mathbb{E}(\rho \mid X) = 0$，它使得**任何** X 的函数当工具都满足无条件矩 $\mathbb{E}(A(X)\rho) = 0$。方差模型错只是把工具用错了（损失效率），但用错的工具仍是 X 的函数，矩条件在 θ_0 处照样成立，识别不受影响。

Claim: yes, $\widehat{\theta}_{\mathrm{FGLS}}$ remains consistent.

Proof for Theorem

Step 1 —the estimated weights converge to a function of X. Even if $\mathbb{E}\big(U^2 \mid X\big) \neq W(X)'\alpha_0$, the first step regresses $\widetilde{U}_i^2$ on $W(X_i)$, whose probability limit $\widehat{\alpha} \xrightarrow{p} \alpha^*$ is the population least-squares coefficient $\alpha^* = \big(\mathbb{E}(WW')\big)^{-1}\mathbb{E}\big(W\,U^2\big)$ (a well-defined constant; $\widetilde{\theta} \xrightarrow{p} \theta_0$ so residuals are asymptotically U_i). Hence

$$\widehat{\widehat{A}}(X) \;\xrightarrow{p}\; A^*(X) := \frac{X}{W(X)'\alpha^*},$$

which is some fixed (possibly "wrong") function of X only —it does *not* depend on Y.

Step 2 —the population moment still holds at θ_0 for any X-measurable instrument. Take *any* instrument $A(X)$ that is a function of X alone. By the law of iterated expectations and the (still valid) conditional mean restriction $\mathbb{E}(\rho(Z,\theta_0) \mid X) = 0$,

$$\mathbb{E}(A(X)\,\rho(Z,\theta_0)) = \mathbb{E}(A(X)\,\mathbb{E}(\rho(Z,\theta_0) \mid X)) = \mathbb{E}(A(X)\cdot 0) = 0.$$

In particular this holds for $A = A^*$. So θ_0 is a solution of the population moment $\mathbb{E}(A^*(X)(Y - X'\theta)) = 0$.

Step 3 —identification and convergence. The limiting estimating equation is

$$\mathbb{E}\big(A^*(X)\,(Y - X'\theta)\big) = 0 \iff \mathbb{E}\big(A^*(X)X'\big)\,(\theta_0 - \theta) = 0,$$

using $Y - X'\theta = (Y - X'\theta_0) + X'(\theta_0 - \theta) = \rho(Z,\theta_0) + X'(\theta_0 - \theta)$ and Step 2. Provided $M^* := \mathbb{E}(A^*(X)X') = \mathbb{E}(XX'/(W(X)'\alpha^*))$ is nonsingular (the usual rank/identification condition, holding whenever $\mathbb{E}(XX')$ is PD and the weights are bounded away from $0,\infty$), the unique solution is $\theta = \theta_0$. The sample moment $\frac{1}{n}\sum_{i=1}^n \widehat{\widehat{A}}(X_i)(Y_i - X_i'\theta)$ converges uniformly (in θ over compact Θ) to this population moment by a ULLN, and $\widehat{\widehat{A}} \xrightarrow{p} A^*$; standard extremum-estimator consistency (8) then gives

$$\boxed{\widehat{\theta}_{\mathrm{FGLS}} \;\xrightarrow{p}\; \theta_0.}$$

一句话总结：方差误设只让你「选错了工具的权重」，因此丢掉效率（不再达到 Chamberlain 界），但工具依旧是 X 的函数，靠 $\mathbb{E}(\rho \mid X) = 0$ 仍正交于扰动 → 矩条件在 θ_0 成立 → 一致性保住。这正是条件矩 GMM 对二阶矩误设稳健的根本原因。

Chapter 2 2018 资格考精解（Worked Solutions）

导读 · Roadmap（先读这里，不含公式）

2018 年的卷子由两份合订：前半是 Guggenberger 的 Econ 501（概率 / 渐近 / 测度 / 检验 / MLE-GMM），后半是 Aradillas-Lopez 的 Econ 510（对称误差线性模型的识别、GMM、最优工具）。本章逐题给出*双语精解*：中文负责思路与直觉，英文负责形式推导与最终（盒装）答案。每题先复述原题，再给完整解答，并*指回*对应方法章节——很多题在前面章节里已是范例，这里只做调用，不重复造轮子。

501 半的暗线是 Guggenberger 的偏好：条件期望走联合正态线性公式、五检验比较强调"没有 UMP"、测度题走标准"指示函数 → 简单函数 → 单调收敛"三步机器、边界参数 MLE（似然平台 / 集值解）、以及*从第一性原理*（不调用极值估计定理）推单矩 GMM 的 plim 与极限分布。510 半的核心戏眼是：对称性把 $\mathbb{E}(\varepsilon \mid Z) = c$ 钉死，于是*截距与 c 纠缠、不可单独识别*——这是贯穿 (i)(ii)(iii) 的主线。

试卷结构与考点地图

题号	分值	考点	方法章节
501-Q1(a)	10	条件期望 / 二元正态	3
501-Q1(b)	10	WLLN + 连续映射（T_n 的 plim）	4
501-Q2(a)	10	五检验定临界值（size $= \alpha$）	6
501-Q2(b)	10	功效函数（$n = 2$）	6
501-Q2(c)	10	无 UMP（功效不可一致占优）	6
501-Q3(a)	6	积分定义的测度（可数可加）	1
501-Q3(b)	9	$\int g\,dv = \int gf\,d\mu$（标准机器）	1
501-Q4(a)	5	边界参数 MLE（$U[-\theta,\theta]$）	5
501-Q4(b)	5	似然平台 / 集值 MLE	5
501-Q4(c)	15	单矩 GMM：plim 与极限分布（第一性原理）	4, 8
501-Q5	10	有界均值 ⇒ 几乎处处有界（独立性必要）	1, 4
510-(i)	20	对称 ⇒ 条件零均值；截距不可识别	7
510-(ii)	40	GMM 构造、渐近分布、截距 plim	8, 7
510-(iii)	40	最优工具（Chamberlain 1987）	8

第一部分 · Econ 501（Guggenberger）

2.1　Q1 ——测量误差下的条件期望与极限

归属：ECON 501（原卷标签）　**重要度**：高（必会）　**再考判断**：高
核心考点：测量误差下的条件正态、LLN 与 CMT
历史复现：条件正态见 2017、2020、2021、2024、2025
掌握方式：背条件正态公式；极限部分现场用 LIE–LLN–CMT

原题 501-Q1（20 分，verbatim）

Suppose the true performance of student i on this exam, X_i, satisfies $X_i \sim N(\mu, \sigma^2)$. Suppose, however, that the grading process records Y_i, where

$$Y_i = X_i + u_i + v_i,$$

$u_i \sim N(0,1)$ is the measurement error from the first grader, $v_i \sim N(0,1)$ is the error from the second grader, and X_i, u_i, v_i are mutually independent.

(a) (10) Calculate $X_i^* = E(X_i \mid Y_i)$ and $\text{var}(X_i \mid Y_i)$.

(b) (10) Assume $i = 1, \ldots, n$. Derive the probability limits of $\overline{X}^* = n^{-1}\sum_{i=1}^n X_i^*$ and of $T_n = \log\big(n^{-1}\sum_{i=1}^n (X_i^* - \overline{X}^*)^2\big)$ as $n \to \infty$. You can apply any theorem we proved in class.

Solution.

(a) 是教科书式“联合正态 ⇒ 条件期望是线性、条件方差是常数”的直接套用，见 3。关键只是把 Y 的方差算对：$Y = X+u+v$，三者独立，$\mathrm{Var}(Y) = \sigma^2+1+1 = \sigma^2+2$。(b) 注意 X_i^* 是 Y_i 的函数、彼此 i.i.d.，所以对它们用 WLLN（样本均值、样本二阶矩各自收敛到总体对应矩），再用连续映射（log）得 T_n 的 plim。陷阱：T_n 收敛到 $\log\big(\mathrm{Var}(X_i^*)\big)$，而 不是 $\log\big(\mathrm{Var}(X_i)\big)$——条件期望“收缩”了方差。

(a) The pair (X_i, Y_i) is jointly normal (a linear map of the independent normals X_i, u_i, v_i). Apply the bivariate-normal conditional formula of Chapter 3. We have

$$\mathbb{E}(X_i) = \mu, \quad \mathbb{E}(Y_i) = \mu, \quad \mathrm{Var}(X_i) = \sigma^2, \quad \mathrm{Var}(Y_i) = \sigma^2+2, \quad \mathrm{Cov}(X_i, Y_i) = \mathrm{Var}(X_i) = \sigma^2,$$

the last because $Y_i = X_i + u_i + v_i$ with u_i, v_i independent of X_i. Hence

$$\boxed{X_i^* = \mathbb{E}(X_i \mid Y_i) = \mu + \frac{\sigma^2}{\sigma^2+2}\,(Y_i - \mu)} \qquad \boxed{\mathrm{Var}(X_i \mid Y_i) = \sigma^2 - \frac{(\sigma^2)^2}{\sigma^2+2} = \frac{2\sigma^2}{\sigma^2+2}}$$

The conditional variance is a constant (independent of Y_i), as always under joint normality.

(b) Write $\rho := \dfrac{\sigma^2}{\sigma^2+2}$, so $X_i^* = \mu + \rho(Y_i - \mu)$. The X_i^* are i.i.d. (functions of the i.i.d. Y_i) with finite moments.

plim of $\overline{X}^*$. By the law of iterated expectations $\mathbb{E}(X_i^*) = \mathbb{E}(\mathbb{E}(X_i \mid Y_i)) = \mathbb{E}(X_i) = \mu$. By the WLLN (Chapter 4),

$$\boxed{\overline{X}^* \xrightarrow{p} \mu}$$

plim of T_n. The inner average is the (uncentered-by-$\overline{X}^*$) sample variance of X_i^*. Both $n^{-1}\sum_{i=1}^n (X_i^*)^2$ and $\overline{X}^*$ converge in probability (WLLN), so by the continuous mapping theorem

$$n^{-1}\sum_{i=1}^n (X_i^* - \overline{X}^*)^2 \xrightarrow{p} \mathrm{Var}(X_i^*)\,.$$

Now $X_i^* - \mu = \rho(Y_i - \mu)$ gives $\mathrm{Var}(X_i^*) = \rho^2\,\mathrm{Var}(Y_i) = \dfrac{\sigma^4}{(\sigma^2+2)^2}(\sigma^2+2) = \dfrac{\sigma^4}{\sigma^2+2}$. Applying $\log(\cdot)$ (continuous on $(0,\infty)$) once more,

$$\boxed{T_n \xrightarrow{p} \log(\mathrm{Var}(X_i^*)) = \log\left(\frac{\sigma^4}{\sigma^2+2}\right) = 4\log\sigma - \log\left(\sigma^2+2\right)}$$

核对：$\mathrm{Var}(X_i^*) = \mathrm{Var}(X_i) - \mathbb{E}(\mathrm{Var}(X_i \mid Y_i)) = \sigma^2 - \frac{2\sigma^2}{\sigma^2+2} = \frac{\sigma^4}{\sigma^2+2}$，与直接算一致（方差分解：条件期望的方差 = 总方差 − 条件方差均值）。

2.2　Q2 ——五个检验：临界值、功效、无 UMP

归属：ECON 501（原卷标签；2025 进入当前 510）　**重要度**：高（必会）　**再考判断**：高
核心考点：五个正态均值检验、功效与无 UMP
历史复现：2018、2025；UMP 讨论见 2017、2019、2023
掌握方式：临界值与功效公式需熟练；无 UMP 用相反方向反例

原题 501-Q2（30 分，verbatim）

Suppose that $X_i \sim N(\mu_i, 1)$ for $i = 1, \ldots, n$ are independent. Consider the following five tests of the null hypothesis $H_0 : \mu_1 = \cdots = \mu_n = 0$ versus the alternative H_1 : "not H_0" with acceptance regions

$$\begin{aligned}
A_1 &= \{(x_1, \ldots, x_n) : \textstyle\sum_{i=1}^n x_i \le k_1\},\\
A_2 &= \{(x_1, \ldots, x_n) : |\textstyle\sum_{i=1}^n x_i| \le k_2\},\\
A_3 &= \{(x_1, \ldots, x_n) : x_i^2 \le k_3 \text{ for all } i = 1, \ldots, n\},\\
A_4 &= \{(x_1, \ldots, x_n) : x_1^2 \le k_4\},\\
A_5 &= \{(x_1, \ldots, x_n) : \textstyle\sum_{i=1}^n x_i^2 \le k_5\},
\end{aligned}$$

where k_j, $j = 1, \ldots, 5$ are constants (that possibly depend on the sample size n) chosen to ensure that the tests have size equal to α.

(a) (10) Determine k_j, $j = 1, \ldots, 5$.

(b) (10) Let $n = 2$ from now on. Determine the power functions of the first and the fourth test.

(c) (10) Show that none of the five tests is (weakly) more powerful than all the other tests considered here uniformly over the alternative space $\{(\mu_1, \mu_2) : -\infty < \mu_1, \mu_2 < \infty\}$.

Solution.

这是 6 的"五检验"范例的原题落点。每个临界值都靠"在 H_0 下把拒绝概率压到 α"定出来，关键是认出每个统计量在 H_0 下的分布：$\sum x_i \sim N(0, n)$、单个 $x_i^2 \sim \chi_1^2$、$\sum x_i^2 \sim \chi_n^2$。记号约定：$z_{1-\alpha} = \Phi^{-1}(1-\alpha)$（标准正态分位数），$\chi^2_{d,1-\alpha}$ 是自由度 d 的 χ^2 的 $1-\alpha$ 分位数。

(a) The test rejects on the *complement* of A_j; "size $= \alpha$" means $\mathrm{P}_{H_0}(\text{reject}) = \alpha$. Under H_0 all $\mu_i = 0$.

A_1 *(one-sided sum).* $\sum_{i=1}^n x_i \sim N(0, n)$, reject when $\sum_{i=1}^n x_i > k_1$. So $\mathrm{P}_{H_0}(\sum_{i=1}^n x_i > k_1) = 1 - \Phi\big(k_1/\sqrt{n}\big) = \alpha$, giving

$$\boxed{k_1 = \sqrt{n}\, z_{1-\alpha}}$$

A_2 (two-sided sum). Reject when $|\sum_{i=1}^n x_i| > k_2$. By symmetry $\mathrm{P}_{H_0}(|\sum_{i=1}^n x_i| > k_2) = 2\big(1-\Phi(k_2/\sqrt{n})\big) = \alpha$, so

$$\boxed{k_2 = \sqrt{n}\, z_{1-\alpha/2}}$$

A_3 (all coordinates). Each $x_i^2 \sim \chi_1^2$ independently. Accept iff every $x_i^2 \le k_3$, so $\mathrm{P}_{H_0}(A_3) = \big[\mathrm{P}(x_1^2 \le k_3)\big]^n = (1-\alpha)$. Hence $\mathrm{P}(\chi_1^2 \le k_3) = (1-\alpha)^{1/n}$ and

$$\boxed{k_3 = \chi^2_{1,\,(1-\alpha)^{1/n}}} \quad \text{(the } (1-\alpha)^{1/n} \text{ quantile of } \chi_1^2\text{)}.$$

A_4 (first coordinate only). $x_1^2 \sim \chi_1^2$, accept iff $x_1^2 \le k_4$, so

$$\boxed{k_4 = \chi^2_{1,\,1-\alpha} = \big(z_{1-\alpha/2}\big)^2}$$

A_5 (sum of squares). $\sum_{i=1}^n x_i^2 \sim \chi_n^2$ under H_0, accept iff $\le k_5$, so

$$\boxed{k_5 = \chi^2_{n,\,1-\alpha}}$$

(b) Set $n = 2$.

Test 1. $x_1 + x_2 \sim N(\mu_1+\mu_2, 2)$, reject when $x_1 + x_2 > k_1 = \sqrt{2}\, z_{1-\alpha}$. The power function is

$$\boxed{\beta_1(\mu_1,\mu_2) = \mathrm{P}(x_1+x_2 > k_1) = 1 - \Phi\left(\frac{k_1 - (\mu_1+\mu_2)}{\sqrt{2}}\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu_1+\mu_2}{\sqrt{2}}\right)}$$

It depends on (μ_1,μ_2) only through $\mu_1+\mu_2$, and is *monotone increasing* in $\mu_1+\mu_2$ (one-sided): power $< \alpha$ when $\mu_1+\mu_2 < 0$.

Test 4. $x_1 \sim N(\mu_1, 1)$, reject when $x_1^2 > k_4 = z^2_{1-\alpha/2}$, i.e. $|x_1| > z_{1-\alpha/2}$. The power depends on μ_1 only:

$$\boxed{\beta_4(\mu_1,\mu_2) = \mathrm{P}\big(|x_1| > z_{1-\alpha/2}\big) = 1 - \Phi\big(z_{1-\alpha/2} - \mu_1\big) + \Phi\big(-z_{1-\alpha/2} - \mu_1\big)}$$

Equivalently $\beta_4 = \mathrm{P}(\chi_1^2(\delta) > k_4)$ with noncentrality $\delta = \mu_1^2$; it ignores μ_2 entirely.

(c) 最短且完全严格的证明用 Neyman–Pearson 唯一性：先找一个简单备择，使 Test 1 是唯一 MP，从而一次排除 Tests 2–5；再找反方向备择排除 Test 1。

Fix any $t > 0$ and consider the simple alternative $\mu^+ = (t,t)$. Relative to $H_0 : (0,0)$, its likelihood ratio is

$$\frac{f_{\mu^+}(x_1,x_2)}{f_0(x_1,x_2)} = \exp\{t(x_1+x_2) - t^2\},$$

which is strictly increasing in $x_1 + x_2$. By the Neyman–Pearson lemma, the unique (up to H_0-null sets) most-powerful size-α test rejects when

$$x_1 + x_2 > \sqrt{2}\, z_{1-\alpha}.$$

This is exactly Test 1. Each of Tests 2–5 has a rejection region that differs from Test 1 on a set of positive H_0 probability; because the likelihood ratio is continuous, NP uniqueness makes the inequality strict:

$$\beta_1(t,t) > \beta_j(t,t), \qquad j = 2,3,4,5.$$

Therefore none of Tests 2–5 can be weakly more powerful than Test 1 uniformly over the alternative space.

It remains to rule out Test 1 itself. At the opposite alternatives $\mu^- = (-t,-t)$,

$$\beta_1(-t,-t) \longrightarrow 0, \qquad \beta_2(-t,-t) \longrightarrow 1 \quad \text{as } t \to \infty,$$

because the one-sided upper-tail test looks in the wrong direction while the two-sided sum test detects an arbitrarily large negative sum. Hence, for all sufficiently large t, $\beta_2(-t,-t) > \beta_1(-t,-t)$, so Test 1 does not uniformly dominate Test 2.

Combining the two comparisons,

$$\boxed{\text{none of the five tests is uniformly weakly more powerful than all the others.}}$$

This proof is valid for every $0 < \alpha < 1$ and needs no numerical power comparison.

□

2.3 Q3 ——由积分定义的测度（Radon–Nikodym 密度）

归属：ECON 501（原卷标签）　**重要度**：高　**再考判断**：高
核心考点：由密度定义测度与换测度积分恒等式
历史复现：2018；同一四步机器见 2017、2019、2024
掌握方式：背可列可加证明和 indicators–simple–MCT 延拓

原题 501-Q3（15 分，verbatim）

Assume f and g are measurable functions $\Omega \to [0,+\infty]$, where $(\Omega,\mathcal{F},\mu)$ is a measure space. Define $v(A) = \int_A f\,d\mu$ for $A \in \mathcal{F}$. Show that

(a) (6) v is a measure,

(b) (9) $\int g\,dv = \int gf\,d\mu$. For b) a sketch of the proof is sufficient.

Solution.

这是 1 的两个标准结论：(a) 验三条（∅、非负、可数可加），可数可加靠把 $\mathbb{1}_{\cup A_k} = \sum_k \mathbb{1}_{A_k}$ 代入并用单调收敛定理交换 $\sum$ 与 $\int$；(b) 是“标准机器”——先指示函数、再简单函数（线性）、再非负可测 g（单调收敛）。f 就是 $dv/d\mu$ 的密度。

(a) Recall $v(A) = \int_A f\,d\mu = \int_\Omega \mathbb{1}_A f\,d\mu$.

(i) $v(\varnothing) = 0$: $\mathbb{1}_\varnothing = 0$, so $\int 0 \cdot f\,d\mu = 0$.

(ii) Nonnegativity: $f \geq 0$ and $\mathbb{1}_A \geq 0$ give $v(A) = \int \mathbb{1}_A f\, d\mu \geq 0$, with values in $[0, +\infty]$.

(iii) Countable additivity: let $\{A_k\}_{k\geq 1} \subseteq \mathcal{F}$ be pairwise disjoint, $A = \bigcup_k A_k$. Then $\mathbb{1}_A = \sum_{k=1}^{\infty} \mathbb{1}_{A_k}$ pointwise, and the partial sums $s_N := \sum_{k=1}^{N} \mathbb{1}_{A_k} f$ increase to $\mathbb{1}_A f$. By the *Monotone Convergence Theorem* (the partial sums are nonnegative and nondecreasing),

$$v(A) = \int \mathbb{1}_A f\, d\mu = \int \lim_N s_N\, d\mu = \lim_N \int s_N\, d\mu = \lim_N \sum_{k=1}^{N} \int \mathbb{1}_{A_k} f\, d\mu = \sum_{k=1}^{\infty} v(A_k).$$

Hence v is a measure on $(\Omega, \mathcal{F})$. □

(b) (Sketch —the “standard machine”.) We prove $\int g\, dv = \int gf\, d\mu$ in four steps.

Step 1 (indicators). For $g = \mathbb{1}_A$, $A \in \mathcal{F}$: $\int \mathbb{1}_A\, dv = v(A) = \int_A f\, d\mu = \int \mathbb{1}_A f\, d\mu$, by the very definition of v.

Step 2 (simple functions). For $g = \sum_{j=1}^{m} c_j \mathbb{1}_{A_j}$ with $c_j \geq 0$: by linearity of the integral and Step 1, $\int g\, dv = \sum_j c_j v(A_j) = \sum_j c_j \int \mathbb{1}_{A_j} f\, d\mu = \int gf\, d\mu$.

Step 3 (nonnegative measurable g). Take simple $0 \leq g_m \uparrow g$ (such a sequence exists for any nonnegative measurable g). Then $g_m f \uparrow gf$ as well. Applying MCT on *both* sides and Step 2,

$$\int g\, dv = \lim_m \int g_m\, dv = \lim_m \int g_m f\, d\mu = \int gf\, d\mu.$$

Step 4 (general / remark). For signed or extended-real g with $\int |g|\, dv < \infty$, split $g = g^+ - g^-$ and apply Step 3 to each part. This is exactly the statement that $f = \dfrac{dv}{d\mu}$ is the Radon–Nikodym density of v w.r.t. μ. □

2.4 Q4 ——边界参数 MLE 与单矩 GMM

归属：ECON 501（原卷标签，含 GMM 小问） **重要度**：中高 **再考判断**：中高
核心考点：边界/集合值 MLE、矩反演、delta 法
历史复现：边界 MLE 见 2018；MLE 与 delta 多年反复
掌握方式：先画参数可行集；GMM 小问按 WLLN–反函数–delta

原题 501-Q4（25 分，verbatim）

This question is about MLE and GMM.

(a) (5) Assume X_i are i.i.d. uniform $[-\theta, \theta]$, $i = 1, \ldots, n$ for some $\theta > 0$. Find the MLE of θ.

(b) (5) Suppose $Y_i = \theta + \varepsilon_i$, $i = 1, \ldots, n$, where ε_i are i.i.d. uniform $[-1, 1]$ and

$\theta \in (0, 1)$. What can you say about the MLE of θ?

(c) (15) In b) compute $E|Y_i|$. Use this calculation to suggest a GMM estimator (with nonrandom weighting "matrix" equal to 1) for the unknown parameter θ. Derive the probability limit and the limiting distribution of this estimator. (It is easier to work from first principles here rather than using the theory of extremum estimators developed in class.)

Solution.

(a)(b) 是 5 的"边界参数"范例：似然不可微，靠支撑约束直接求最大。(a) 单峰、解唯一 $= \max |X_i|$；(b) 似然在一整段 θ 上是平的（常数 $(1/2)^n$），MLE 不唯一——集值。(c) 是本卷计算量最大的一题：先算 $\mathbb{E}(|Y_i|)$ 得到一个把 θ 钉死的矩条件，反解出估计量，再从第一性原理（WLLN+CLT+delta 法，不调用极值估计定理）推 plim 与极限分布。

(a) The density of $U[-\theta, \theta]$ is $\frac{1}{2\theta}\mathbb{1}\{|x| \leq \theta\}$. The likelihood is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2\theta}\mathbb{1}\{|X_i| \leq \theta\} = \left(\frac{1}{2\theta}\right)^n \mathbb{1}\{\theta \geq \max_i |X_i|\}.$$

On its support $\theta \geq \max_i |X_i|$, L is strictly *decreasing* in θ, so it is maximized at the smallest admissible θ:

$$\boxed{\widehat{\theta}_{\text{MLE}} = \max_{1\leq i\leq n} |X_i|}$$

(b) Now $\varepsilon_i \sim U[-1, 1]$ so $Y_i = \theta + \varepsilon_i \in [\theta - 1, \theta + 1]$, density $\frac{1}{2}$ on that interval. The likelihood is

$$L(\theta) = \left(\tfrac{1}{2}\right)^n \mathbb{1}\{\theta - 1 \leq Y_i \leq \theta + 1\ \forall i\} = \left(\tfrac{1}{2}\right)^n \mathbb{1}\left\{\max_i Y_i - 1 \leq \theta \leq \min_i Y_i + 1\right\}.$$

On the (nonempty) interval $\left[\max_i Y_i - 1,\ \min_i Y_i + 1\right] \cap (0, 1)$ the likelihood is *flat* (constant $(1/2)^n$). Hence

$$\boxed{\text{the MLE is } \textit{not unique}\text{: every } \theta \in \left[\max_i Y_i - 1,\ \min_i Y_i + 1\right] \cap (0, 1) \text{ maximizes } L.}$$

要点是识别出"平台似然 $\Rightarrow$ 集值 MLE"。任取该区间内一点都是 MLE（例如区间中点 $\frac{1}{2}(\max_i Y_i + \min_i Y_i)$）；MLE 在边界参数 / 平台似然下可以不是单点。

(c) *Step 1: compute* $\mathbb{E}(|Y_i|)$. Since $\theta \in (0, 1)$, the support $[\theta - 1, \theta + 1]$ straddles 0 ($\theta - 1 < 0 < \theta + 1$). With density $\frac{1}{2}$,

$$\mathbb{E}(|Y_i|) = \frac{1}{2}\int_{\theta-1}^{\theta+1} |y|\, dy = \frac{1}{2}\left[\int_{\theta-1}^{0} (-y)\, dy + \int_{0}^{\theta+1} y\, dy\right] = \frac{1}{2}\left[\frac{(\theta-1)^2}{2} + \frac{(\theta+1)^2}{2}\right] = \frac{\theta^2 + 1}{2}.$$

So the moment function is $m(\theta) := \mathbb{E}(|Y_i|) = \dfrac{\theta^2+1}{2}$, strictly increasing on $(0,1)$ (from $\frac{1}{2}$ to 1), hence invertible: $\theta = \sqrt{2m-1}$ (positive root, as $\theta > 0$).

Step 2: the GMM estimator. With the single moment condition $\mathbb{E}(|Y_i| - m(\theta)) = 0$ and weight 1, GMM sets the sample analogue to zero. Define $\overline{W} := n^{-1}\sum_{i=1}^n |Y_i|$. Solving $\overline{W} = \dfrac{\widehat{\theta}^2+1}{2}$:

$$\boxed{\widehat{\theta}_{\text{GMM}} = \sqrt{2\overline{W}-1} = \sqrt{\frac{2}{n}\sum_{i=1}^n |Y_i| - 1}}$$

(well-defined w.p.a.1 since $\overline{W} \xrightarrow{p} m(\theta) = \frac{\theta^2+1}{2} > \frac{1}{2}$).

Step 3: probability limit (first principles). $|Y_i|$ are i.i.d. with mean $m(\theta)$ and finite variance, so by the WLLN $\overline{W} \xrightarrow{p} m(\theta) = \frac{\theta^2+1}{2}$. The map $w \mapsto \sqrt{2w-1}$ is continuous at $w = m(\theta)$, so by the continuous mapping theorem

$$\boxed{\widehat{\theta}_{\text{GMM}} \xrightarrow{p} \sqrt{2m(\theta)-1} = \sqrt{\theta^2} = \theta} \qquad \text{(consistent)}.$$

Step 4: limiting distribution (CLT + delta method). Let $\sigma_W^2 := \operatorname{Var}(|Y_i|)$. Now

$$\mathbb{E}\big(|Y_i|^2\big) = \mathbb{E}\big(Y_i^2\big) = \operatorname{Var}(Y_i) + (\mathbb{E}(Y_i))^2 = \frac{2^2}{12} + \theta^2 = \frac{1}{3} + \theta^2,$$

so $\sigma_W^2 = \mathbb{E}\big(Y_i^2\big) - \big(\mathbb{E}(|Y_i|)\big)^2 = \frac{1}{3} + \theta^2 - \left(\dfrac{\theta^2+1}{2}\right)^2$. By the Lindeberg–Lévy CLT,

$$\sqrt{n}\,\big(\overline{W} - m(\theta)\big) \xrightarrow{d} N\big(0,\ \sigma_W^2\big).$$

Let $g(w) = \sqrt{2w-1}$, so $\widehat{\theta} = g(\overline{W})$ and $g'(m) = \dfrac{1}{\sqrt{2m-1}} = \dfrac{1}{\theta}$. By the delta method (Chapter 4),

$$\boxed{\sqrt{n}\,\Big(\widehat{\theta}_{\text{GMM}} - \theta\Big) \xrightarrow{d} N\left(0,\ \frac{\sigma_W^2}{\theta^2}\right), \qquad \sigma_W^2 = \frac{1}{3} + \theta^2 - \left(\frac{\theta^2+1}{2}\right)^2.}$$

这正是 8 的“矩条件 → 估计量 → 一致性 → 渐近正态”模板在单参数单矩下的最简版：consistency 走 WLLN+CMT，asymptotic normality 走 CLT+delta 法，无需调用极值估计的一般理论。数值核对：$\theta = 0.4$ 时 $\sigma_W^2/\theta^2 \approx 0.98$，与蒙特卡洛模拟一致。

2.5 Q5 ——有界样本均值蕴含几乎处处有界（独立性必要）

归属：ECON 501（原卷标签） **重要度**：中 **再考判断**：中
核心考点：有界样本均值与独立性作用
历史复现：2018；独立性/收敛反例在 2020–2022 高频
掌握方式：理解反例构造，不背具体随机变量

原题 501-Q5（10 分，verbatim）

Let $\{X_i : 1 \le i \le n\}$ be i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, P)$. Let $M > 0$ be a finite constant.

(a) (6) Show that

$$P\left[\frac{1}{n}\sum_{i=1}^{n}|X_i| \le M\right] = 1 \;\Rightarrow\; P\big[|X_i| \le M\big] = 1.$$

(b) (4) Show that the implication in a) is in general not true if the independence assumption is dropped.

Solution.

这题答案地图漏掉了——它是一道纯测度 / 概率题（1 与独立性）。(a) 反证：若某个 $|X_i| > M$ 有正概率，用 i.i.d. 让"所有坐标同时都很大"有正概率，从而平均超过 M，与前提矛盾。最干净的写法是把事件 $\{|X_1| > M\}$ 复制 n 次、用独立性取交集。(b) 给一个相关的反例：让大值互相抵消，使平均恒等于一个 $\le M$ 的常数，但单个变量可以超过 M。

(a) Suppose, for contradiction, $P(|X_1| > M) =: p > 0$. By the i.i.d. assumption all X_i have the same law, so $P(|X_i| > M) = p$ for each i, and the events $\{|X_i| > M\}$ are independent. Hence

$$P\left(\bigcap_{i=1}^{n}\{|X_i| > M\}\right) = \prod_{i=1}^{n} P(|X_i| > M) = p^n > 0.$$

On the event $\bigcap_{i=1}^{n}\{|X_i| > M\}$ every $|X_i| > M$, so $\frac{1}{n}\sum_{i=1}^{n}|X_i| > M$. Therefore

$$P\left[\frac{1}{n}\sum_{i=1}^{n}|X_i| > M\right] \;\ge\; p^n > 0,$$

contradicting the hypothesis $P\big[\frac{1}{n}\sum_{i=1}^{n}|X_i| \le M\big] = 1$ (equivalently $P\big[\frac{1}{n}\sum_{i=1}^{n}|X_i| > M\big] = 0$). Hence $p = 0$, i.e.

$$\boxed{P\big[|X_i| \le M\big] = 1}$$

☐注：题给的是每个 n（固定 n）的有界性即可推出结论；上面对该固定 n 直接构造正概率的“全大”事件即可，不需要取极限。

(b) Drop independence (fix any $n \geq 2$; for $n = 1$ the hypothesis already forces $|X_1| \leq M$ a.s., so no counterexample exists). The construction makes “exactly one coordinate is large” a.s., so the absolute values never pile up in the average. Let $U \sim U[0,1]$ and set, for $i = 1, \dots, n$,

$$X_i = nM\,\mathbb{1}\left\{\tfrac{i-1}{n} \leq U < \tfrac{i}{n}\right\}.$$

Each X_i equals nM on an event of probability $\frac{1}{n}$ and 0 otherwise (identically distributed, with $P(|X_i| > M) = \frac{1}{n} > 0$ since $nM > M$), and they are *dependent*: exactly one of them is nonzero. Then

$$\frac{1}{n}\sum_{i=1}^{n}|X_i| = \frac{1}{n}\cdot nM = M \quad \text{a.s.},$$

so $P\left[\frac{1}{n}\sum_{i=1}^{n}|X_i| \leq M\right] = 1$, yet $P\left[|X_i| \leq M\right] = 1 - \frac{1}{n} < 1$. The implication fails. ☐反例的灵魂：相依结构让“恰好一个变量很大、其余为零”，把质量分散到不同坐标，使平均被摊平到 $\leq M$，而任一单个变量仍能以正概率越过 M。独立性恰恰排除了这种“轮流爆发”的协调。

第二部分 · Econ 510（Aradillas-Lopez）

2.6 对称误差线性模型：识别、GMM、最优工具

原题 510 模型设定（verbatim）

Consider the following model $Y_i = X_i'\beta_0 + \varepsilon_i$, where $X_i \in \mathbb{R}^{k_1}$ (assume X_i includes an intercept). ε_i is an unobserved latent variable. Suppose there exists an observable $Z_i \in \mathbb{R}^{k_2}$ such that the distribution of ε_i conditional on Z_i satisfies

$$\mathrm{P}(\varepsilon_i \leq c - \epsilon \mid Z_i) = \mathrm{P}(\varepsilon_i \geq c - \epsilon \mid Z_i) \quad \forall \epsilon \in \mathbb{R},$$

almost surely in Z_i, where $c \in \mathbb{R}$ is an unknown constant. (The conditional distribution of ε_i given Z_i is symmetric around the unknown constant c.) For simplicity, assume that the distribution of ε_i conditional on Z_i is absolutely continuous w.r.t. Lebesgue measure (no point-masses conditional on Z_i). The econometrician observes an i.i.d. sample $(Y_i, X_i, Z_i)_{i=1}^{n}$, and $k_2 \geq k_1$. In what follows, feel free to describe any additional assumptions you think are necessary. However, incorrect assumptions will subtract points from your answers.

归属：ECON 510（原卷标签，旧考法）　**重要度**：中高　**再考判断**：中高
核心考点：对称误差、条件矩识别与截距混淆
历史复现：2018；识别在 2020–2025 连续出现
掌握方式：先把对称中心写成条件均值，再检查截距与位置常数

原题 510-(i)（20 分，verbatim）

Show that the previous assumptions imply a conditional mean-zero restriction. From here, characterize a system of unconditional (in Z_i) mean-zero restrictions that can be used to identify and estimate β_0. Can you identify the intercept in β_0 under the assumptions described above?

Solution.

这是 7 的对称误差识别原题。核心三步：(1) 条件对称 ⇒ $\mathbb{E}(\varepsilon \mid Z) = c$（对称分布关于对称中心的均值就是中心）；(2) 把 c 当成一个待估常数并入模型，得到条件零均值矩 $\mathbb{E}(Y - X'\beta_0 - c \mid Z) = 0$，再用任意 $A(Z)$ 乘出无条件矩；(3) **关键结论**——X 含截距，而模型里又冒出一个独立的位移常数 c，两者完全纠缠，**截距不可单独识别**，只有"截距 $+c$"这个和可识别（斜率可识别）。

Step 1: conditional symmetry ⇒ conditional mean. The printed condition has an evident typo (the right side should read $c+\epsilon$, not $c-\epsilon$ —taken literally it would force $\mathrm{P}(\varepsilon_i \le a \mid Z_i) = \frac{1}{2}$ for every a, an impossible CDF). Reading it as symmetry about c, per the parenthetical in the prompt, the condition says that, conditional on Z_i, the law of ε_i is symmetric about c: for every t,

$$\mathrm{P}(\varepsilon_i \le c - t \mid Z_i) = \mathrm{P}(\varepsilon_i \ge c + t \mid Z_i)\,,$$

i.e. $\varepsilon_i - c$ and $-(\varepsilon_i - c)$ have the same conditional distribution given Z_i. Absolute continuity alone does *not* guarantee that a mean exists, so add the necessary integrability condition $\mathbb{E}(|\varepsilon_i| \mid Z_i) < \infty$ a.s. (or $\mathbb{E}(|\varepsilon_i|) < \infty$). Conditional symmetry then makes the integrable odd variable $\varepsilon_i - c$ have conditional mean zero, and hence

$$\boxed{\mathbb{E}(\varepsilon_i \mid Z_i) = c \quad \text{a.s.}}$$

Step 2: conditional mean-zero restriction. Substitute $\varepsilon_i = Y_i - X_i'\beta_0$:

$$\mathbb{E}\big(\, Y_i - X_i'\beta_0 - c \mid Z_i \,\big) = 0 \quad \text{a.s.}$$

Equivalently, defining the *augmented* residual $\tilde{\varepsilon}_i := \varepsilon_i - c = Y_i - X_i'\beta_0 - c$, we have the conditional mean-zero restriction

$$\boxed{\mathbb{E}\big(\, Y_i - X_i'\beta_0 - c \mid Z_i \,\big) = 0 \quad \text{a.s. in } Z_i.}$$

Step 3: unconditional moment system. A conditional mean-zero restriction generates infinitely many unconditional ones: for *any* measurable $A(\cdot) : \mathbb{R}^{k_2} \to \mathbb{R}^{k_1}$ with $\mathbb{E}(\|A(Z_i)\tilde{\varepsilon}_i\|) < \infty$, the law of iterated expectations gives

$$\boxed{\mathbb{E}\big(\, A(Z_i)\,\big(Y_i - X_i'\beta_0 - c\big)\,\big) = 0.}$$

A practical finite system stacks at least k_1 instrument functions in $A(Z_i)$ (e.g. a low-order polynomial / spline basis in Z_i, including a constant). These moments can identify the k_1-dimensional combination $(\beta_{0,1} + c, \tilde{\beta}_0')'$, subject to a full-rank condition; they cannot identify the $k_1 + 1$ primitive coordinates (β_0, c) jointly, as Step 4 shows.

Step 4: the intercept is NOT identified. Write $X_i = (1,\, \tilde{X}_i')'$ and $\beta_0 = (\beta_{0,1},\, \tilde{\beta}_0')'$ with intercept $\beta_{0,1}$. Then

$$Y_i - X_i'\beta_0 - c = Y_i - \tilde{X}_i'\tilde{\beta}_0 - (\beta_{0,1} + c).$$

Only the sum $\beta_{0,1} + c$ enters the moment conditions —the intercept $\beta_{0,1}$ and the symmetry center c are perfectly confounded. Any $(\beta_{0,1}^*, c^*)$ with $\beta_{0,1}^* + c^* = \beta_{0,1} + c$ produces an observationally equivalent model. Therefore

$$\boxed{\text{the slopes } \tilde{\beta}_0 \text{ are identified; the intercept } \beta_{0,1} \text{ is } \textit{not} \text{ separately identified}}$$

(only $\beta_{0,1} + c$ is). 直觉：对称性只约束形状，不约束位置。ε 的对称中心 c 与回归截距都是“纵向平移”，数据无法区分把这点平移记在 c 头上还是截距头上。这与 7 中“正态均值 $\mu_1 + \mu_2$ 可识别、各自不可识别”是同型现象。

归属：ECON 510（原卷标签，旧考法） **重要度**：中高 **再考判断**：中
核心考点：GMM 构造、渐近方差与伪真截距
历史复现：2017–2022 GMM 主线；2024 渐近模板
掌握方式：背 GMM sandwich，识别不足时明确 estimand

原题 510-(ii)（40 分，verbatim）

Describe a GMM estimator based on the unconditional moment restrictions you described in part (i). Characterize the asymptotic distribution of this GMM estimator. What is the probability-limit of the estimated intercept?

Solution.
标准 GMM 模板（8）：把待估参数取为 可识别 的组合 $\gamma_0 = (\beta_{0,1} + c,\ \tilde{\beta}_0')' \in \mathbb{R}^{k_1}$（截距位置上放“截距 $+c$”），样本矩 $g_n(\gamma) = n^{-1}\sum A(Z_i)(Y_i - X_i'\gamma)$，最小化 $g_n'Wg_n$。因为对 γ 而言矩是线性的，估计量有闭式（加权 IV 形式），sandwich 协方差直接套。最后一问 plim：估计出的“截距”收敛到 $\beta_{0,1} + c$，不是 $\beta_{0,1}$——被 c 偏移。

Estimator. Reparametrize by the identified vector $\gamma_0 := \big(\beta_{0,1}{+}c,\ \tilde{\beta}_0'\big)' \in \mathbb{R}^{k_1}$, so that $Y_i - X_i'\beta_0 - c = Y_i - X_i'\gamma_0$ (the regressor on the intercept is the constant

1). Let $A(Z_i)$ be a chosen $k_1 \times 1$ (or $L \times 1$, $L \geq k_1$) instrument vector, and define the sample moments

$$g_n(\gamma) = \frac{1}{n}\sum_{i=1}^{n} A(Z_i)\left(Y_i - X_i'\gamma\right).$$

For a positive-definite weight W (the problem's special case $W = I$ / scalar 1 when $L = k_1$), the GMM estimator is

$$\boxed{\hat{\gamma} = \arg\min_{\gamma}\ g_n(\gamma)'\, W\, g_n(\gamma).}$$

Because the moments are linear in γ, the FOC $G_n' W g_n(\hat{\gamma}) = 0$ with $G_n = -\frac{1}{n}\sum A(Z_i)X_i'$ yields the closed form

$$\hat{\gamma} = \left(\widehat{S}_{AX}'\, W\, \widehat{S}_{AX}\right)^{-1}\widehat{S}_{AX}'\, W\, \widehat{S}_{AY}, \quad \widehat{S}_{AX} = \tfrac{1}{n}\sum A(Z_i)X_i',\ \widehat{S}_{AY} = \tfrac{1}{n}\sum A(Z_i)Y_i,$$

the GMM/IV estimator. (Exactly identified $L = k_1$: $\hat{\gamma} = \widehat{S}_{AX}^{-1}\widehat{S}_{AY}$, and W is irrelevant.)

Asymptotic distribution. Standard GMM theory (Chapter 8). Let

$$G = \mathbb{E}\big(A(Z_i)X_i'\big)\ \ (= -\operatorname{plim} G_n), \qquad \Omega = \mathbb{E}\big(A(Z_i)A(Z_i)'\,\tilde{\varepsilon}_i^2\big), \quad \tilde{\varepsilon}_i = Y_i - X_i'\gamma_0,$$

with G of full column rank k_1 (the identification/relevance condition $k_2 \geq k_1$, rank $G = k_1$). Then

$$\boxed{\sqrt{n}\,(\hat{\gamma} - \gamma_0) \xrightarrow{d} N\big(0,\ V\big), \quad V = (G'WG)^{-1}G'W\,\Omega\,WG\,(G'WG)^{-1}.}$$

With the efficient weight $W = \Omega^{-1}$ this collapses to the optimal-GMM form $V = (G'\Omega^{-1}G)^{-1}$.

plim of the estimated intercept. The first coordinate of $\hat{\gamma}$ estimates the first coordinate of γ_0, which is $\beta_{0,1} + c$, *not* $\beta_{0,1}$. Hence

$$\boxed{\operatorname{plim}\ \widehat{\text{intercept}} = \beta_{0,1} + c}$$

i.e. the estimated intercept converges to the true intercept *plus* the unknown symmetry center c —it is asymptotically biased for $\beta_{0,1}$ by exactly c, the inevitable price of the non-identification in part (i). The slope estimates are consistent for $\tilde{\beta}_0$. 应试落点：(ii) 的“惊喜”就在最后一问。前面一致 / 渐近正态都是模板套用，真正考辨析力的是“估计出的截距 plim 是多少”——答 $\beta_{0,1} + c$，并点明根因是 (i) 的不可识别。

归属：ECON 510（原卷标签，旧考法） **重要度**：中高 **再考判断**：中
核心考点：条件矩的有效工具
历史复现：2017、2018、2020、2021、2022
掌握方式：从 $D(Z)/V(Z)$ 通式推导，不背本题符号

原题 510-(iii)（40 分，verbatim）

Let us limit attention to instrument-functions of the type $A(Z)$, where $A(Z)$ is a $k_1 \times 1$ vector of functions of Z (recall that k_1 denotes the dimension of β_0). Describe how to find the efficient choice of $A(Z)$ given the conditional mean-zero restriction you found in part (i).

Solution.

这是条件矩约束下的最优工具问题——Chamberlain (1987) 的经典结论（8 的高效工具一节）。给定条件零均值矩 $\mathbb{E}(\rho_i(\gamma_0) \mid Z) = 0$，在“$A(Z)$ 型工具”这一类里，使渐近方差最小的工具是

$$A^*(Z) = D(Z)' \, \Sigma(Z)^{-1},$$

其中 $D(Z) = \mathbb{E}(\partial\rho_i/\partial\gamma \mid Z)$ 是残差对参数导数的条件均值、$\Sigma(Z) = \mathrm{Var}(\rho_i \mid Z)$ 是条件方差。本题残差 $\rho_i = Y_i - X_i'\gamma$ 标量、导数 $= -X_i'$、条件方差 $= \mathbb{E}\big(\varepsilon_i^2 \mid Z\big)$（去中心后即 $\tilde{\varepsilon}_i^2$ 的条件期望）。

Setup. The conditional moment restriction is $\mathbb{E}(\rho_i(\gamma_0) \mid Z_i) = 0$ with scalar residual $\rho_i(\gamma) = Y_i - X_i'\gamma$ (using the identified γ_0 of part (ii)). Restricting to instruments of the form $A(Z_i) \in \mathbb{R}^{k_1}$, the resulting estimator solves $\mathbb{E}(A(Z_i)\rho_i(\gamma_0)) = 0$ and has, by the part-(ii) sandwich with W irrelevant in the just-identified case,

$$\mathrm{Avar}(\hat{\gamma}_A) = \big(G_A'\Omega_A^{-1}G_A\big)^{-1} \text{ form, with } \quad G_A = \mathbb{E}\big(A(Z)X'\big), \ \ \Omega_A = \mathbb{E}\big(A(Z)A(Z)'\,\sigma^2(Z)\big),$$

where $\sigma^2(Z) := \mathbb{E}\big(\rho_i^2 \mid Z\big) = \mathrm{Var}(\varepsilon_i \mid Z_i)$ (the conditional variance of the residual).

Chamberlain (1987) optimal instrument. Define

$$D(Z) := \mathbb{E}\left(\frac{\partial\rho_i(\gamma)}{\partial\gamma} \,\Big|\, Z\right) = -\mathbb{E}\big(X_i' \mid Z_i\big) \in \mathbb{R}^{1\times k_1}, \qquad \sigma^2(Z) := \mathbb{E}\big(\rho_i^2 \mid Z_i\big) = \mathbb{E}\big(\varepsilon_i^2 \mid Z_i\big) - c^2 \text{ (conditional}$$

The variance-minimizing instrument within the class $A(Z)$ is

$$\boxed{A^*(Z_i) = \frac{D(Z_i)'}{\sigma^2(Z_i)} = -\frac{\mathbb{E}(X_i \mid Z_i)}{\mathrm{Var}(\varepsilon_i \mid Z_i)}}$$

(a $k_1 \times 1$ vector), achieving the semiparametric efficiency bound for this conditional moment restriction:

$$\operatorname{Avar}(\hat{\gamma}_{A^*}) = \left(\mathbb{E}\left(\frac{\mathbb{E}(X_i \mid Z_i)\,\mathbb{E}(X_i \mid Z_i)'}{\operatorname{Var}(\varepsilon_i \mid Z_i)}\right)\right)^{-1}.$$

Implementation (feasible version). A^* depends on the unknown $\mathbb{E}(X_i \mid Z_i)$ and $\operatorname{Var}(\varepsilon_i \mid Z_i)$. The standard two-step / feasible-efficient procedure:

1. Get a consistent first-step $\hat{\gamma}$ (any valid $A(Z)$, e.g. $A(Z) = Z$ truncated to k_1 components, or a polynomial basis).
2. Nonparametrically estimate $\widehat{m}(Z) = \widehat{\mathbb{E}(X \mid Z)}$ (series / kernel regression of X on Z) and $\widehat{\sigma}^2(Z) = \widehat{\mathbb{E}(\hat{\rho}^2 \mid Z)}$ from first-step residuals $\hat{\rho}_i = Y_i - X_i'\hat{\gamma}$.
3. Form $\widehat{A}^*(Z_i) = -\widehat{m}(Z_i)/\widehat{\sigma}^2(Z_i)$ and re-estimate $\hat{\gamma}$ using $\widehat{A}^*$ as instrument.

Under standard regularity (and convergence rates for the nonparametric first step), the feasible estimator attains the same efficiency bound as the infeasible A^*. 记忆口诀（8）：*最优工具 = “斜率方向”除以“噪声大小”*——分子是残差对参数的导数的条件均值（这里 $-\mathbb{E}(X \mid Z)$，即把内生 / 含噪的 X 投影到工具 Z 上），分母是条件方差（异方差加权）。同方差 $\sigma^2(Z) \equiv \sigma^2$ 时退化为 $A^* \propto -\mathbb{E}(X \mid Z)$，正是把 X 用其对 Z 的最优预测替换的“最优 IV”。

本年小结 · Year Wrap-Up

501（Guggenberger）的可迁移套路：(1) 联合正态 ⇒ 条件期望线性、条件方差常数（Q1a），配上 LIE + WLLN + 连续映射做 plim（Q1b）；(2) 五检验比临界值靠认对 H_0 下分布（$N(0,n)$ / χ^2_1 / χ^2_n），“无 UMP”靠给每个检验找一个失明方向（Q2）；(3) 测度题走“指示 → 简单 → MCT”三步机器（Q3）；(4) 边界参数 MLE 看支撑约束、平台似然给集值解（Q4ab）；(5) 单矩 GMM 从第一性原理：WLLN+CMT 得 plim、CLT+delta 得极限分布（Q4c）；(6) i.i.d. 下“全大”事件正概率 p^n 的反证、相依下“轮流爆发”反例（Q5）。

510（Aradillas-Lopez）的戏眼：对称性只钉死*形状*不钉死*位置*，于是 c 与截距纠缠、*截距不可识别*（这条贯穿 i/ii/iii）；GMM 套模板，但估计出的截距 plim $= \beta_{0,1} + c$；最优工具走 Chamberlain：$A^* = -\mathbb{E}(X \mid Z) / \operatorname{Var}(\varepsilon \mid Z)$。

Chapter 3　2019 资格考精解（Worked Solutions）

导读 · Roadmap（先读这里，不含公式）

本章把 2019 年两场资格考逐题做成“原题 + 完整解”：上半场是 ECON 510（Aradillas-Lopez）的一道大题，主角是**两步 2SLS** 与**忽略内生性的 OLS**——考点全部落在“两步估计量的影响函数”和“OLS 在内生时收敛到哪里”。下半场是 Guggenberger 的综合考（501 基础），六道小题串起：联合密度 → 边缘/条件/条件期望、测度论换元（推前测度）、样本均值与样本方差的联合 CLT、截断法证 WLLN、二项 MLE + Neyman–Pearson + 单边 UMP、以及二阶 delta 法。每道题先**照录题面**（个别处略作排版整理，不改数学内容），再给**中文思路**（先想清楚再动笔）与 *English formal derivation*（带方框最终答案）。能引用正文方法的地方一律 `\ref` 指过去、就地套用，不重复造轮子。

3.1　ECON 510 · Aradillas-Lopez：两步 2SLS 与内生 OLS

［优先级 ●●● 高 High］　资格考足迹：2019(510)

归属：ECON 510（原卷标签，旧考法）　**重要度**：中高　**再考判断**：中
核心考点：两步 2SLS/generated regressor 的影响函数
历史复现：2019；两步修正在 GMM、Heckit、bootstrap 中复现
掌握方式：掌握 stacked expansion，不能漏一阶段估计误差

原题 ECON 510 Q1（verbatim, August 2019）

Consider the structural equation

$$Y_{1i} = X_i'\beta_0 + \delta_0 Y_{2i} + \varepsilon_i. \tag{1}$$

Let $\theta_0 \equiv (\beta_0', \delta_0)'$ denote the structural parameters. Y_{2i} is a scalar random variable and X_i is a random vector. Assume X_i is exogenous: $\mathbb{E}(X_i\varepsilon_i) = 0$, but $\mathbb{E}(Y_{2i}\varepsilon_i) \neq 0$ (i.e., Y_{2i} may be endogenous). Model the endogeneity of Y_{2i}

through the reduced-form equation

$$Y_{2i} = Z_i'\gamma_0 + \eta_i. \tag{2}$$

Z_i is exogenous: $\mathbb{E}(Z_i\eta_i) = 0$ and $\mathbb{E}(Z_i\varepsilon_i) = 0$; also $\mathbb{E}(X_i\eta_i) = 0$. Z_i contains at least one variable excluded from X_i. Allow $\mathbb{E}(\eta_i\varepsilon_i) \neq 0$, so η_i is the source of endogeneity. Plugging (2) into (1),

$$Y_{1i} = X_i'\beta_0 + \delta_0\, Z_i'\gamma_0 + \nu_i, \qquad \nu_i \equiv \delta_0\eta_i + \varepsilon_i. \tag{3}$$

We have an iid sample $(Y_{1i}, Y_{2i}, X_i, Z_i)_{i=1}^n$.

(1) [60 pts] Estimate θ_0 by a two-step procedure. *Step 1:* estimate γ_0 in (2) by OLS, call it $\hat\gamma$. *Step 2:* estimate θ_0 by OLS of (3) with γ_0 replaced by $\hat\gamma$. Call the result $\widehat{\theta}^{2SLS}$. Describe the asymptotic properties of $\widehat{\theta}^{2SLS}$ by characterizing its influence function.

Solution.

中文思路。这是“生成回归元 / 两步估计”的经典题。第二步要回归的真模型是 (3)：$Y_{1i} = \tilde W_i'\theta_0 + \nu_i$，其中 $\tilde W_i = (X_i',\, Z_i'\gamma_0)'$。关键观察：在 (3) 里 $\tilde W_i$ 与 ν_i 正交 ($\mathbb{E}(X_i\nu_i) = \delta_0\mathbb{E}(X_i\eta_i) + \mathbb{E}(X_i\varepsilon_i) = 0$; $\mathbb{E}((Z_i'\gamma_0)\nu_i) = \gamma_0'(\delta_0\mathbb{E}(Z_i\eta_i) + \mathbb{E}(Z_i\varepsilon_i)) = 0$)，所以 (3) 的总体 OLS 就识别 θ_0。但第二步用的是 $\hat W_i = (X_i',\, Z_i'\hat\gamma)'$ ——回归元里塞进了第一步的估计 $\hat\gamma$。于是要把第一步的抽样误差通过 $\hat\gamma$ 传到第二步：这就是 generated-regressor 修正项。整道题就是把这两段影响函数“串”起来，按 8 的两步极值估计/M 估计模板做链式展开。

Notation. Write $\tilde W_i = (X_i',\, Z_i'\gamma_0)'$ and $\hat W_i = (X_i',\, Z_i'\hat\gamma)'$, both (k_x+1)-vectors. Let

$$Q \equiv \mathbb{E}\left(\tilde W_i \tilde W_i'\right), \qquad M_{ZZ} \equiv \mathbb{E}\left(Z_i Z_i'\right),$$

both assumed nonsingular. The selector that picks out the δ-component is $e \equiv (0_{k_x}',\, 1)'$, so the last entry of $\hat W_i$ is $e'\hat W_i = Z_i'\hat\gamma$ and $\hat W_i = \tilde W_i + e\, Z_i'(\hat\gamma - \gamma_0)$.

Step 1 influence function (OLS of (2)). By the standard OLS expansion (see 4),

$$\sqrt{n}(\hat\gamma - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{ZZ}^{-1} Z_i\eta_i + o_p(1), \qquad \mathbb{E}\left(M_{ZZ}^{-1} Z_i\eta_i\right) = 0.$$

Step 2 expansion. The second-step estimator solves the normal equations $\widehat{\theta}^{2SLS} = \left(\frac{1}{n}\sum_{i=1}^n \hat W_i \hat W_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n \hat W_i Y_{1i}$, equivalently

$$\frac{1}{n}\sum_{i=1}^n \hat W_i \left(Y_{1i} - \hat W_i'\, \widehat{\theta}^{2SLS}\right) = 0.$$

Substitute $Y_{1i} = \tilde{W}_i'\theta_0 + \nu_i$ and $\hat{W}_i = \tilde{W}_i + e\,Z_i'(\hat{\gamma} - \gamma_0)$. Two perturbations enter: (a) the residual ν_i (second-stage sampling noise), and (b) the generated regressor through $\hat{\gamma} - \gamma_0$. A first-order (mean-value) expansion gives

$$\sqrt{n}(\widehat{\theta}^{2SLS} - \theta_0) = Q^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_i\nu_i \;+\; Q^{-1}\,J\,\sqrt{n}(\hat{\gamma} - \gamma_0) \;+\; o_p(1),$$

where J is the derivative of the population second-step moment $\mathbb{E}\Big(\hat{W}_i(Y_{1i} - \hat{W}_i'\theta_0)\Big) = \mathbb{E}\Big(\hat{W}_i\nu_i(\gamma)\Big)$ with respect to γ, evaluated at γ_0. 这个 J 就是生成回归元修正的“链条接头”：第一步误差怎样改变第二步的总体矩。

Computing J. Treat γ as a parameter inside $W_i(\gamma) = (X_i', Z_i'\gamma)'$ and let $r_i(\gamma) = Y_{1i} - W_i(\gamma)'\theta_0$. Then $\mathbb{E}(W_i(\gamma)\,r_i(\gamma))$ has γ-derivative (using $\partial W_i/\partial\gamma' = eZ_i'$ and $\partial r_i/\partial\gamma' = -\theta_0'\,\partial W_i/\partial\gamma' = -\delta_0 Z_i'$, since only the δ_0-loaded component of W_i depends on γ):

$$J = \mathbb{E}\big(\,e\,Z_i'\,r_i(\gamma_0)\,\big) - \mathbb{E}\big(\,W_i(\gamma_0)\,\delta_0 Z_i'\,\big) = \underbrace{\mathbb{E}\big(e\,Z_i'\nu_i\big)}_{=\,e\,\mathbb{E}(Z_i\nu_i)'=0} - \delta_0\,\mathbb{E}\Big(\tilde{W}_i Z_i'\Big) = -\,\delta_0\,\mathbb{E}\Big(\tilde{W}_i Z_i'\Big),$$

because $\mathbb{E}(Z_i\nu_i) = \delta_0\mathbb{E}(Z_i\eta_i) + \mathbb{E}(Z_i\varepsilon_i) = 0$. 第一块消失——总体上 Z_i 与 ν_i 正交，所以“残差对 γ 的直接依赖”不贡献；剩下的是回归元本身被 δ_0 放大的那一项。

Assemble the influence function. Define

$$\boxed{\psi_i^{2SLS} = Q^{-1}\Big[\tilde{W}_i\nu_i \;-\; \delta_0\,\mathbb{E}\Big(\tilde{W}_i Z_i'\Big)\;M_{ZZ}^{-1}Z_i\eta_i\Big].}$$

Then

$$\boxed{\sqrt{n}(\widehat{\theta}^{2SLS} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i^{2SLS} + o_p(1) \;\xrightarrow{d}\; N\big(0,\;\mathbb{E}\big(\psi_i^{2SLS}\psi_i^{2SLS\prime}\big)\big).}$$

Equivalently, with $A \equiv \delta_0\,\mathbb{E}\Big(\tilde{W}_i Z_i'\Big)\,M_{ZZ}^{-1}$ (the generated-regressor correction matrix),

$$\psi_i^{2SLS} = Q^{-1}\big(\tilde{W}_i\nu_i - A\,Z_i\eta_i\big).$$

The asymptotic variance is $V_{2SLS} = Q^{-1}\,\mathbb{E}\Big((\tilde{W}_i\nu_i - AZ_i\eta_i)(\tilde{W}_i\nu_i - AZ_i\eta_i)'\Big)\;Q^{-1}$.

Remark (为什么必须修正第一步).

> 如果错误地把 $\hat{\gamma}$ 当作已知 γ_0，会漏掉 $-AZ_i\eta_i$ 这一项，方差会算错（一般偏小）。这正是“生成回归元 / 两步估计”题的得分点：必须把第一步的影响函数 $M_{ZZ}^{-1}Z_i\eta_i$ 经 $J=-\delta_0\mathbb{E}\left(\tilde{W}_iZ_i'\right)$ 传到第二步。 The correction vanishes only in the knife-edge cases $\delta_0=0$ (no endogenous regressor) or $\mathbb{E}\left(\tilde{W}_iZ_i'\right)=0$. The general two-step / generated-regressor template is exactly that of 8: stack the two moment conditions and read off the joint influence function.

归属：ECON 510（原卷标签，旧考法） **重要度**：中 **再考判断**：中
核心考点：内生 OLS 的概率极限与比较
历史复现：2019；IV 识别与弱工具见 2023–2025
掌握方式：从 reduced form 代入后现场算 plim

原题 ECON 510 Q2（verbatim）

(2) [**40 pts**] Suppose we ignore the endogeneity of Y_{2i} and estimate θ_0 by OLS of the structural equation (1); call it $\widehat{\theta}^{OLS}$.

(i) [**20 pts**] Using (2), describe the probability limit of $\widehat{\theta}^{OLS}$, and label it θ^*.

(ii) [**20 pts**] Once again, using (2), describe the linear representation of $\widehat{\theta}^{OLS}$.

Note: parts (i)–(ii) require you to use the reduced-form equation (2).

Solution.

中文思路。直接对 (1) 跑 OLS，把 Y_{2i} 当外生。回归元 $R_i=(X_i',Y_{2i})'$ 与误差 ε_i 不正交（因为 $\mathbb{E}(Y_{2i}\varepsilon_i)\neq 0$），所以 OLS 不再一致——它收敛到一个被污染的 θ^*。题目特意要求“用 (2) 表达”：意思是把 $\mathbb{E}(Y_{2i}\varepsilon_i)$ 用 η_i 写出来，$\mathbb{E}(Y_{2i}\varepsilon_i)=\mathbb{E}((Z_i'\gamma_0+\eta_i)\varepsilon_i)=\mathbb{E}(\eta_i\varepsilon_i)$（因为 $\mathbb{E}(Z_i\varepsilon_i)=0$）。这样偏差就只剩 $\sigma_{\eta\varepsilon}=\mathbb{E}(\eta_i\varepsilon_i)$ 这一个标量在驱动。

Setup. Stack the structural regressors as $R_i\equiv(X_i',Y_{2i})'$, so (1) is $Y_{1i}=R_i'\theta_0+\varepsilon_i$ and $\widehat{\theta}^{OLS}=\left(\frac{1}{n}\sum_{i=1}^n R_iR_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n R_iY_{1i}$. Let $Q_R\equiv\mathbb{E}(R_iR_i')$ (nonsingular).

(i) Probability limit. By the LLN and the continuous-mapping/Slutsky argument of 4,

$$\widehat{\theta}^{OLS}\xrightarrow{p}\theta^*=\theta_0+Q_R^{-1}\,\mathbb{E}(R_i\varepsilon_i)\,.$$

Now $\mathbb{E}(R_i\varepsilon_i)=\left(\mathbb{E}(X_i\varepsilon_i)',\ \mathbb{E}(Y_{2i}\varepsilon_i)\right)'=\left(0',\ \mathbb{E}(Y_{2i}\varepsilon_i)\right)'$ since X_i is exogenous. Using (2) and $\mathbb{E}(Z_i\varepsilon_i)=0$,

$$\mathbb{E}(Y_{2i}\varepsilon_i)=\mathbb{E}\big((Z_i'\gamma_0+\eta_i)\varepsilon_i\big)=\gamma_0'\underbrace{\mathbb{E}(Z_i\varepsilon_i)}_{=0}+\mathbb{E}(\eta_i\varepsilon_i)=\sigma_{\eta\varepsilon},\qquad \sigma_{\eta\varepsilon}\equiv\mathbb{E}(\eta_i\varepsilon_i)\,.$$

Hence, with $e = (0'_{k_x}, 1)'$,

$$\boxed{\theta^* = \theta_0 + \sigma_{\eta\varepsilon}\, Q_R^{-1} e = \theta_0 + \mathbb{E}(\eta_i \varepsilon_i)\ \big(\mathbb{E}(R_i R_i')\big)^{-1} (0, \ldots, 0, 1)'.}$$

偏差方向是 $Q_R^{-1}e$，大小是 $\sigma_{\eta\varepsilon}$。当 $\sigma_{\eta\varepsilon} = 0$（即 η 与 ε 不相关、Y_{2i} 其实外生）时 $\theta^* = \theta_0$，OLS 一致——和直觉吻合。注意偏差一般会“漏”进 β 分量，不只污染 δ，因为 Q_R^{-1} 不是对角的。

(ii) Linear (asymptotically linear) representation. Center at the (biased) plim θ^*, not at θ_0. Define the population residual at θ^*:

$$u_i^* \equiv Y_{1i} - R_i'\theta^* = \varepsilon_i - R_i'(\theta^* - \theta_0) = \varepsilon_i - \sigma_{\eta\varepsilon}\, R_i' Q_R^{-1} e,$$

which satisfies $\mathbb{E}(R_i u_i^*) = \mathbb{E}(R_i \varepsilon_i) - \sigma_{\eta\varepsilon}\mathbb{E}(R_i R_i')\, Q_R^{-1} e = \sigma_{\eta\varepsilon} e - \sigma_{\eta\varepsilon} e = 0$ by construction. Then by the OLS expansion of 4,

$$\boxed{\sqrt{n}(\widehat{\theta}^{OLS} - \theta^*) = Q_R^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_i u_i^* + o_p(1) \ \xrightarrow{d}\ N\big(0,\ Q_R^{-1}\, \mathbb{E}\big(u_i^{*2} R_i R_i'\big)\ Q_R^{-1}\big),}$$

with influence function $\psi_i^{OLS} = Q_R^{-1} R_i u_i^*$ (mean zero) and $u_i^* = \varepsilon_i - \sigma_{\eta\varepsilon} R_i' Q_R^{-1} e$ written via (2) through $\sigma_{\eta\varepsilon} = \mathbb{E}(\eta_i \varepsilon_i)$.

Remark (与 (i) 的衔接).

> (ii) 的影响函数中心在 θ^*（不是 θ_0）——这是“OLS 在内生时”这类题最容易扣分的地方：错误地以为 $\sqrt{n}(\widehat{\theta} - \theta_0)$ 渐近正态会发散（其实 $\widehat{\theta} - \theta_0 \xrightarrow{p} \theta^* - \theta_0 \neq 0$，乘 $\sqrt{n}$ 爆掉）。只有围绕 θ^* 居中、用 u_i^* 而非 ε_i，才是合法的渐近线性表示。

3.2 Comprehensive ·Guggenberger：501 基础六题

[优先级 ●●● 高 High] 资格考足迹：2019(Comp)

归属：ECON 501（按内容归类） **重要度**：中高 **再考判断**：中高
核心考点：联合密度、边缘/条件密度与条件期望
历史复现：2019、2021、2024
掌握方式：熟练积分与相除；不背具体常数

原题 Comp Q1（verbatim, 9 pts）

Let the joint density of X and Y be

$$f(x, y) = \begin{cases} \dfrac{2}{(1 + x + y)^3} & 0 < x < \infty,\ 0 < y < \infty, \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Compute the marginal density of X. **(b)** Compute the conditional density of Y given $X = x$ for $x > 0$. **(c)** Compute $\mathbb{E}(Y \mid X = x)$ for $x > 0$.

Solution.

中文思路。纯积分练习（见 2 的边缘/条件密度）。三步：积掉 y 得边缘；相除得条件；再积 $y \cdot f_{Y|X}$ 得条件期望。技巧：对 $\int(1+x+y)^{-3}\,dy$ 直接当 $u = 1+x+y$ 处理。

(a) For $x > 0$, integrate out y:

$$f_X(x) = \int_0^\infty \frac{2}{(1+x+y)^3}\,dy = 2\Big[-\tfrac{1}{2}(1+x+y)^{-2}\Big]_{y=0}^{\infty} = \frac{1}{(1+x)^2}.$$

$$\boxed{f_X(x) = \frac{1}{(1+x)^2}, \qquad x > 0.}$$

(b) For $x > 0$, $y > 0$,

$$f_{Y\,|\,X}(y\,|\,x) = \frac{f(x,y)}{f_X(x)} = \frac{2/(1+x+y)^3}{1/(1+x)^2} = \boxed{\frac{2(1+x)^2}{(1+x+y)^3}}.$$

(c) Let $a \equiv 1+x > 0$ and substitute $u = a + y$:

$$\mathbb{E}(Y\,|\,X=x) = \int_0^\infty y\,\frac{2a^2}{(a+y)^3}\,dy = 2a^2\int_a^\infty \frac{u-a}{u^3}\,du = 2a^2\Big(\tfrac{1}{a} - a\cdot\tfrac{1}{2a^2}\Big) = 2a^2\cdot\frac{1}{2a} = a.$$

$$\boxed{\mathbb{E}(Y\,|\,X=x) = 1+x, \qquad x > 0.}$$

检验：$\mathbb{E}(Y\,|\,X=x)$ 随 x 线性增长，说明 X, Y 正相关——和联合密度沿 $x+y$ 递减、质量集中在小 $x+y$ 区域的形状一致。

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：推前测度与 LOTUS 的测度论证明
历史复现：2017、2018、2019、2024
掌握方式：逐行背四步标准机器

原题 Comp Q2（verbatim, 9 pts）

Suppose X is a random p-vector on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and $g : \mathbb{R}^p \to \mathbb{R}$ is a measurable function such that $\int |g(x)|\,d\mathrm{P} < \infty$. Recall $\mathbb{E}(g(X)) = \int g(x)\,d\mathrm{P}$. Show that $\mathbb{E}(g(X)) = \int g\,d\mathrm{P}_X$, where P_X is the distribution (push-forward) of X.

Solution.

中文思路。这是“无意识统计学家定律”（LOTUS）的测度论证明，标准套路是 1 的**标准机器**：先对示性函数验，再线性扩到简单函数，再单调收敛扩到非负可测，最后拆正负部到一般可积函数。核心恒等式只有一条：推前测度的定义 $\mathrm{P}_X(B) = \mathrm{P}(X^{-1}(B))$。

Recall the push-forward $\mathrm{P}_X(B) = \mathrm{P}(X \in B) = \mathrm{P}(X^{-1}(B))$ for Borel $B \subseteq \mathbb{R}^p$. We prove $\int_\Omega g(X)\,d\mathrm{P} = \int_{\mathbb{R}^p} g\,d\mathrm{P}_X$ by the standard machine (see 1).

Proof for Theorem

[STRUCTURE —know the steps, fudge details]

Step 1 (indicators). Let $g = \mathbb{1}_B$ for Borel B. Then $g(X) = \mathbb{1}_B(X) = \mathbb{1}_{\{X\in B\}}$, so

$$\int_\Omega \mathbb{1}_B(X)\,d\mathrm{P} = \mathrm{P}(X \in B) = \mathrm{P}_X(B) = \int_{\mathbb{R}^p} \mathbb{1}_B\,d\mathrm{P}_X,$$

which is exactly the definition of P_X.

Step 2 (simple functions). Let $g = \sum_{k=1}^m a_k \mathbb{1}_{B_k}$ be nonnegative simple. By linearity of the integral and Step 1,

$$\int_\Omega g(X)\,d\mathrm{P} = \sum_k a_k\,\mathrm{P}_X(B_k) = \int_{\mathbb{R}^p} g\,d\mathrm{P}_X.$$

Step 3 (nonnegative measurable). For $g \geq 0$ measurable, take simple $0 \leq g_n \uparrow g$ pointwise. Then $g_n(X) \uparrow g(X)$ pointwise on Ω. By the Monotone Convergence Theorem applied on both sides and Step 2,

$$\int_\Omega g(X)\,d\mathrm{P} = \lim_n \int_\Omega g_n(X)\,d\mathrm{P} = \lim_n \int_{\mathbb{R}^p} g_n\,d\mathrm{P}_X = \int_{\mathbb{R}^p} g\,d\mathrm{P}_X.$$

Step 4 (general integrable). For general measurable g with $\int |g(X)|\,d\mathrm{P} < \infty$, split $g = g^+ - g^-$. By Step 3, $\int_\Omega g^\pm(X)\,d\mathrm{P} = \int_{\mathbb{R}^p} g^\pm\,d\mathrm{P}_X < \infty$ (finiteness from $\int |g|\,d\mathrm{P}_X = \int |g(X)|\,d\mathrm{P} < \infty$, itself Step 3 applied to $|g|$). Subtracting,

$$\mathbb{E}(g(X)) = \int_\Omega g(X)\,d\mathrm{P} = \int_{\mathbb{R}^p} g^+\,d\mathrm{P}_X - \int_{\mathbb{R}^p} g^-\,d\mathrm{P}_X = \int_{\mathbb{R}^p} g\,d\mathrm{P}_X. \qquad \square$$

要点：整道证明只在 Step 1 用到 P_X 的定义，其余全是“标准机器”的机械推进。考场上把四步骨架写清楚即可拿满。

归属：ECON 501（按内容归类）　**重要度**：高（必会）　**再考判断**：高
核心考点：样本均值/方差联合 CLT 与一致性
历史复现：2019、2020、2021、2025
掌握方式：背线性化；能写二维 CLT 的协方差矩阵

原题 Comp Q3（verbatim, 12 pts）

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2, and let $C = \sigma/\mu$. Define $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, and $\widehat{C}_n = \hat{\sigma}_n/\bar{X}_n$. **(a)** Establish consistency of $\widehat{C}_n$ for C (you may use the LLN, rules for convergence in probability, and Slutsky). State assumptions. **(b)** Find the asymptotic distribution of $n^{1/2}\big(\bar{X}_n - \mu,\ \hat{\sigma}_n^2 - \sigma^2\big)'$. State assumptions.

Solution.

中文思路。 (a) 是 LLN + 连续映射/Slutsky 的标准一致性（4）；唯一要点是分母不能为 0，所以假设 $\mu \neq 0$，再加二阶矩有限。(b) 是样本均值与样本方差的**联合** CLT：先把 $\hat{\sigma}_n^2$ 换成以 μ 为中心的版本（差一个 $O_p(1/n)$ 可忽略），对二维向量 $\big(X_i - \mu, (X_i - \mu)^2 - \sigma^2\big)$ 直接上多元 CLT，协方差里出现三阶、四阶中心矩，所以需要四阶矩有限。

Assumptions. X_i i.i.d.; $\mu \neq 0$ and $\mathbb{E}\big(X_i^2\big) < \infty$ for (a); additionally $\mathbb{E}\big(X_i^4\big) < \infty$ for (b). Write central moments $\mu_3 = \mathbb{E}\big((X_i - \mu)^3\big)$, $\mu_4 = \mathbb{E}\big((X_i - \mu)^4\big)$.

(a) Consistency. By the WLLN, $\bar{X}_n \xrightarrow{p} \mu$ and $n^{-1}\sum_{i=1}^n X_i^2 \xrightarrow{p} \mathbb{E}\big(X_i^2\big)$. Since $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n X_i^2 - \bar{X}_n^2 \xrightarrow{p} \mathbb{E}\big(X_i^2\big) - \mu^2 = \sigma^2$ (CMT on the difference), and the square root is continuous on $[0,\infty)$, $\hat{\sigma}_n \xrightarrow{p} \sigma$. With $\mu \neq 0$, $\bar{X}_n \xrightarrow{p} \mu \neq 0$, so by Slutsky

$$\widehat{C}_n = \frac{\hat{\sigma}_n}{\bar{X}_n} \xrightarrow{p} \frac{\sigma}{\mu} = C. \qquad \boxed{\widehat{C}_n \xrightarrow{p} C.}$$

(b) Joint CLT. First, $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n (X_i - \mu)^2 - (\bar{X}_n - \mu)^2$, and $(\bar{X}_n - \mu)^2 = O_p(1/n)$, so $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \sqrt{n}\big(n^{-1}\sum_{i=1}^n[(X_i - \mu)^2 - \sigma^2]\big) + o_p(1)$. Apply the multivariate CLT (4) to the i.i.d., mean-zero vectors $V_i = \big(X_i - \mu,\ (X_i - \mu)^2 - \sigma^2\big)'$:

$$\sqrt{n}\begin{pmatrix}\bar{X}_n - \mu \\ \hat{\sigma}_n^2 - \sigma^2\end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_i + o_p(1) \xrightarrow{d} N(0, \Sigma), \qquad \Sigma = \mathbb{E}\big(V_i V_i'\big).$$

The entries of Σ:

$$\begin{aligned}
\Sigma_{11} &= \mathrm{Var}(X_i) = \sigma^2, \\
\Sigma_{12} &= \mathbb{E}\big((X_i - \mu)\big[(X_i - \mu)^2 - \sigma^2\big]\big) = \mu_3, \\
\Sigma_{22} &= \mathrm{Var}\big((X_i - \mu)^2\big) = \mu_4 - \sigma^4.
\end{aligned}$$

Hence

$$\boxed{\sqrt{n}\begin{pmatrix}\bar{X}_n - \mu \\ \hat{\sigma}_n^2 - \sigma^2\end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4\end{pmatrix}\right).}$$

若分布对称，$\mu_3 = 0$，均值与方差渐近独立——这是常考的特例（如正态时 $\mu_3 = 0$，$\mu_4 = 3\sigma^4$，给出 $\Sigma = \mathrm{diag}(\sigma^2, 2\sigma^4)$）。注意除以 n 还是 $n-1$ 不影响极限。

归属：ECON 501（按内容归类）　**重要度**：中高　**再考判断**：中高
核心考点：只有一阶矩时的截断 WLLN
历史复现：2019；重尾与收敛率见 2023、2025
掌握方式：理解先固定截断再放宽截断的双重极限

原题 Comp Q4（verbatim, 10 pts）

Let $X_1, \ldots, X_n$ be i.i.d. mean-zero random variables with a distribution symmetric about zero and variance not necessarily finite. Prove the law of large numbers by a truncation argument, using

$$\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i \mathbb{1}(|X_i| \le c) + n^{-1}\sum_{i=1}^{n} X_i \mathbb{1}(|X_i| > c), \qquad c > 0.$$

Solution.

中文思路。方差可能无穷，不能直接上 L^2 弱大数。截断法（4 的截断 WLLN）：固定 c 把 X 切成“截断部分”（有界、二阶矩有限，用 Chebyshev 控制波动）和“尾部”（用可积性把它的均值压小）。关键资源是**均值零**（题目说 mean-zero）$\Rightarrow$ **读作** $\mathbb{E}(X) = 0$，故 $\mathbb{E}(|X|) < \infty$（一阶绝对矩有限），于是 $\mathbb{E}(|X|\mathbb{1}(|X| > c)) \to 0$（控制收敛）。对称性让截断部分自动中心化（$\mathbb{E}(X\mathbb{1}(|X| \le c)) = 0$），证明更干净。目标：$\bar{X}_n \xrightarrow{p} 0$。

Goal. $\bar{X}_n \xrightarrow{p} 0$. Fix $\varepsilon > 0$; we show $\mathrm{P}\big(|\bar{X}_n| > \varepsilon\big) \to 0$.

Proof for Theorem

[REPRODUCE —memorize this proof]

Decomposition. For $c > 0$ write $X_i = Y_i + R_i$ with $Y_i = X_i \mathbb{1}(|X_i| \le c)$, $R_i = X_i \mathbb{1}(|X_i| > c)$, so $\bar{X}_n = \bar{Y}_n + \bar{R}_n$.

Symmetry centers each part. Symmetry about 0 makes $x \mapsto x\mathbb{1}(|x| \le c)$ and $x \mapsto x\mathbb{1}(|x| > c)$ odd integrable functions, hence

$$\mathbb{E}(Y_i) = \mathbb{E}(X_i \mathbb{1}(|X_i| \le c)) = 0,$$
$$\mathbb{E}(R_i) = \mathbb{E}(X_i \mathbb{1}(|X_i| > c)) = 0.$$

(We read "mean-zero" as $\mathbb{E}(X_i) = 0$ in the usual sense, i.e. $\mathbb{E}\left(X_i^+\right)$ and $\mathbb{E}\left(X_i^-\right)$ are both finite and equal; equivalently $\mathbb{E}(|X_i|) < \infty$ with $\mathbb{E}(X_i) = 0$. This first-absolute-moment finiteness is exactly what makes Y_i, R_i integrable and drives the tail control below. The symmetric infinite-mean case—e.g. Cauchy, where $\mathbb{E}(|X_i|) = \infty$ and no centering is even defined—is therefore excluded by hypothesis; the truncation WLLN as stated relies on $\mathbb{E}(|X_i|) < \infty$.)

Truncated part —Chebyshev. $|Y_i| \le c$, so $\operatorname{Var}(Y_i) \le \mathbb{E}\left(Y_i^2\right) \le c^2 < \infty$. By independence and Chebyshev,

$$\mathrm{P}\left(|\bar{Y}_n| > \varepsilon/2\right) \le \frac{\operatorname{Var}\left(\bar{Y}_n\right)}{(\varepsilon/2)^2} = \frac{\operatorname{Var}(Y_1)}{n(\varepsilon/2)^2} \le \frac{4c^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Tail part —integrability. Using $\mathbb{E}(R_i) = 0$ and Markov on $|\bar{R}_n|$ via the triangle bound $\mathbb{E}\left(|\bar{R}_n|\right) \le \mathbb{E}(|R_1|) = \mathbb{E}(|X_1|\mathbb{1}(|X_1| > c))$,

$$\mathrm{P}\left(|\bar{R}_n| > \varepsilon/2\right) \le \frac{\mathbb{E}\left(|\bar{R}_n|\right)}{\varepsilon/2} \le \frac{2}{\varepsilon}\,\mathbb{E}(|X_1|\mathbb{1}(|X_1| > c))\,.$$

Since $\mathbb{E}(|X_1|) < \infty$, dominated convergence gives $\delta(c) \equiv \mathbb{E}(|X_1|\mathbb{1}(|X_1| > c)) \to 0$ as $c \to \infty$.

Combine (double limit). For any $\varepsilon > 0$,

$$\mathrm{P}\left(|\bar{X}_n| > \varepsilon\right) \le \mathrm{P}\left(|\bar{Y}_n| > \varepsilon/2\right) + \mathrm{P}\left(|\bar{R}_n| > \varepsilon/2\right) \le \frac{4c^2}{n\varepsilon^2} + \frac{2}{\varepsilon}\,\delta(c).$$

Given $\epsilon' > 0$, first choose c large so that $\frac{2}{\varepsilon}\delta(c) < \epsilon'/2$; then for that fixed c choose N so that $n \ge N \Rightarrow \frac{4c^2}{n\varepsilon^2} < \epsilon'/2$. Hence $\limsup_n \mathrm{P}\left(|\bar{X}_n| > \varepsilon\right) \le \epsilon'$ for every $\epsilon' > 0$, so $\mathrm{P}\left(|\bar{X}_n| > \varepsilon\right) \to 0$. Therefore $\bar{X}_n \xrightarrow{p} 0$. ■

Remark (符号 vs. 一般情形).

对称性是“免费午餐”：它直接给出 $\mathbb{E}(Y_i) = 0$，省掉了“截断均值 $\to 0$”的额外论证。若不对称，需多一步证 $\mathbb{E}(X\mathbb{1}(|X| \le c)) \to \mathbb{E}(X) = 0$ 并把它从 $\bar{Y}_n$ 中减去。先固定 c 让 $n \to \infty$、再让 $c \to \infty$ 的“双重极限”次序是这类证明的命门。

归属：ECON 501（按内容归类） **重要度**：高（必会） **再考判断**：高
核心考点：Binomial MLE、NP、MLR 与单边 UMP
历史复现：2017、2019、2023、2025
掌握方式：MLE/CLT 可现场推；NP 与 Karlin–Rubin 骨架需背

原题 Comp Q5（verbatim, 12 pts）

Suppose $X \sim \text{bin}(n,p)$ with pmf $f(x,p) = \binom{n}{x}p^x(1-p)^{n-x}$, $x = 0,\dots,n$, $p \in [0,1]$. One observes a single X. **(a)** MLE of p? **(b)** Asymptotic distribution of the ML estimator as $n \to \infty$? **(c)** State the Neyman–Pearson Lemma. **(d)** Level-α UMP test of $H_0 : p = p_0$ vs. $H_1 : p = p_1$ for $p_1 > p_0$ (simplify; treat the critical value as if continuous). **(e)** Level-α UMP test of $H_0 : p \le p_0$ vs. $H_1 : p > p_0$.

Solution.

中文思路。(a)(b) 是二项 MLE 与其 CLT（5）；(c)(d)(e) 是经典检验（6）：NP 引理 → 单点对单点的似然比检验 → 由 MLR 推广到单边复合假设（Karlin–Rubin）。诀窍：二项似然比关于 X 单调，故拒绝域统一是 $\{X > k\}$。

(a) MLE. Log-likelihood $\ell(p) = \log\binom{n}{X} + X\log p + (n-X)\log(1-p)$. Setting $\ell'(p) = \frac{X}{p} - \frac{n-X}{1-p} = 0$ gives $X(1-p) = (n-X)p \Rightarrow X = np$, so

$$\boxed{\hat{p} = X/n} \quad (\ell'' < 0,\ \text{interior max; boundary handled by } X \in \{0, n\}).$$

(b) Asymptotics. $X = \sum_{i=1}^n B_i$ with $B_i \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$, so $\hat{p} = \bar{B}_n$. By the CLT,

$$\boxed{\sqrt{n}(\hat{p} - p) \xrightarrow{d} N\big(0,\ p(1-p)\big),}$$

matching $I(p)^{-1} = p(1-p)$ since the Fisher information per observation is $I(p) = \frac{1}{p(1-p)}$.

(c) Neyman–Pearson Lemma. For testing simple $H_0 : \theta = \theta_0$ vs. simple $H_1 : \theta = \theta_1$ with densities f_0, f_1, consider the likelihood-ratio test that rejects when $f_1(x)/f_0(x) > k$ (and randomizes on $= k$) chosen so that the size is exactly α. Then: *(Existence)* such a test of size α exists; *(Sufficiency)* any test of this LR form is most powerful among all level-α tests; *(Necessity)* any most powerful level-α test must be of this LR form (a.e. where $f_0 + f_1 > 0$), up to behavior on $\{f_1 = kf_0\}$.

(d) Simple-vs-simple ($p_1 > p_0$). The likelihood ratio is

$$\Lambda(X) = \frac{f(X,p_1)}{f(X,p_0)} = \left(\tfrac{1-p_1}{1-p_0}\right)^n \left(\tfrac{p_1(1-p_0)}{p_0(1-p_1)}\right)^X.$$

Since $p_1 > p_0$ implies $\frac{p_1(1-p_0)}{p_0(1-p_1)} > 1$, $\Lambda(X)$ is strictly increasing in X. So $\{\Lambda > k\} =$

$\{X > c\}$ for some c. By NP, the level-α MP test is

$$\boxed{\phi(X) = \mathbb{1}\{X > c\}, \quad c \text{ chosen so that } \mathrm{P}_{p_0}(X > c) = \alpha.}$$

(Treating the critical value as continuous, no randomization is needed.)

(e) One-sided composite. The binomial family has *monotone likelihood ratio* in X (the ratio $f(x, p')/f(x, p)$ is increasing in x whenever $p' > p$). By the Karlin–Rubin theorem, the test

$$\boxed{\phi(X) = \mathbb{1}\{X > c\} \text{ with } \mathrm{P}_{p_0}(X > c) = \alpha}$$

is UMP level-α for $H_0 : p \le p_0$ vs. $H_1 : p > p_0$. 同一个拒绝域：MLR 保证 (d) 的临界值不依赖具体 p_1，又保证功效在 H_1 上一致最优、且第一类错误率在 $p \le p_0$ 上由 $p = p_0$ 处取到上确界 α。

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：一阶退化时的二阶 delta 法
历史复现：2019、2022、2025 的渐近题同族
掌握方式：记住率从 $\sqrt{n}$ 变为 n，并现场 Taylor 展开

原题 Comp Q6（verbatim, 8 pts）

Let Y_n satisfy $\sqrt{n}(Y_n - \theta) \xrightarrow{d} Z \sim N(0, \sigma^2)$ as $n \to \infty$, for scalars θ and $\sigma^2 > 0$. Let $g \in C^2(U)$ on an open neighborhood U of θ, scalar-valued, with $g'(\theta) = 0$. Work out the limiting distribution of $g(Y_n)$ (first find the correct normalization and centering). State any additional assumptions.

Solution.

中文思路。$g'(\theta) = 0$ 时一阶 delta 法退化（一阶项消失），要用**二阶 delta 法**（见 4）。一阶项没了，主导项变成二阶：$g(Y_n) - g(\theta) \approx \frac{1}{2} g''(\theta)(Y_n - \theta)^2$。$(Y_n - \theta)^2 = O_p(1/n)$，所以正确的放大率是 n（不是 $\sqrt{n}$），极限是 χ^2_1 的标量倍。额外假设：$g''(\theta) \neq 0$（否则二阶也退化，极限退化为点质量 0）。

Additional assumption. $g''(\theta) \neq 0$ (otherwise the leading term vanishes and $n\,(g(Y_n) - g(\theta)) \xrightarrow{p} 0$).

Derivation. A second-order Taylor expansion of $g \in C^2$ about θ gives, for Y_n near θ,

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \tfrac{1}{2} g''(\tilde{\theta}_n)(Y_n - \theta)^2,$$

with $\tilde{\theta}_n$ between Y_n and θ. Since $g'(\theta) = 0$,

$$n\big(g(Y_n) - g(\theta)\big) = \tfrac{1}{2}\, g''(\tilde{\theta}_n) \left[\sqrt{n}(Y_n - \theta)\right]^2.$$

From $\sqrt{n}(Y_n - \theta) \xrightarrow{d} Z$ we get $Y_n \xrightarrow{p} \theta$, so $\tilde{\theta}_n \xrightarrow{p} \theta$ and by continuity of g'',

$g''(\tilde{\theta}_n) \xrightarrow{p} g''(\theta)$. By the CMT, $[\sqrt{n}(Y_n - \theta)]^2 \xrightarrow{d} Z^2 = \sigma^2\chi^2_1$. By Slutsky,

$$\boxed{n\big(g(Y_n) - g(\theta)\big) \ \xrightarrow{d}\ \tfrac{1}{2}\, g''(\theta)\, Z^2 \ \overset{d}{=}\ \tfrac{1}{2}\, g''(\theta)\, \sigma^2\, \chi^2_1.}$$

形状解读: 极限是 χ^2_1 的标量倍，故**不再渐近正态**，且当 $g''(\theta) > 0$ 时支撑在 $[0, \infty)$（$Y_n = \theta$ 是 g 的局部极小，偏离一律使 g 增大）；$g''(\theta) < 0$ 时支撑在 $(-\infty, 0]$。这正是"一阶导为零"的特征信号——记住放大率从 $\sqrt{n}$ 升到 n。

Chapter 4　2020 资格考精解（Worked Solutions）

导读 · Roadmap（先读这里，不含公式）

2020 年是**单卷十二题**：Q1–Q7 是 501 风格的概率与渐近（条件期望即 L^2 投影、随机变量变换、阶统计量与概率积分变换、用联合 MGF 定正态、混合/Roy 识别、样本方差与样本均值平方的极限、五种收敛模式辨析、独立/均值独立/不相关三层例子），Q8–Q11 是 510 风格（参数相关支撑下的截断回归、固定阈值有序 probit、矩条件的 rank 识别、乘性异方差下的最优工具），Q12 是一道玩笑题。

本章每题先给**中文思路**（点出该用哪一把刀），再给 **English formal derivation** 与 **boxed final answer**。特别注意两处反直觉陷阱：Q8 的截断 MLE 是非正则边界问题，Q10 的矩条件虽然数量多却仍然降秩欠识别。需要的定理（delta 法、CMT、Markov、最优工具公式）都在前面章节，本章只**指向并套用**，不重讲。

4.1　Q1 —条件期望（L^2 投影）、变换、与残差的相关

归属：ECON 501（按内容归类）　**重要度**：高　**再考判断**：高
核心考点：L^2 投影、变量变换与残差正交
历史复现：2020、2021、2024；条件期望多年高频
掌握方式：投影正交式要背；密度变换逐支求和

原题 Q1 (verbatim)

Suppose that (X, Y) has a joint probability density function $f_{XY}(x, y) = 8xy\,\mathbb{1}(0 \le x \le y \le 1)$.

(a) Obtain the function $g : (0, 1) \to \mathbb{R}$ such that for any measurable function $h : (0, 1) \to \mathbb{R}$, we have $\mathbb{E}\Big(\{Y - g(X)\}^2\Big) \le \mathbb{E}\Big(\{Y - h(X)\}^2\Big)$.

(b) Let $Z = X^2$. Obtain the probability density function of Z.

(c) Compute the correlation coefficient between $Y - g(X)$ and Z, where g and Z are defined in parts (a) and (b) respectively.

Solution.

思路：(a) 问的「对任意 h 都最小化 MSE 的 g」按定义就是条件期望 $g(x) = \mathbb{E}(Y \mid X = x)$——这正是 3 里讲的「条件期望 = L^2 最优预测」。先在三角形支撑 $x \le y \le 1$ 上对 y 积出 X 的边缘密度，再求条件密度。(b) 单调变换 $z = x^2$，套换元公式 (2)。(c) 关键不是硬算，而是看出 $Z = X^2$ 是 X 的函数，而残差 $Y - g(X)$ 对 X 条件均值为零——于是协方差为 0，相关系数为 0。

(a) Conditional mean as the L^2-optimal predictor. By the projection characterization (3), the minimizer over all measurable h is $g(x) = \mathbb{E}(Y \mid X = x)$. Compute the marginal of X: for $x \in (0,1)$, integrating y over $[x, 1]$,

$$f_X(x) = \int_x^1 8xy\,\mathrm{d}y = 8x \cdot \tfrac{1-x^2}{2} = 4x(1-x^2), \qquad x \in (0,1).$$

Hence the conditional density on $y \in [x,1]$ is

$$f_{Y \mid X}(y \mid x) = \frac{8xy}{4x(1-x^2)} = \frac{2y}{1-x^2},$$

and

$$g(x) = \mathbb{E}(Y \mid X = x) = \int_x^1 y \cdot \frac{2y}{1-x^2}\,\mathrm{d}y = \frac{2}{1-x^2} \cdot \frac{1-x^3}{3} = \boxed{\frac{2}{3}\,\frac{1-x^3}{1-x^2}} = \frac{2}{3}\,\frac{1+x+x^2}{1+x}.$$

(b) Density of $Z = X^2$. The map $z = x^2$ is strictly increasing on $(0,1)$ with $x = \sqrt{z}$, $\frac{\mathrm{d}x}{\mathrm{d}z} = \frac{1}{2\sqrt{z}}$. By the change-of-variables formula (2),

$$f_Z(z) = f_X(\sqrt{z}) \left| \frac{\mathrm{d}x}{\mathrm{d}z} \right| = 4\sqrt{z}\,(1-z) \cdot \frac{1}{2\sqrt{z}} = \boxed{2(1-z), \quad z \in (0,1)}$$

i.e. $Z \sim \mathrm{Beta}(1,2)$.

(c) Correlation of the residual with Z. Let $R := Y - g(X)$. By construction $\mathbb{E}(R \mid X) = \mathbb{E}(Y \mid X) - g(X) = 0$ almost surely. Since $Z = X^2$ is a (measurable) function of X, the tower property gives

$$\mathrm{Cov}(R, Z) = \mathbb{E}(R\,Z) - \mathbb{E}(R)\,\mathbb{E}(Z) = \mathbb{E}(Z\,\mathbb{E}(R \mid X)) - 0 = \mathbb{E}(Z \cdot 0) = 0.$$

Therefore the correlation coefficient is

$$\mathrm{Corr}(R, Z) = \frac{\mathrm{Cov}(R, Z)}{\sqrt{\mathrm{Var}(R)\,\mathrm{Var}(Z)}} = \boxed{0}$$

(provided $\mathrm{Var}(R) > 0$ and $\mathrm{Var}(Z) > 0$, which hold here).

要点：根本不用算 $\mathrm{Var}(R)$ 或 $\mathbb{E}(RZ)$ 的具体数值。「残差与任何 X 的函数不相关」是条件期望/L^2 投影的正交性的直接推论——考场上写出 $\mathbb{E}(R \mid X) = 0 \Rightarrow \mathrm{Cov}(R, m(X)) = 0$ 即可拿满分。

4.2 Q2 —指数分布（单位风险率）、最小值、概率积分变换

归属：ECON 501（按内容归类） **重要度**：中高 **再考判断**：中高
核心考点：危险率刻画指数分布、最小值与 PIT
历史复现：2020、2021、2025 的指数分布题同族
掌握方式：背 exponential memoryless/hazard；order statistic 现场推

原题 Q2 (verbatim)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Suppose that $\mathrm{P}(X_1 \le 0) = 0$ and $\lim_{h \downarrow 0} \mathrm{P}(X_1 \le x + h \mid X_1 > x)/h = 1$ for all $x > 0$. Further, let $X_{(j)}$ be the jth smallest variable among $X_1, \ldots, X_n$.

(a) Please compute $\mathbb{E}(X_1)$ and $\mathbb{E}\big(X_{(1)}\big)$.

(b) Let F be the CDF of X_1. Please derive the CDF and the pdf of $Y_{(j)} := F(X_{(j)})$ for $j = 1, 2, \ldots, n$.

Solution.

思路：(a) 给的极限就是**风险率** $\lambda(x) = f(x)/\bar{F}(x) \equiv 1$。支撑在 $(0, \infty)$ 上的常值单位风险率唯一刻画 $\mathrm{Exp}(1)$。n 个独立 $\mathrm{Exp}(1)$ 的最小值是 $\mathrm{Exp}(n)$。(b) 概率积分变换：$U_i := F(X_i) \sim \mathrm{Unif}(0, 1)$，其第 j 阶统计量是 $\mathrm{Beta}(j, n-j+1)$。这两点都在 2 的阶统计量一节。

(a) The hazard rate is

$$\lambda(x) = \lim_{h \downarrow 0} \frac{\mathrm{P}(x < X_1 \le x + h)}{h\ \mathrm{P}(X_1 > x)} = \lim_{h \downarrow 0} \frac{1}{h} \mathrm{P}(X_1 \le x + h \mid X_1 > x) = 1 \quad (x > 0).$$

With $\mathrm{P}(X_1 \le 0) = 0$, the survival function solves $\frac{\mathrm{d}}{\mathrm{d}x}\big[-\log \bar{F}(x)\big] = \lambda(x) = 1$, $\bar{F}(0) = 1$, so $\bar{F}(x) = e^{-x}$ and $X_1 \sim \mathrm{Exp}(1)$. Hence

$$\boxed{\mathbb{E}(X_1) = 1.}$$

For the minimum, $\mathrm{P}\big(X_{(1)} > t\big) = \mathrm{P}(X_1 > t)^n = e^{-nt}$, so $X_{(1)} \sim \mathrm{Exp}(n)$ and

$$\boxed{\mathbb{E}\big(X_{(1)}\big) = \frac{1}{n}.}$$

(b) By the probability integral transform, $U_i := F(X_i) \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, 1)$, and because F is nondecreasing, $Y_{(j)} = F(X_{(j)}) = U_{(j)}$, the jth order statistic of n

i.i.d. uniforms. Thus $Y_{(j)} \sim \mathrm{Beta}(j, n-j+1)$:

$$\boxed{f_{Y_{(j)}}(y) = \frac{n!}{(j-1)!\,(n-j)!}\,y^{\,j-1}(1-y)^{\,n-j}, \quad y \in (0,1),}$$

with CDF (regularized incomplete beta / binomial tail)

$$F_{Y_{(j)}}(y) = \mathrm{P}\big(Y_{(j)} \le y\big) = \sum_{k=j}^{n} \binom{n}{k} y^k (1-y)^{n-k}, \qquad y \in [0,1].$$

$Y_{(j)}$ 的分布**完全不依赖** F——这正是概率积分变换的威力，也是阶统计量分布理论里所有 distribution-free 结论的根。

4.3 Q3 —用联合 MGF 定正态 + 线性事件下的条件期望

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：MGF 识别正态与截断线性条件期望
历史复现：2020、2021、2024、2025
掌握方式：先做线性变换协方差，再用逆 Mills 公式

原题 Q3 (verbatim)

Suppose that X and Y are random variables with $\mathbb{E}(\exp(tX+sY)) = \exp\{(t^2+s^2)/2\}$ for all $t, s \in \mathbb{R}$. Please compute $\mathbb{E}(2X+Y \mid X+2Y>0)$.

Solution.
思路：联合 MGF $=\exp\{(t^2+s^2)/2\}$ 恰是二元正态的 MGF，对照标准式 $\exp\{\mu' t + \frac{1}{2}t'\Sigma t\}$ 读出 $\mu=0$、$\Sigma = I_2$，即 $X, Y \overset{i.i.d.}{\sim} N(0,1)$。令 $A = 2X+Y$、$B = X+2Y$，二者联合正态。对联合正态用 $\mathbb{E}(A \mid B>0) = \mathbb{E}(A) + \frac{\mathrm{Cov}(A,B)}{\mathrm{Var}(B)}\mathbb{E}(B \mid B>0)$（先把 A 对 B 做线性投影，残差与 B 独立，故与 $\{B>0\}$ 独立）。半正态均值 $\mathbb{E}(B \mid B>0) = \sqrt{2\,\mathrm{Var}(B)/\pi}$。详见 3 多元正态一节。

Identify the law. The MGF $\exp\{\frac{1}{2}(t^2+s^2)\}$ matches the bivariate-normal MGF $\exp\{\mu^\top(t,s) + \frac{1}{2}(t,s)\Sigma(t,s)^\top\}$ with $\mu = 0$ and $\Sigma = I_2$. Hence $X, Y \overset{i.i.d.}{\sim} N(0,1)$ (MGF determines the distribution, 2).

Set up. Let $A = 2X+Y$, $B = X+2Y$. Then (A,B) is jointly normal with

$$\mathbb{E}(A) = \mathbb{E}(B) = 0, \quad \mathrm{Var}(A) = 4{+}1 = 5, \quad \mathrm{Var}(B) = 1{+}4 = 5, \quad \mathrm{Cov}(A,B) = 2{\cdot}1{+}1{\cdot}2 = 4.$$

Project A on B: $A = \dfrac{\mathrm{Cov}(A,B)}{\mathrm{Var}(B)}\,B + \eta = \dfrac{4}{5}B + \eta$, where η is jointly normal with B and $\mathrm{Cov}(\eta, B) = 0$, hence $\eta \perp B$. Therefore

$$\mathbb{E}(A \mid B>0) = \frac{4}{5}\,\mathbb{E}(B \mid B>0) + \mathbb{E}(\eta) = \frac{4}{5}\,\mathbb{E}(B \mid B>0)\,.$$

For $B \sim N(0,5)$, the half-normal mean is $\mathbb{E}(B \mid B>0) = \sqrt{2\,\mathrm{Var}(B)/\pi} = \sqrt{10/\pi}$. Thus

$$\mathbb{E}(2X+Y \mid X+2Y>0) = \frac{4}{5}\sqrt{\frac{10}{\pi}} = \boxed{4\sqrt{\frac{2}{5\pi}}} \approx 1.427.$$

核验：数值 $4\sqrt{2/(5\pi)} \approx 1.427$，与「正相关 ⇒ 条件在 $B>0$ 上 A 期望为正」的符号一致。

4.4 Q4 —从 (Y,T) 识别潜在结果的边缘分布 / 单调性下的联合

归属：ECON 501（按内容归类） **重要度**：中 **再考判断**：中
核心考点：潜在结果边际识别与单调性
历史复现：2020、2021 的识别题；当前 510 仍重识别逻辑
掌握方式：用 observable cells 逐项识别；联合分布不可擅推

原题 Q4 (verbatim)

Suppose that (Y_1, Y_0) and T are independent, where T is Bernoulli. Define $Y := Y_1 T + Y_0(1-T)$.

(a) Can we identify the (marginal) distribution functions of Y_0 and Y_1 from the joint distribution of (Y,T)? Prove or disprove.

(b) Suppose that Y_1 and Y_0 are also Bernoulli and that they are dependent such that $Y_1 \geq Y_0$ with probability one. Can we identify $\mathrm{P}(Y_1 = 1 \mid Y_0 = 0)$ from the joint distribution of (Y,T)? Prove or disprove.

Solution.

思路：这是 Roy/潜在结果模型的识别题，关键全靠 $T \perp (Y_1, Y_0)$。(a) 在 $T=1$ 上观测到的 Y 就是 Y_1，且因独立性，条件分布 = 边缘分布，故两个边缘都识别（只要 $0 < \mathrm{P}(T=1) < 1$）。(b) 加上单调性 $Y_1 \geq Y_0$ 几乎处处，二元 0/1 的联合被两个边缘**唯一**钉死，于是目标概率识别。识别框架见 7。

(a) Yes, both marginals are identified. Assume $0 < p := \mathrm{P}(T=1) < 1$ (otherwise one arm is never observed). For any Borel set C,

$$\mathrm{P}(Y \in C,\ T=1) = \mathrm{P}(Y_1 \in C,\ T=1) = \mathrm{P}(Y_1 \in C)\ \mathrm{P}(T=1),$$

where the last step uses $T \perp (Y_1, Y_0)$. Hence the *observable* left-hand side gives

$$\mathrm{P}(Y_1 \in C) = \frac{\mathrm{P}(Y \in C,\ T=1)}{\mathrm{P}(T=1)} = \mathrm{P}(Y \in C \mid T=1).$$

Symmetrically $\mathrm{P}(Y_0 \in C) = \mathrm{P}(Y \in C \mid T = 0)$. Both are functionals of the observed law of (Y, T), so the marginal CDFs F_{Y_1}, F_{Y_0} are point-identified. $\boxed{\text{Identified.}}$

反例警告：若去掉独立性（仅 $Y = Y_T$），$T = 1$ 上看到的是 $Y_1 \mid T = 1$，未必等于边缘——这正是选择偏差的来源。这里独立性是救命稻草。

(b) Yes, $\mathrm{P}(Y_1 = 1 \mid Y_0 = 0)$ **is identified.** From part (a) the marginals are identified, so we know

$$p_1 := \mathrm{P}(Y_1 = 1)\,, \qquad p_0 := \mathrm{P}(Y_0 = 1)\,.$$

Monotonicity $Y_1 \geq Y_0$ a.s. rules out the cell $(Y_1, Y_0) = (0, 1)$, so the joint pmf is pinned down:

$$\mathrm{P}(Y_1 = 1, Y_0 = 1) = p_0, \quad \mathrm{P}(Y_1 = 1, Y_0 = 0) = p_1 - p_0, \quad \mathrm{P}(Y_1 = 0, Y_0 = 0) = 1 - p_1,$$

(and necessarily $p_1 \geq p_0$). Therefore

$$\mathrm{P}(Y_1 = 1 \mid Y_0 = 0) = \frac{\mathrm{P}(Y_1 = 1, Y_0 = 0)}{\mathrm{P}(Y_0 = 0)} = \boxed{\frac{p_1 - p_0}{1 - p_0}}$$

which is a function of identified quantities, hence identified. $\boxed{\text{Identified.}}$ 直觉：0/1 二元 + 单调性 ⇒ 三种「类型」(always-1、switcher、always-0)，边缘一定就能反推每种的比例。这正是 LATE/单调性识别的离散原型。

4.5 Q5 —样本方差与样本均值平方的极限分布（delta 法）

归属：ECON 501（按内容归类） **重要度**：高（必会） **再考判断**：高
核心考点：样本方差与均值平方的极限分布
历史复现：2019、2020、2021、2025
掌握方式：样本方差线性化和 delta 法都要能白纸写出

原题 Q5 (verbatim)

Suppose that $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}\big(X_1^4\big) < \infty$. Let $\hat{\mu} = n^{-1} \sum_{i=1}^n X_i$ and $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n X_i^2 - \hat{\mu}^2$. You may assume $\sigma^2 > 0$ and $\mathbb{E}\big((X_1 - \mu)^4\big) > \sigma^4$, where $\mu = \mathbb{E}(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1)$.

(a) Derive the (non-degenerate) limit distribution of $n^r(\hat{\sigma}^2 - \sigma^2)$ for the appropriate $r > 0$.

(b) Derive the (non-degenerate) limit distribution of $n^r(\hat{\mu}^2 - \mu^2)$ for the appropriate $r > 0$.

Solution.

思路：两问都是 CLT + delta 法（4）。(a) 把 $\hat{\sigma}^2$ 写成对 (X_i, X_i^2) 样本均值的光滑函数，雅可比一乘即得；标准 $\sqrt{n}$ 率，渐近方差恰是 $\mu_4 - \sigma^4$，题给条件 $\mu_4 > \sigma^4$ 就是保证它非退化。(b) 对 $g(m) = m^2$ 用 delta 法：$g'(\mu) = 2\mu$。$\mu \neq 0$ 时是 $\sqrt{n}$ 率正态；$\mu = 0$ 时一阶项消失**退化**，要升到 $r = 1$，极限是 $\sigma^2\chi_1^2$（二阶 delta 法，见 4 的二阶 delta 法）。

(a) Write $\hat{\sigma}^2 = \widehat{m_2} - \hat{\mu}^2$ with $\widehat{m_2} = n^{-1}\sum_{i=1}^n X_i^2$. A clean route: center at μ. Let $W_i := (X_i - \mu)^2$. Then

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - (\hat{\mu} - \mu)^2 = \bar{W} - (\hat{\mu} - \mu)^2.$$

Since $(\hat{\mu} - \mu)^2 = O_p(n^{-1}) = o_p(n^{-1/2})$, it is asymptotically negligible at the $\sqrt{n}$ scale, and $\mathbb{E}(W_i) = \sigma^2$. By the CLT,

$$\sqrt{n}\,(\hat{\sigma}^2 - \sigma^2) = \sqrt{n}\,(\bar{W} - \sigma^2) + o_p(1) \xrightarrow{d} N\big(0,\ \mathrm{Var}(W_1)\big).$$

Here $\mathrm{Var}(W_1) = \mathbb{E}\big((X_1 - \mu)^4\big) - \sigma^4 = \mu_4 - \sigma^4$, where $\mu_4 := \mathbb{E}\big((X_1 - \mu)^4\big)$. So $r = \frac{1}{2}$ and

$$\boxed{\sqrt{n}\,(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N\big(0,\ \mu_4 - \sigma^4\big),}$$

non-degenerate because $\mu_4 > \sigma^4$ by assumption.

(b) Apply the delta method to $g(m) = m^2$ at $m = \mu$, using $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Case $\mu \neq 0$: $g'(\mu) = 2\mu \neq 0$, so $r = \frac{1}{2}$ and

$$\boxed{\sqrt{n}\,(\hat{\mu}^2 - \mu^2) \xrightarrow{d} N\big(0,\ 4\mu^2\sigma^2\big).}$$

Case $\mu = 0$: the first-order term vanishes; the limit degenerates and one needs $r = 1$. Directly,

$$n\,(\hat{\mu}^2 - 0) = \big(\sqrt{n}\,\hat{\mu}\big)^2 \xrightarrow{d} \big(N(0, \sigma^2)\big)^2 = \sigma^2\,\chi_1^2.$$

$$\boxed{n\,\hat{\mu}^2 \xrightarrow{d} \sigma^2\,\chi_1^2 \quad (\text{when } \mu = 0).}$$

满分要点：(b) 必须**分情形**。$\mu = 0$ 是 delta 法一阶导为零的退化点，率从 $n^{1/2}$ 跳到 n^1、极限从正态变成卡方——这是「二阶 delta 法」的经典考点。

4.6 Q6 —收敛模式辨析（五个判断题）

归属： ECON 501（按内容归类） **重要度：** 高（必会） **再考判断：** 高
核心考点： 收敛模式、O_p、CMT 与 L^2
历史复现： 2020、2021、2022、2024、2025
掌握方式： 每个错误蕴含配一个最短反例；正命题写定理条件

原题 Q6 (verbatim)

Prove or disprove each of the following statements. You may assume that all the random variables are defined on the same probability space.

(a) If $X_n \xrightarrow{a.s.} 0$, then we must have $\mathbb{E}(X_n) \to 0$.

(b) If $\mathbb{E}\left(|X_n|^3\right) = O(n)$, then we must have $X_n = O_p(n^{1/3})$.

(c) If $X_n \xrightarrow{d} X$, then we must have $h(X_n) \xrightarrow{d} h(X)$ whenever h is a continuous function.

(d) If $X_n \xrightarrow{d} 0$, then we must have $X_n \xrightarrow{p} 0$ as well.

(e) If $\mathbb{E}\left((X_n - Y)^2\right) \to 0$, where $\mathbb{E}\left(Y^2\right) < \infty$ and $\mathbb{E}\left(X_n^2\right) < \infty$ for all n, then we must have $\mathbb{E}\left(X_n^2\right) \to \mathbb{E}\left(Y^2\right)$.

Solution.

思路：这五问是收敛模式的「会不会反推」标准考点（全在 4）。(a) 假——缺一致可积性的经典反例。(b) 真——Markov 不等式。(c) 真——CMT。(d) 真——退化极限下 $\xrightarrow{d}$ 等价 $\xrightarrow{p}$。(e) 真——L^2 范数连续（反三角不等式）。

(a) Disprove (FALSE). Let $U \sim \text{Unif}(0,1)$ and $X_n = n\,\mathbb{1}(U \le 1/n)$. Then $X_n(\omega) \to 0$ for every ω with $U(\omega) > 0$, i.e. $X_n \xrightarrow{a.s.} 0$; yet $\mathbb{E}(X_n) = n \cdot \frac{1}{n} = 1 \not\to 0$. Almost-sure convergence does not control the means without uniform integrability (mass escapes to $+\infty$). $\boxed{\text{False.}}$

(b) Prove (TRUE). By Markov's inequality, for any $M > 0$,

$$\mathrm{P}\left(|X_n| > M\,n^{1/3}\right) \le \frac{\mathbb{E}\left(|X_n|^3\right)}{M^3 n} = \frac{O(n)}{M^3 n} = \frac{O(1)}{M^3}.$$

Given $\eta > 0$, the $O(1)$ constant C satisfies $\sup_n \mathbb{E}\left(|X_n|^3\right)/n \le C$, so choosing $M = (C/\eta)^{1/3}$ makes $\sup_n \mathrm{P}\left(|X_n| > Mn^{1/3}\right) \le \eta$. Hence $\{X_n/n^{1/3}\}$ is bounded in probability, i.e. $X_n = O_p(n^{1/3})$. $\boxed{\text{True.}}$

(c) Prove (TRUE). This is the Continuous Mapping Theorem (4). If h is continuous (the discontinuity set has P_X-measure zero suffices), then $X_n \xrightarrow{d} X \Rightarrow h(X_n) \xrightarrow{d} h(X)$. *Sketch:* by Skorokhod's representation there exist $\tilde{X}_n \overset{d}{=} X_n$, $\tilde{X} \overset{d}{=}$

X with $\tilde{X}_n \to \tilde{X}$ a.s.; continuity gives $h(\tilde{X}_n) \to h(\tilde{X})$ a.s., hence in distribution, and distribution is preserved under $\stackrel{d}{=}$. $\boxed{\text{True.}}$

(d) Prove (TRUE). Convergence in distribution to the constant 0 implies convergence in probability to 0. The limiting CDF is $F(x) = \mathbb{1}(x \geq 0)$, continuous at every $x \neq 0$. For $\varepsilon > 0$,

$$\mathrm{P}(|X_n| > \varepsilon) \leq \mathrm{P}(X_n \leq -\varepsilon) + \mathrm{P}(X_n > \varepsilon) \leq F_n(-\varepsilon) + \big(1 - F_n(\varepsilon)\big) \to F(-\varepsilon) + (1 - F(\varepsilon)) = 0 + 0.$$

Hence $X_n \xrightarrow{p} 0$. $\boxed{\text{True.}}$ 这是「极限是常数时 $\xrightarrow{d} \Leftrightarrow \xrightarrow{p}$」的一半，资格考反复考。

(e) Prove (TRUE). $\mathbb{E}\big((X_n - Y)^2\big) \to 0$ means $\|X_n - Y\|_2 \to 0$. By the reverse triangle inequality in L^2,

$$\big|\, \|X_n\|_2 - \|Y\|_2 \,\big| \leq \|X_n - Y\|_2 \to 0,$$

so $\|X_n\|_2 \to \|Y\|_2$, i.e. $\mathbb{E}\big(X_n^2\big) = \|X_n\|_2^2 \to \|Y\|_2^2 = \mathbb{E}\big(Y^2\big)$. $\boxed{\text{True.}}$

4.7 Q7 —独立 / 条件均值独立 / 不相关：三个反例

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：独立、条件均值独立与不相关的反例
历史复现：2020、2021、2023、2024
掌握方式：记三层蕴含关系和一组可复用反例

原题 Q7 (verbatim)

Create a joint distribution of two random variables x, y with finite variances for each of the following three situations:

(a) x, y are independent;

(b) y is conditionally mean independent of x but x, y are *not* independent;

(c) y and x are uncorrelated but y is *not* conditionally mean independent of x.

You can choose whether x, y are discrete, continuous, or mixed and whether you provide distribution functions or probability mass/density functions.

Solution.

思路：这题在区分三层逐渐变弱的「无关」：独立 ⇒ 均值独立 ⇒ 不相关。要造的是**严格落在两层之间**的例子。(b) 要均值独立但不独立 ⇒ 让方差随 x 变（异方差）。(c) 要不相关但非均值独立 ⇒ 让 $\mathbb{E}(y\,|\,x)$ 是 x 的**偶函数**（如 x^2），从而与 x 的协方差为零。讲解见 3。

(a) Let $x, y \stackrel{i.i.d.}{\sim} N(0,1)$ (or i.i.d. Bernoulli(1/2)). Finite variances, and the joint density factors $f(x, y) = \phi(x)\phi(y)$, so $x \perp y$. ✓

(b) Mean-independent but not independent. Let $x \sim N(0,1)$ and $\eta \sim N(0,1)$ with $\eta \perp x$, and set $y = x\,\eta$. Then

$$\mathbb{E}(y \,|\, x) = x\,\mathbb{E}(\eta) = 0 \quad \text{(does not depend on } x\text{): mean-independent,}$$

but $\operatorname{Var}(y \,|\, x) = x^2 \operatorname{Var}(\eta) = x^2$ depends on x, so the conditional distribution varies with x and $x \not\perp y$. Variances are finite ($\operatorname{Var}(y) = \mathbb{E}\big(x^2\big)\,\mathbb{E}\big(\eta^2\big) = 1$). ✓

(c) Uncorrelated but not mean-independent. Let $x \sim N(0,1)$ and $y = x^2 - 1$. Then

$$\operatorname{Cov}(x, y) = \mathbb{E}\big(x(x^2-1)\big) = \mathbb{E}\big(x^3\big) - \mathbb{E}(x) = 0 \quad \text{(uncorrelated),}$$

because all odd moments of $N(0,1)$ vanish, yet

$$\mathbb{E}(y \,|\, x) = x^2 - 1 \quad \text{depends on } x \;\Rightarrow\; \text{not conditionally mean-independent.}$$

Finite variances ($\operatorname{Var}(y) = \operatorname{Var}\big(x^2\big) = 2$). ✓口诀：要「不相关但非均值独立」，让 $\mathbb{E}(y \,|\, x)$ 关于 x 对称（偶）；偶函数与 x 这种对称分布上的奇函数正交，协方差自然为 0。

4.8 Q8 —对称截断误差下的 OLS 一致性、渐近分布、ML 与效率

归属：ECON 510（按内容归类） **重要度**：中 **再考判断**：中
核心考点：截断正态样本中的 OLS、MLE 与效率
历史复现：2020；截断/选择模型在 2023–2025 以 Heckit 再现
掌握方式：先区分 truncation 与 selection；推导条件似然归一化项

原题 Q8 (verbatim)

Consider $y_i^* = x_i^\top \beta_0 + u_i$, where x_i, u_i are mutually independent, data are i.i.d., and $\mathbb{E}\big(x_i x_i^\top\big) > 0$. Further, $u_i \sim N(0,1)$. The variable y_i^* is unobserved. Instead, you observe $y_i = y_i^*$ if and only if $|u_i| \le 1$.

(a) Show that the OLS estimator $\hat\beta$ of β_0 is consistent.

(b) Figure out what the asymptotic distribution of $\hat\beta$ is in this specific case. (Only this specific case is needed.)

(c) Suppose one would instead use ML. Write down the loglikelihood function for this specific case (you only observe y_i when $|u_i| \le 1$).

(d) Would the MLE in this case be more or less efficient than absent the truncation? Explain.

Solution.
思路：选择规则是 $|u_i| \le 1$——一个**关于** u **对称、且与** x **独立**的集合。两点救命：(i) 对称 ⇒ 截断后 $\mathbb{E}(u \mid |u| \le 1, x) = 0$，矩条件 $\mathbb{E}(xu) = 0$ 在选择样本里仍成立 ⇒ OLS 一致；(ii) 选择概率 $\mathrm{P}(|u| \le 1)$ 不依赖 x（因 $u \perp x$）⇒ 选择样本里 $\mathbb{E}(xx^\top)$ 不变。(b) 同方差三明治退化成 $\mathrm{Var}(u \mid |u| \le 1) \cdot (\mathbb{E}(xx^\top))^{-1}$，截断方差要算。(c) 截断正态似然。(d) 信息比较是题眼，老师明说没那么直观。

(a) Consistency. Conditioning is on the event $S_i := \{|u_i| \le 1\}$. Since $u_i \perp x_i$ and the truncated $N(0,1)$ on $[-1,1]$ is symmetric about 0,

$$\mathbb{E}(u_i \mid S_i, x_i) = \mathbb{E}(u_i \mid |u_i| \le 1) = 0,$$

so in the selected population $\mathbb{E}(x_i u_i \mid S_i) = \mathbb{E}(x_i\,\mathbb{E}(u_i \mid S_i, x_i)) = 0$. OLS on the observed sample solves $\frac{1}{n}\sum_{i:S_i} x_i(y_i - x_i^\top \hat\beta) = 0$, giving

$$\hat\beta - \beta_0 = \Big(\frac{1}{n}\sum_{i:S_i} x_i x_i^\top\Big)^{-1} \frac{1}{n}\sum_{i:S_i} x_i u_i \xrightarrow{p} \Big(\mathbb{E}\big(x_i x_i^\top \mid S_i\big)\Big)^{-1}\mathbb{E}(x_i u_i \mid S_i) = 0.$$

(The selected design $\mathbb{E}\big(x_i x_i^\top \mid S_i\big) = \mathbb{E}\big(x_i x_i^\top\big) > 0$ because $S_i \perp x_i$.) Hence $\hat\beta \xrightarrow{p} \beta_0$. $\boxed{\text{Consistent.}}$

(b) Asymptotic distribution. Put $S_i := \mathbb{1}(|u_i| \le 1)$, $p_S := \mathrm{P}(S_i = 1) = 2\Phi(1) - 1$, and $Q := \mathbb{E}\big(x_i x_i^\top\big)$. On the selected sample the errors are homoskedastic with variance $\sigma_T^2 := \mathrm{Var}(u_i \mid S_i = 1)$. Conditional on the number $n_S = \sum_i S_i$ of observed units, the usual OLS CLT gives

$$\sqrt{n_S}\,(\hat\beta - \beta_0) \xrightarrow{d} N\Big(0,\ \sigma_T^2\,\Big(\mathbb{E}\big(x_i x_i^\top\big)\Big)^{-1}\Big).$$

Compute the truncated variance for $u \sim N(0,1)$ restricted to $[-1,1]$ (mean 0 by symmetry):

$$\sigma_T^2 = 1 - \frac{2\phi(1)}{2\Phi(1) - 1} \approx 1 - \frac{2(0.24197)}{0.68269} \approx 0.2911.$$

$$\boxed{\sqrt{n_S}\,(\hat\beta - \beta_0) \xrightarrow{d} N\Big(0,\ \Big(1 - \tfrac{2\phi(1)}{2\Phi(1)-1}\Big)\Big(\mathbb{E}\big(x_i x_i^\top\big)\Big)^{-1}\Big).}$$

If n denotes the number of latent draws before selection, then $n_S/n \xrightarrow{p} p_S$ and Slutsky equivalently gives

$$\boxed{\sqrt{n}\,(\hat\beta - \beta_0) \xrightarrow{d} N\left(0, \frac{\sigma_T^2}{p_S} Q^{-1}\right).}$$

$\sigma_T^2 \approx 0.2911$；以保留下来的样本量 n_S 归一化时用 $\sigma_T^2 Q^{-1}$，以最初抽取数 n 归一化时还要除以保留率 $p_S \approx 0.6827$。

(c) Log-likelihood (truncated normal). Conditional on being observed, y_i has density equal to the $N(x_i^\top\beta, 1)$ density renormalized over the region $\left|y_i - x_i^\top\beta\right| \le 1$:

$$f(y_i \mid x_i, S_i = 1; \beta) = \frac{\phi(y_i - x_i^\top\beta)}{p_S}\mathbb{1}\left\{\left|y_i - x_i^\top\beta\right| \le 1\right\}.$$

Hence, over the *fixed set of observed units*, with $\log 0 = -\infty$,

$$\boxed{\ell_n(\beta) = \sum_{i:S_i=1}\left[\log\phi(y_i - x_i^\top\beta) - \log p_S + \log\mathbb{1}\left\{\left|y_i - x_i^\top\beta\right| \le 1\right\}\right].}$$

If the missingness indicators for all latent draws are recorded, their contribution is $\sum_{i:S_i=0}\log(1-p_S)$, also constant in β. Thus the MLE minimizes the sum of squared residuals *subject to all support restrictions simultaneously*:

$$\boxed{\hat\beta_{\mathrm{ML}} = \arg\min_{\beta\in\mathcal{B}_n}\sum_{i:S_i=1}(y_i - x_i^\top\beta)^2, \qquad \mathcal{B}_n := \bigcap_{i:S_i=1}\{\beta : |y_i - x_i^\top\beta| \le 1\}.}$$

The set of summation indices must not depend on the candidate β: a candidate that violates even one observed unit's support has likelihood zero. Therefore the MLE is *not* generally the unconstrained OLS estimator.

(d) Efficiency vs. no truncation. 题眼不是把截断方差塞进普通 Fisher-information 公式，而是看见支撑依赖参数。这违反正则 MLE 理论，边界极值会带来比 $\sqrt{n}$ 更快的信息。

The support $[x_i^\top\beta - 1, x_i^\top\beta + 1]$ moves with β. Consequently, the score does not satisfy the usual regularity identity obtained by differentiating under a fixed support, and a conventional Cramér–Rao/Fisher-information comparison is invalid. The support restrictions themselves reveal β_0 very sharply.

For example, in the intercept-only case $x_i = 1$, every observed draw satisfies $y_i = \beta_0 + u_i$ with $u_i \in [-1, 1]$. The likelihood is positive only when

$$\max_{i:S_i=1}(y_i - 1) \le \beta \le \min_{i:S_i=1}(y_i + 1).$$

Equivalently, this random feasible interval has endpoints

$$\beta_0 + u_{(n_S)} - 1 \quad\text{and}\quad \beta_0 + u_{(1)} + 1.$$

The truncated-normal density is positive at both endpoints -1 and 1, so standard extreme-value calculations give

$$1 - u_{(n_S)} = O_p(n_S^{-1}), \qquad u_{(1)} + 1 = O_p(n_S^{-1}).$$

Thus the entire feasible interval, and hence the MLE inside it, is within $O_p(n^{-1})$

of β_0. With sufficiently rich regressors, analogous support inequalities pin down every direction of β_0 at the same nonregular rate. By contrast, without truncation the Gaussian-regression MLE has only the usual $\sqrt{n}$ rate.

> The truncated MLE is asymptotically more informative (typically n-consistent), but this is a nonregular support-boundary result, not a smaller regular Fisher variance.

Its normalized limit is generally extreme-value/non-normal and depends on the design. OLS from part (b) ignores this support information and remains only $\sqrt{n}$-consistent.

4.9 Q9 —固定阈值双删失有序 probit：识别与 MLE 极限分布

归属：ECON 510（按内容归类） **重要度**：中 **再考判断**：中
核心考点：双删失 ordered probit 的识别与 MLE
历史复现：2020；离散选择识别见 2021/2022，Heckit 见 2024/2025
掌握方式：写三格概率、score 与信息矩阵；不背本题长式

原题 Q9 (verbatim)

Assume i.i.d. data and consider $y_i^* = x_i^\top \beta_0 + u_i$, where $u_i \,|\, x_i = x \sim N(0, \sigma_0^2)$ for some $0 < \sigma_0^2 < \infty$ and $0 < \mathbb{E}\left(x_i x_i^\top\right) < \infty$ where the support of x_i includes zero. The variable y_i^* is latent. Instead, you observe

$$y_i = \begin{cases} -1, & y_i^* \leq -1, \\ 0, & |y_i^*| \leq 1, \\ 1, & y_i^* \geq 1. \end{cases}$$

(a) Argue whether or not β_0 is identified. (If β_0 is not identified then assume $\sigma_0 = 1$ below.)

(b) What is the limit distribution of the MLE of β_0? Show the steps.

Solution.

思路：这是**阈值已知固定**（± 1）的三类有序 probit。和普通有序 probit（阈值自由 $\Rightarrow$ 只识别 β/σ）不同：阈值钉死在 ± 1 提供了**尺度锚**，所以 β_0 与 σ_0 都识别（前提 x 有变化、且支撑含 0 给出截距/水平的锚）。(b) 把似然写成三格多项 logit-型，套极大似然渐近 $\sqrt{n}(\hat\beta - \beta_0) \xrightarrow{d} N(0, I^{-1})$，见 5。

(a) Identification. The three cell probabilities given x are

$$\mathrm{P}(y=-1\,|\,x)=\Phi\Big(\tfrac{-1-x^\top\beta_0}{\sigma_0}\Big),\quad \mathrm{P}(y=1\,|\,x)=1-\Phi\Big(\tfrac{1-x^\top\beta_0}{\sigma_0}\Big),$$

$$\mathrm{P}(y=0\,|\,x)=\Phi\Big(\tfrac{1-x^\top\beta_0}{\sigma_0}\Big)-\Phi\Big(\tfrac{-1-x^\top\beta_0}{\sigma_0}\Big).$$

The thresholds are fixed and known (± 1), not free parameters. They anchor the scale: the two known cutpoints ± 1 are observed in the same units as y^*, so σ_0 is separately identified (unlike free-cutpoint ordered probit, where only β/σ is identified). Concretely, Φ^{-1} of the cell probabilities gives two equations

$$\frac{-1-x^\top\beta_0}{\sigma_0},\qquad \frac{1-x^\top\beta_0}{\sigma_0}$$

as identified functions of x; their *difference* is $2/\sigma_0$ (free of x and of β_0), which identifies σ_0; then $x^\top\beta_0$ is identified, and since the support of x includes 0 and $\mathbb{E}\big(xx^\top\big)>0$ (full rank, with an intercept-like variation from the support including zero), β_0 is identified.

$$\boxed{\beta_0 \text{ (and } \sigma_0\text{) are identified —the fixed thresholds } \pm 1 \text{ fix the scale.}}$$

(b) Limit distribution of the MLE. The model is a three-cell multinomial. Let $p_k(x;\theta)$, $k\in\{-1,0,1\}$, be the cell probabilities above with $\theta=(\beta',\sigma)'$. The log-likelihood is

$$\ell(\theta)=\sum_{i=1}^{n}\sum_{k\in\{-1,0,1\}}\mathbb{1}(y_i=k)\,\log p_k(x_i;\theta).$$

This is a smooth, correctly specified MLE with fixed support $\{-1,0,1\}$. Under the usual interiority, domination, and nonsingular-information conditions,

$$\boxed{\sqrt{n}\,(\hat\theta-\theta_0)\xrightarrow{d}N\big(0,\mathcal{I}(\theta_0)^{-1}\big)\,,\qquad \mathcal{I}(\theta_0)=\mathbb{E}\left(\sum_{k=-1}^{1}\frac{\dot p_k(x;\theta_0)\dot p_k(x;\theta_0)'}{p_k(x;\theta_0)}\right).}$$

Here $\dot p_k=\partial p_k/\partial\theta$. To make every derivative explicit, put

$$a=\frac{-1-x'\beta}{\sigma},\qquad b=\frac{1-x'\beta}{\sigma}.$$

Then

$$\partial_\beta p_{-1}=-\phi(a)x/\sigma,\quad \partial_\beta p_0=\{\phi(a)-\phi(b)\}x/\sigma,\quad \partial_\beta p_1=\phi(b)x/\sigma,$$
$$\partial_\sigma p_{-1}=-a\phi(a)/\sigma,\quad \partial_\sigma p_0=\{a\phi(a)-b\phi(b)\}/\sigma,\quad \partial_\sigma p_1=b\phi(b)/\sigma.$$

Because σ_0 is estimated jointly, the requested marginal law for $\hat{\beta}$ is

$$\boxed{\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{d} N\big(0,\, [\mathcal{I}(\theta_0)^{-1}]_{\beta\beta}\big)\,, \qquad [\mathcal{I}^{-1}]_{\beta\beta} = \big(\mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta\sigma}\mathcal{I}_{\sigma\sigma}^{-1}\mathcal{I}_{\sigma\beta}\big)^{-1}.}$$

If σ_0 were known instead, the covariance would reduce to $\mathcal{I}_{\beta\beta}^{-1}$. The nuisance-adjusted Schur complement is essential when σ_0 is unknown.

4.10 Q10 —GMM 识别计数、过度识别的好处、最优工具类比

归属：ECON 510（按内容归类） **重要度**：中高 **再考判断**：中高
核心考点：GMM 识别计数、过度识别与 J 检验
历史复现：2020、2021、2022；GMM 为长期地基
掌握方式：背参数数目/矩数目逻辑和 J 检验自由度

原题 Q10 (verbatim)

Assume i.i.d. data and consider the model $y_i^* = x_i^\top\beta_0 + q_i\gamma_0 + u_i$, $q_i = w_i^\top\delta_0 + v_i$, where y_i^* is latent, u_i, v_i are mean-zero jointly normal and independent of $z_i = [x_i^\top, w_i^\top]^\top$, u_i has variance one. w_i, x_i have multiple elements with none in common. You observe $y_i = \mathbb{1}(y_i^* \geq 0)$. It can be shown that

$$\mathbb{E}(y_i \,|\, z_i = z) = \Phi\Big(\tfrac{x^\top\beta_0 + w^\top\delta_0\gamma_0}{\sqrt{1+\gamma_0^2\sigma_0^2+2\gamma_0\sigma_0\rho_0}}\Big),$$

$\sigma_0^2 = \mathrm{Var}(v_i)$, $\rho_0 = \mathrm{Corr}(u_i, v_i)$. You form the moment conditions

$$\mathbb{E}\Big(z_i\Big(y_i - \Phi(\cdots)\Big)\Big) = 0, \quad \mathbb{E}\Big(z_i(q_i - w_i^\top\delta_0)\Big) = 0, \quad \mathbb{E}\Big((q_i - w_i^\top\delta_0)^2 - \sigma_0^2\Big) = 0.$$

(a) Do the moment conditions provide under-, exact, or over-identification? Explain briefly.

(b) Name one advantage of using GMM with an overidentified system compared to using an asymptotically equally efficient GMM estimator based on an exactly identified system.

(c) Finish the sentence: "Exactly identified GMM using a given fixed vector of instruments z_i is to OLS what GMM with optimal instruments is to ..."

Solution.

思路：(a) 这题故意让「矩数多于参数数」，但 order count 不是 rank condition。先把后两组矩能识别的 δ_0, σ_0 固定，再检查第一组 probit index 到底能不能把 $\beta_0, \gamma_0, \rho_0$ 分开。答案是不能：它只识别两个缩放组合，留下一个连续等价方向。

(a) The system is underidentified. Let

$$s_0 := \sqrt{1+\gamma_0^2\sigma_0^2+2\gamma_0\sigma_0\rho_0} > 0.$$

Under the stated full-rank/variation conditions, the second block identifies δ_0 and the third identifies σ_0^2. The first block can at most identify the coefficients in the conditional probit index,

$$b_x := \frac{\beta_0}{s_0}, \qquad b_w := \frac{\delta_0\gamma_0}{s_0}.$$

If at least one component of δ_0 is nonzero, b_w and the known δ_0 identify only the scalar

$$a := \frac{\gamma_0}{s_0},$$

not γ_0 and s_0 separately. Thus the observable moments determine $(b_x, a, \delta_0, \sigma_0^2)$, but not the primitive triple $(\beta_0, \gamma_0, \rho_0)$.

To display the observational equivalence explicitly, suppose $a \neq 0$. For any candidate γ in a sufficiently small neighborhood of γ_0 with the same sign, set

$$s(\gamma) := \frac{\gamma}{a}, \qquad \beta(\gamma) := s(\gamma)b_x, \qquad \rho(\gamma) := \frac{s(\gamma)^2 - 1 - \gamma^2\sigma_0^2}{2\gamma\sigma_0}.$$

Then

$$s(\gamma)^2 = 1+\gamma^2\sigma_0^2+2\gamma\sigma_0\rho(\gamma), \qquad \frac{\beta(\gamma)}{s(\gamma)} = b_x, \qquad \frac{\delta_0\gamma}{s(\gamma)} = \delta_0 a.$$

Hence every such triple generates exactly the same probit index and leaves the other two moment blocks unchanged. If $|\rho_0| < 1$, continuity makes $|\rho(\gamma)| < 1$ for all nearby γ, so this is a genuine local continuum of admissible parameters. If $a = 0$ (in particular $\gamma_0 = 0$), ρ_0 drops out altogether, so identification still fails.

Naive counting gives $2(k_x+k_w)+1$ moments versus k_x+k_w+3 parameters, but the Jacobian of the population moments is rank-deficient by at least one. Therefore

$$\boxed{\text{the stated moment system is underidentified; excess moment count does not imply overidentification.}}$$

(b) Advantage of over-identification. A genuinely identified system with $q > p$ admits a test of the overidentifying restrictions (Hansen's J / Sargan test): under the usual full-rank conditions and correct specification, $n\,\bar g_n^\top \hat W \bar g_n \xrightarrow{d} \chi^2_{q-p}$ (10). This is a *specification test* of model validity that an exactly identified system cannot provide—even when the point estimator is asymptotically equally efficient, exact identification leaves no residual moments to test.

$$\boxed{\text{Over-identification yields a specification (Hansen } J\text{) test of the model.}}$$

这里 (b) 是一般性的假设比较。对 (a) 中这个欠识别系统，不能直接把「矩数减参

数数」当作 J 检验自由度；必须先补充能消除 γ–ρ 等价方向的识别限制。

(c) Analogy.

$$\boxed{\text{... GLS (generalized least squares).}}$$

直觉：固定工具的恰好识别 GMM ≡ OLS（不管条件方差结构）；最优工具用条件方差 $\Omega(z)$ 加权，正像 GLS 用方差结构改进 OLS。下一题 Q11(b) 就是把这个最优工具具体推出来。

4.11 Q11 —乘性异方差 IV：识别矩条件与最优工具

归属：ECON 510（按内容归类） **重要度**：中高 **再考判断**：中
核心考点：乘性异方差 IV 的识别矩与最优工具
历史复现：2020、2021；最优工具在 2017–2022 高频
掌握方式：从独立性生成矩，再由 D/V 求工具

原题 Q11 (verbatim)

Suppose that $y_i = x_i^\top\beta_0 + u_i$, where $x_i = \Pi_0^\top z_i + v_i$ and $u_i = \exp(z_i^\top\gamma_0)e_i$, with e_i mean-zero unit-variance, e_i and v_i independent of z_i, and all variables have finite fourth moments.

(a) Write down some moment conditions that provide identification under reasonable additional conditions, and state what those conditions are.

(b) Derive the optimal instruments.

Solution.

思路：(a) 因 $\mathbb{E}(u\,|\,z) = \exp(z^\top\gamma_0)\mathbb{E}(e) = 0$，$z$（及其函数）是合法工具：$\mathbb{E}\big(z(y - x^\top\beta_0)\big) = 0$。识别要 relevance（$\Pi_0$ 满秩）+ 阶条件 $\dim z \geq \dim x$ + $\mathbb{E}\big(zz^\top\big)$ 非奇异。γ_0 由平方残差的矩 $\mathbb{E}\big([\exp(-2z^\top\gamma_0)(y - x^\top\beta)^2 - 1]\,m(z)\big) = 0$ 识别。(b) 条件矩模型的最优工具 $= D(z)\Omega(z)^{-1}$，其中 $D(z) = \mathbb{E}(\partial g/\partial\beta\,|\,z) = -\mathbb{E}(x\,|\,z) = -\Pi_0^\top z$，$\Omega(z) = \mathrm{Var}(u\,|\,z) = \exp(2z^\top\gamma_0)$。见 14/最优工具与 11 条件矩效率界。

(a) Identifying moment conditions. Since $e_i \perp z_i$ with $\mathbb{E}(e_i) = 0$,

$$\mathbb{E}(u_i\,|\,z_i) = \exp(z_i^\top\gamma_0)\,\mathbb{E}(e_i\,|\,z_i) = 0,$$

so any function of z_i is a valid instrument. The basic conditions are

$$\boxed{\mathbb{E}\Big(z_i\big(y_i - x_i^\top\beta_0\big)\Big) = 0}$$

(and more generally $\mathbb{E}\big(A(z_i)(y_i - x_i^\top\beta_0)\big) = 0$ for any matrix function $A(\cdot)$). For γ_0, use the unit-variance normalization $\mathbb{E}\big(e_i^2\big) = 1$, i.e. $\mathbb{E}\big(u_i^2\,|\,z_i\big) = \exp(2z_i^\top\gamma_0)$, giving

$$\boxed{\mathbb{E}\Big(\Big(\exp(-2z_i^\top\gamma_0)\,(y_i - x_i^\top\beta_0)^2 - 1\Big)\,m(z_i)\Big) = 0}$$

for instruments $m(z_i)$ (e.g. $m(z_i) = (1, z_i^\top)^\top$).

For feasibility one may also identify the reduced-form coefficient Π_0 from

$$\boxed{\mathbb{E}\Big(z_i\big(x_i - \Pi_0^\top z_i\big)^\top\Big) = 0,}$$

using the additional normalization $\mathbb{E}(v_i) = 0$ (together with $v_i \perp z_i$).

Additional conditions for identification:

- **Reduced form and relevance:** $\mathbb{E}(v_i) = 0$, $\mathbb{E}\big(z_i z_i^\top\big)$ is nonsingular, and Π_0 has full column rank (equivalently $\mathbb{E}\big(z_i x_i^\top\big) = \mathbb{E}\big(z_i z_i^\top\big)\,\Pi_0$ has rank $\dim x$).
- **Order:** $\dim z_i \geq \dim x_i$, and $\mathbb{E}\big(z_i z_i^\top\big)$ nonsingular.
- **Variance-index rank:** the chosen $m(z_i)$ supplies at least $\dim\gamma_0$ linearly independent equations; locally, the Jacobian $-2\mathbb{E}\big(m(z_i)z_i^\top\big)$ (at the normalized squared-residual moment) has full column rank.
- **Exogeneity:** $e_i, v_i \perp z_i$ (given), ensuring $\mathbb{E}(u_i \,|\, z_i) = 0$ and validity of the squared-residual moment.

Under these, β_0 is identified from the first set and γ_0 from the second.

(b) Optimal instruments. For the conditional-moment model $\mathbb{E}(g_i(\beta_0) \,|\, z_i) = 0$ with $g_i(\beta) = y_i - x_i^\top\beta$, the efficient (optimal) instrument is (14)

$$A^*(z_i) = D(z_i)\,\Omega(z_i)^{-1},$$

where

$$D(z_i) = \mathbb{E}\Big[\frac{\partial g_i(\beta_0)}{\partial\beta^\top}\,\Big|\,z_i\Big] = -\,\mathbb{E}\Big(x_i^\top \,|\, z_i\Big) = -\,(\Pi_0^\top z_i)^\top = -\,z_i^\top\Pi_0,$$

(using $\mathbb{E}(v_i \,|\, z_i) = 0$), and

$$\Omega(z_i) = \mathrm{Var}(u_i \,|\, z_i) = \mathbb{E}\big(u_i^2 \,|\, z_i\big) = \exp(2z_i^\top\gamma_0)\,\mathbb{E}\big(e_i^2\big) = \exp(2z_i^\top\gamma_0).$$

Hence the optimal instrument (as a $\dim x$-vector multiplying the scalar residual) is

$$\boxed{A^*(z_i) \;\propto\; \exp(-2z_i^\top\gamma_0)\;\Pi_0^\top z_i\;,}$$

and the efficient estimator solves $\sum_{i=1}^n A^*(z_i)\,(y_i - x_i^\top\hat\beta) = 0$, i.e.

$$\hat\beta_{\mathrm{opt}} = \Big(\sum_{i=1}^n e^{-2z_i^\top\gamma_0}\Pi_0^\top z_i x_i^\top\Big)^{-1}\sum_{i=1}^n e^{-2z_i^\top\gamma_0}\Pi_0^\top z_i\,y_i.$$

The feasible version plugs in consistent $\hat{\Pi}$ (from regressing x on z) and $\hat{\gamma}$ (from the squared-residual moments). It attains the semiparametric efficiency bound

$$V^* = \Big(\mathbb{E}\Big(D(z_i)^\top \Omega(z_i)^{-1} D(z_i)\Big)\Big)^{-1} = \Big(\mathbb{E}\Big(e^{-2z_i^\top \gamma_0}\, \Pi_0^\top z_i z_i^\top \Pi_0\Big)\Big)^{-1}.$$

这正是 Q10(c) 类比的兑现：最优工具用条件方差 $\Omega(z) = e^{2z^\top \gamma_0}$ 的倒数加权，把异方差「除掉」——就是 IV 版的 GLS。relevance 处 $D(z) = -\Pi_0^\top z$ 用到了第一阶段约简形式。

4.12　Q12 —玩笑题

归属：玩笑题（无课程知识归属）　**重要度**：极低　**再考判断**：极低
核心考点：识别题目是否为可判定数学命题
历史复现：2020、2021 均有类似玩笑题
掌握方式：不投入复习时间

原题 Q12 (verbatim)

True or false? Please explain. (a) This exam is hard.

Solution.

这是卷尾的玩笑题，无数学内容。 **False (conditionally).** Conditional on having internalized the methods compiled in this guide —the L^2-projection characterization of conditional expectation (3), the delta method and modes of convergence (4), MLE/GMM asymptotics (5, 8), identification counting (7), specification testing (10), and optimal instruments (14) —every problem above reduces to “name the tool, then turn the crank.” Unconditionally the statement is a matter of taste, hence not a well-posed proposition. $\boxed{\text{Answer: it depends on the conditioning information.}}$

Chapter 5　2021 资格考精解（Worked Solutions）

本章逐题给出 2021 春资格考的完整解答。中文负责思路与直觉，英文负责形式推导与最终答案（装入方框）。每题先复述原题，再点出对应章节的方法，最后独立重推。2021 与 2020 高度同构：阶统计量、MGF 识别正态、L^2 投影、变量变换、混合识别、样本方差极限、O_p 阶；510 部分考 GMM 识别计数与乘性异方差下的最优工具。

Remark (扫描缺页说明 · Missing pages).

> 原始扫描只含印刷页 1 与 3，缺失印刷页 2（题 7–9）。故本章覆盖 Q1–Q6 与 Q10–Q12；Q7–Q9 无题面，留空待补。

5.1　Q1 · 指数分布的危险率刻画与阶统计量 / Hazard, Exponential, Order Statistics

归属：ECON 501（按内容归类）　**重要度**：中高　**再考判断**：中高
核心考点：危险率刻画指数分布与阶统计量
历史复现：2020、2021、2025 的指数分布题同族
掌握方式：hazard 到生存函数的微分方程要会推；阶统计量记模板

原题 Q1

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables. Suppose that $\mathrm{P}(X_1 \le 0) = 0$ and $\lim_{h \downarrow 0} \mathrm{P}(X_1 \le x + h \mid X_1 > x)/h = 1$ for all $x > 0$. Let $X_{(j)}$ be the j^{th} smallest variable among $X_1, \ldots, X_n$.
(a) Compute $\mathbb{E}(X_1)$ and $\mathbb{E}\big(X_{(1)}\big)$.
(b) Let F be the CDF of X_1. Derive the CDF and PDF of $Y_{(j)} := F(X_{(j)})$ for $j = 1, 2, \ldots, n$.

Solution.

(a) 中文思路. 题给的极限正是**危险率**（hazard rate）$\lambda(x) = f(x)/[1 - F(x)]$。因

为

$$\lim_{h\downarrow 0}\frac{\mathrm{P}(X_1\le x+h\mid X_1>x)}{h}=\lim_{h\downarrow 0}\frac{F(x+h)-F(x)}{h\,[1-F(x)]}=\frac{f(x)}{1-F(x)}=\lambda(x).$$

条件说 $\lambda(x)\equiv 1$（$x>0$），且 $\mathrm{P}(X_1\le 0)=0$ 把分布钉在 $(0,\infty)$ 上。常数危险率 $\equiv 1$ 唯一对应标准指数分布。见 2 的危险率与 2 阶统计量。

Formal. A constant hazard $\lambda(x)=1$ integrates to the survival function

$$S(x)=1-F(x)=\exp\Big(-\int_0^x\lambda(t)\,\mathrm{d}t\Big)=e^{-x},\qquad x>0,$$

so $X_1\sim\mathrm{Exp}(1)$ with density $f(x)=e^{-x}\,\mathbb{1}(x>0)$ and $\mathbb{E}(X_1)=1$. The minimum of n i.i.d. Exp(1) has survival $\mathrm{P}\big(X_{(1)}>x\big)=\prod_{i=1}^n\mathrm{P}(X_i>x)=e^{-nx}$, i.e. $X_{(1)}\sim\mathrm{Exp}(n)$. Hence

$$\boxed{\mathbb{E}(X_1)=1,\qquad \mathbb{E}\big(X_{(1)}\big)=\tfrac{1}{n}.}$$

(b) 中文思路．概率积分变换：$U_i:=F(X_i)\sim\mathrm{Unif}(0,1)$ i.i.d.（F 连续严增）。F 单调不减保序，故 $F(X_{(j)})=U_{(j)}$ 是 n 个 i.i.d. 均匀变量的第 j 个阶统计量，服从 $\mathrm{Beta}(j,n-j+1)$。见 2。

Formal. For $y\in[0,1]$, $Y_{(j)}=U_{(j)}\le y$ iff at least j of the n i.i.d. Uniforms fall below y, so the CDF is

$$F_{Y_{(j)}}(y)=\sum_{k=j}^{n}\binom{n}{k}y^k(1-y)^{n-k},\qquad 0\le y\le 1,$$

(with 0 below and 1 above). Differentiating (the telescoping binomial sum) gives the Beta density

$$\boxed{f_{Y_{(j)}}(y)=\frac{n!}{(j-1)!\,(n-j)!}\,y^{j-1}(1-y)^{n-j}\,\mathbb{1}(0\le y\le 1),\quad Y_{(j)}\sim\mathrm{Beta}(j,\,n-j+1).}$$

5.2 Q2 · MGF 识别二元正态 + 截断条件期望 / MGF ⇒ Normal, Truncated Mean

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：MGF 识别二元正态与截断均值
历史复现：2020、2021、2024、2025
掌握方式：联合正态线性变换和逆 Mills 公式必须熟练

原题 Q2

Suppose that X and Y are random variables with $\mathbb{E}(\exp(tX+sY))=\exp\{(t^2+s^2)/2\}$ for all $t,s\in\mathbb{R}$. Compute $\mathbb{E}(2X+Y\mid X+2Y>0)$.

Solution.

中文思路. 联合 MGF $= e^{(t^2+s^2)/2}$ 正是 $N(0, I_2)$ 的 MGF，由 MGF 唯一性（2）得 $X, Y \overset{i.i.d.}{\sim} N(0,1)$。记 $A = 2X + Y$, $B = X + 2Y$，二者联合正态、均值零。关键两步：(i) 对联合正态用 L^2 投影 $\mathbb{E}(A \mid B) = \frac{\mathrm{Cov}(A,B)}{\mathrm{Var}(B)} B$（3）；(ii) $\mathbb{E}(B \mid B > 0)$ 用半正态均值公式。

Formal. Since the MGF factors as $e^{t^2/2} e^{s^2/2}$, $X \perp Y$ and each is $N(0,1)$. Then

$$\mathrm{Var}(A) = 2^2{+}1 = 5, \quad \mathrm{Var}(B) = 1{+}2^2 = 5, \quad \mathrm{Cov}(A,B) = \mathbb{E}((2X+Y)(X+2Y)) = 2{+}0{+}0{+}2 = 4.$$

For jointly normal mean-zero (A,B), $A = \frac{\mathrm{Cov}(A,B)}{\mathrm{Var}(B)} B + \varepsilon = \frac{4}{5} B + \varepsilon$ with $\varepsilon \perp B$ and $\mathbb{E}(\varepsilon) = 0$, so

$$\mathbb{E}(A \mid B > 0) = \tfrac{4}{5}\, \mathbb{E}(B \mid B > 0)\,.$$

With $B \sim N(0,5)$, the half-normal mean is $\mathbb{E}(B \mid B > 0) = \sqrt{\mathrm{Var}(B)}\,\sqrt{2/\pi} = \sqrt{5}\sqrt{2/\pi} = \sqrt{10/\pi}$. Therefore

$$\boxed{\mathbb{E}(2X + Y \mid X + 2Y > 0) = \frac{4}{5}\sqrt{\frac{10}{\pi}} = \frac{4\sqrt{10}}{5\sqrt{\pi}} \approx 1.427.}$$

半正态均值：$Z \sim N(0, \tau^2) \Rightarrow \mathbb{E}(Z \mid Z > 0) = \tau\sqrt{2/\pi}$。

推导：$\mathbb{E}(Z \mid Z > 0) = \dfrac{\int_0^\infty z\, \tau^{-1} \phi(z/\tau)\, \mathrm{d}z}{1/2} = \tau\sqrt{2/\pi}$。

5.3 Q3 · L^2 投影、变量变换与残差正交 / Best Predictor, Transformation, Orthogonality

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：L^2 投影、变换密度与正交性
历史复现：2020、2021、2024
掌握方式：理解投影定理并能直接验证条件期望

原题 Q3

Suppose (X, Y) has joint pdf $f_{XY}(x, y) = 8xy\,\mathbb{1}(0 \le x \le y \le 1)$.

(a) Obtain $g : (0,1) \to \mathbb{R}$ such that for any measurable $h : (0,1) \to \mathbb{R}$, $\mathbb{E}\big(\{Y - g(X)\}^2\big) \le \mathbb{E}\big(\{Y - h(X)\}^2\big)$.

(b) Let $Z = X^2$. Obtain the pdf of Z.

(c) Compute the correlation coefficient between $Y - g(X)$ and Z.

Solution.

(a) 中文思路. 最小化 MSE 的预测子就是条件期望 $g(x) = \mathbb{E}(Y \mid X = x)$（$L^2$ 投影，3）。注意支撑是三角形 $0 \le x \le y \le 1$，给定 $X = x$ 时 Y 在 $[x, 1]$ 上。

Formal. Marginal $f_X(x) = \int_x^1 8xy\,\mathrm{d}y = 4x(1-x^2)$ on $(0,1)$. Then

$$\boxed{g(x) = \mathbb{E}(Y \mid X = x) = \frac{\int_x^1 y \cdot 8xy\,\mathrm{d}y}{4x(1-x^2)} = \frac{\frac{8x}{3}(1-x^3)}{4x(1-x^2)} = \frac{2}{3} \cdot \frac{1-x^3}{1-x^2}.}$$

化简：$\dfrac{1-x^3}{1-x^2} = \dfrac{(1-x)(1+x+x^2)}{(1-x)(1+x)} = \dfrac{1+x+x^2}{1+x}$，故 $g(x) = \dfrac{2(1+x+x^2)}{3(1+x)}$。

(b) 中文思路. 单调变换 $Z = X^2$，$x = \sqrt{z} \in (0,1)$，用变量变换公式（2）。

Formal. With $x = \sqrt{z}$, $|\mathrm{d}x/\mathrm{d}z| = \frac{1}{2\sqrt{z}}$,

$$f_Z(z) = f_X(\sqrt{z}) \cdot \frac{1}{2\sqrt{z}} = 4\sqrt{z}\,(1-z) \cdot \frac{1}{2\sqrt{z}} = \boxed{2(1-z)\,\mathbb{1}(0<z<1).}$$

(c) 中文思路. 这是“条件均值残差与任何 X 的函数正交”的标准事实（3）。$Z = X^2$ 是 X 的函数，故协方差为 0，相关系数为 0，无需算方差。

Formal. The residual $e := Y - g(X)$ satisfies $\mathbb{E}(e \mid X) = 0$. For any measurable ψ,

$$\mathrm{Cov}(e, \psi(X)) = \mathbb{E}(e\,\psi(X)) - \underbrace{\mathbb{E}(e)}_{=0}\mathbb{E}(\psi(X)) = \mathbb{E}(\mathbb{E}(e \mid X)\,\psi(X)) = 0.$$

Taking $\psi(X) = Z = X^2$ gives $\mathrm{Cov}(e, Z) = 0$. Both $\mathrm{Var}(e) > 0$ (the residual is non-degenerate since Y is not a.s. a function of X on this triangular support) and $\mathrm{Var}(Z) > 0$ ($Z = X^2$ is non-degenerate), so the correlation is well-defined and equals

$$\boxed{\mathrm{Corr}\big(Y - g(X),\, Z\big) = 0.}$$

5.4 Q4 · 潜在结果的边际识别与联合不可识别 / Identification under Bernoulli Selection

归属：ECON 501（按内容归类） **重要度**：中 **再考判断**：中
核心考点：潜在结果的边际识别与联合不可识别
历史复现：2020、2021；识别逻辑贯穿当前 510
掌握方式：从观测分布写可识别对象；不给联合分布强加假设

原题 Q4

Suppose (Y_1, Y_0) and T are independent, where T is Bernoulli. Define $Y := Y_1 T + Y_0(1-T)$.

(a) Can we identify the marginal distribution functions of Y_0 and Y_1 from the joint distribution of (Y, T)? Prove or disprove.

(b) Suppose Y_1, Y_0 are also Bernoulli and dependent with $Y_1 \geq Y_0$ w.p. 1. Can we identify $\mathrm{P}(Y_1 = 1 \mid Y_0 = 0)$ from the joint distribution of (Y, T)? Prove or disprove.

Solution.

(a) 答：可以识别（前提 $0 < \mathrm{P}(T = 1) < 1$**）**。中文思路：$T$ 独立于 (Y_1, Y_0)（随机分配），故在 $\{T = 1\}$ 上观测到的 Y 就是 Y_1，且其条件分布等于 Y_1 的**无条件**分布。见 7 的潜在结果识别。

Formal. On $\{T = 1\}$, $Y = Y_1$; by independence $T \perp (Y_1, Y_0)$,

$$\mathrm{P}(Y \le y \mid T = 1) = \mathrm{P}(Y_1 \le y \mid T = 1) = \mathrm{P}(Y_1 \le y) = F_{Y_1}(y),$$

and symmetrically $\mathrm{P}(Y \le y \mid T = 0) = F_{Y_0}(y)$. Both conditioning events have positive probability, so the left-hand sides are functions of the observed law of (Y, T):

$$\boxed{F_{Y_1}(y) = \mathrm{P}(Y \le y \mid T = 1)\,, \qquad F_{Y_0}(y) = \mathrm{P}(Y \le y \mid T = 0)\,.}$$

(b) 答：可以识别（前提 $p_0 < 1$**）**。中文思路：结果是二值的，加上单调性 $Y_1 \ge Y_0$ a.s. 恰好等价于 $\mathrm{P}(Y_0 = 1, Y_1 = 0) = 0$，这把 $(Y_0, Y_1) \in \{0, 1\}^2$ 的 2×2 联合表的四个格子只用两条边际 p_0, p_1 就唯一钉死（Fréchet 上界 / 同单调耦合）。于是 $\mathrm{P}(Y_1 = 1 \mid Y_0 = 0)$ 是边际的函数，而边际已由 (a) 识别。

Formal. From (a) the data pin down the marginals $p_1 := \mathrm{P}(Y_1 = 1)$ and $p_0 := \mathrm{P}(Y_0 = 1)$, with $p_1 \ge p_0$ forced by $Y_1 \ge Y_0$. Monotonicity $Y_1 \ge Y_0$ a.s. is equivalent to $\mathrm{P}(Y_0 = 1, Y_1 = 0) = 0$, so the 2×2 joint table of (Y_0, Y_1) is uniquely determined by the margins:

$$\mathrm{P}(Y_0 = 1, Y_1 = 1) = p_0, \quad \mathrm{P}(Y_0 = 0, Y_1 = 1) = p_1 - p_0, \quad \mathrm{P}(Y_0 = 0, Y_1 = 0) = 1 - p_1, \quad \mathrm{P}(Y_0 = 1, Y_1 = 0) =$$

(This is the Fréchet upper-bound / comonotone coupling: with $Y_1 \ge Y_0$, the event $\{Y_0 = 1\} \subseteq \{Y_1 = 1\}$.) Hence, for $p_0 < 1$,

$$\boxed{\mathrm{P}(Y_1 = 1 \mid Y_0 = 0) = \frac{\mathrm{P}(Y_0 = 0, Y_1 = 1)}{\mathrm{P}(Y_0 = 0)} = \frac{p_1 - p_0}{1 - p_0} \quad \text{is point-identified.}}$$

直观检验：取 $p_1 = \frac{3}{4}$，$p_0 = \frac{1}{4}$。四格为

$$\mathrm{P}(Y_0{=}1, Y_1{=}1) = \tfrac{1}{4}, \quad \mathrm{P}(Y_0{=}0, Y_1{=}1) = \tfrac{1}{2}, \quad \mathrm{P}(Y_0{=}0, Y_1{=}0) = \tfrac{1}{4}, \quad \mathrm{P}(Y_0{=}1, Y_1{=}0) = 0,$$

给出

$$\mathrm{P}(Y_1{=}1 \mid Y_0{=}0) = \tfrac{1/2}{3/4} = \tfrac{2}{3},$$

与公式 $\frac{p_1 - p_0}{1 - p_0} = \frac{1/2}{3/4} = \frac{2}{3}$ 一致。任何违反 $\mathrm{P}(Y_0{=}1, Y_1{=}0) = 0$ 的表（例如左下角非零）都违反单调性、不在可行集内，故无法用来证伪识别。

Remark (出题意图 · What Q4(b) tests).

> 核心在于识别的边界随结构假设而移动：随机分配单独只识别**边际**潜在结果分布（ATE 可识别），一般的**交叉**量（如 $\mathrm{P}(Y_1 = 1 \mid Y_0 = 0)$）依赖不可观测的联合 copula。但本题是二值结果**再加**单调性 $Y_1 \geq Y_0$——这把 copula 钉成 Fréchet 上界，联合表被边际唯一决定，故该条件概率反而**可**点识别。教训：别把"交叉量一般不可识别"当成口诀套用；在二值 + 单调性下要老老实实数自由度。

5.5 Q5 · 样本方差与 $\hat{\mu}^2$ 的极限分布 / Limit Laws via CLT + Delta Method

归属：ECON 501（按内容归类）　**重要度**：高（必会）　**再考判断**：高
核心考点：样本均值与样本方差联合极限
历史复现：2019、2020、2021、2025
掌握方式：二维 CLT、线性化及 delta 法要能逐步重现

原题 Q5

Suppose $X_1, \ldots, X_n$ are independent and identically distributed with $\mathbb{E}(X_1^4) < \infty$. Let $\hat{\mu} = n^{-1}\sum_{i=1}^n X_i$ and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n X_i^2 - \hat{\mu}^2$. Assume $\sigma^2 > 0$ and $\mathbb{E}((X_1-\mu)^4) > \sigma^4$, where $\mu = \mathbb{E}(X_1)$, $\sigma^2 = \mathrm{Var}(X_1)$.

(a) Derive the (non-degenerate) limit distribution of $n^r(\hat{\sigma}^2 - \sigma^2)$ for an appropriate $r > 0$.

(b) Derive the (non-degenerate) limit distribution of $n^r(\hat{\mu}^2 - \mu^2)$ for an appropriate $r > 0$.

Solution.

(a) 中文思路. $\hat{\sigma}^2$ 是中心二阶矩的样本版。写成对 $W_i := (X_i - \mu)^2$ 的样本均值再减去 $(\hat{\mu}-\mu)^2$ 的修正项；该修正项是 $O_p(1/n)$，被 $\sqrt{n}$ 一乘仍是 $o_p(1)$，故不影响极限。对 W_i 直接用 CLT（4）。$r = 1/2$。

Formal. Algebra gives $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2 - (\hat{\mu}-\mu)^2$. Since $(\hat{\mu}-\mu)^2 = O_p(n^{-1})$,

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n W_i - \sigma^2\Big) - \underbrace{\sqrt{n}(\hat{\mu}-\mu)^2}_{=O_p(n^{-1/2})=o_p(1)}.$$

By the CLT applied to $W_i = (X_i - \mu)^2$ (mean σ^2, finite variance because $\mathbb{E}(X_1^4) < \infty$),

$$\boxed{\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N\big(0,\ \mathrm{Var}\big((X_1-\mu)^2\big)\big) = N\big(0,\ \mathbb{E}\big((X_1-\mu)^4\big) - \sigma^4\big), \quad r = \tfrac{1}{2}.}$$

The assumption $\mathbb{E}\big((X_1-\mu)^4\big) > \sigma^4$ guarantees a strictly positive (non-degenerate) variance.

(b) 中文思路. 对 $\hat{\mu}$ 作 $x \mapsto x^2$ 的 delta 法（4）。导数 2μ 在 $\mu = 0$ 处消失，须分情形——这正是出题陷阱（与 2020 同构）。

Formal. CLT: $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Case $\mu \neq 0$. First-order delta method with $\mathrm{d}(x^2)/\mathrm{d}x = 2\mu$:

$$\sqrt{n}(\hat{\mu}^2 - \mu^2) \xrightarrow{d} N\big(0,\ (2\mu)^2\sigma^2\big) = N\big(0,\ 4\mu^2\sigma^2\big), \qquad r = \tfrac{1}{2}.$$

Case $\mu = 0$. The first derivative vanishes, so the $\sqrt{n}$-limit is degenerate; rescale by n (second-order / quadratic):

$$n\,\hat{\mu}^2 = \big(\sqrt{n}\hat{\mu}\big)^2 \xrightarrow{d} \big(N(0,\sigma^2)\big)^2 = \sigma^2\,\chi^2_1, \qquad r = 1.$$

By the continuous mapping theorem $(x \mapsto x^2)$. Collecting:

$$\boxed{n^r(\hat{\mu}^2 - \mu^2) \xrightarrow{d} \begin{cases} N(0,\ 4\mu^2\sigma^2), & \mu \neq 0,\ r = \frac{1}{2}, \\ \sigma^2\,\chi^2_1, & \mu = 0,\ r = 1. \end{cases}}$$

$\mu = 0$ 时 $\hat{\mu}^2 - \mu^2 = \hat{\mu}^2 \geq 0$，极限非对称（卡方），不可能是正态——这是“delta 法一阶系数为零须升阶”的标志性例子。

5.6 Q6 · 阶的判断：O_p 与 a.s. 收敛不蕴含 L^1 / Markov $\Rightarrow O_p$; UI Gap

归属：ECON 501（按内容归类） **重要度**：高 **再考判断**：高
核心考点：O_p、a.s. 与 L^1 的蕴含/反例
历史复现：2020、2021、2022、2024、2025
掌握方式：背最短尖峰反例和 Markov 证明

原题 Q6

Prove or disprove (all r.v. on a common probability space):

(a) If $\mathbb{E}\big(|X_n|^3\big) = O(n)$, then $X_n = O_p(n^{1/3})$.

(b) If $X_n \xrightarrow{a.s.} 0$, then $\mathbb{E}(X_n) \to 0$.

Solution.

(a) 真（True）。中文思路：由 Markov/Chebyshev 不等式把矩界翻译成尾概率界，再凑成 O_p 的定义（4）。

Formal. By Markov's inequality, for any $M > 0$,

$$\mathrm{P}\Big(|X_n| > M\,n^{1/3}\Big) \leq \frac{\mathbb{E}\big(|X_n|^3\big)}{M^3\,n} = \frac{O(n)}{M^3 n} = \frac{O(1)}{M^3}.$$

Given $\varepsilon > 0$, the $O(1)$ bound (say $\le C$ for all n) lets us pick $M = (C/\varepsilon)^{1/3}$ so that $\sup_n \mathrm{P}\big(|X_n| > Mn^{1/3}\big) \le \varepsilon$. Hence $X_n/n^{1/3}$ is bounded in probability, i.e.

$$\boxed{X_n = O_p(n^{1/3}).}$$

(b) 假（False）——反例. 中文思路：a.s. 收敛 $\not\Rightarrow$ L^1 收敛，缺的是一致可积（UI）/ 控制条件（4）。经典“质量逃逸”反例。

Formal. Let $U \sim \mathrm{Unif}(0,1)$ and set $X_n = n \cdot \mathbb{1}(U \le 1/n)$. For a.e. fixed ω with $U(\omega) > 0$, eventually $1/n < U(\omega)$ so $X_n(\omega) = 0$; thus $X_n \overset{a.s.}{\to} 0$. But

$$\mathbb{E}(X_n) = n \cdot \mathrm{P}(U \le 1/n) = n \cdot \tfrac{1}{n} = 1 \quad \text{for all } n,$$

so $\mathbb{E}(X_n) \to 1 \ne 0$. Therefore the statement is $\boxed{\text{false}}$: a.s. convergence does not imply convergence of means without uniform integrability (or a dominating function, as DCT requires).

5.7 Q7–Q9 · 扫描缺页 / Missing from scan

原始扫描缺印刷页 2，题 7–9 无题面。若日后补到题面，按本章体例补解。

5.8 Q10 · Probit 样本选择模型的 GMM 识别计数 / Identification Counting in GMM

归属：ECON 510（按内容归类） **重要度**：中高 **再考判断**：中高
核心考点：Probit 选择模型的 GMM 识别计数
历史复现：2020、2021、2022；Heckit 见 2024/2025
掌握方式：先数参数与矩，再写 rank，而不是只比较维数

原题 Q10

Model: $y_i^* = x_i^\top \beta_0 + q_i\gamma_0 + u_i$, $q_i = w_i^\top \delta_0 + v_i$, with (u_i, v_i) mean-zero jointly normal, independent of $z_i = [x_i^\top, w_i^\top]^\top$, $\mathrm{Var}(u_i) = 1$; w_i, x_i each have multiple elements $(k, m \ge 2)$ with no elements in common. Observe $y_i = \mathbb{1}(y_i^* \ge 0)$, and

$$\mathbb{E}(y_i \mid z_i = z) = \Phi\Big(\tfrac{x_i^\top \beta_0 + w_i^\top \delta_0 \gamma_0}{\sqrt{1+\gamma_0^2\sigma_0^2 + 2\gamma_0\sigma_0\rho_0}}\Big),$$

$\sigma_0^2 = \mathrm{Var}(v_i)$, $\rho_0 = \mathrm{Corr}(u_i, v_i)$. Moment conditions:

$$\mathbb{E}(z_i\{y_i - \Phi(\cdot)\}) = 0, \quad \mathbb{E}\Big(z_i(q_i - w_i^\top \delta_0)\Big) = 0, \quad \mathbb{E}\Big((q_i - w_i^\top \delta_0)^2 - \sigma_0^2\Big) = 0.$$

(a) Underidentification, exact identification, or overidentification? Explain.
(b) Name one advantage of overidentified GMM vs an asymptotically equally

efficient exactly-identified GMM.
(c) Finish: "Exactly identified GMM using a given fixed z_i is to OLS what GMM with optimal instruments is to ________."

Solution.

(a) 答：欠识别。 不能只数方程。后两组矩能钉住 δ_0, σ_0^2，但第一组 probit 矩只识别缩放后的 index coefficients，$\beta_0, \gamma_0, \rho_0$ 之间还剩一个连续等价方向。

Write

$$s_0 := \sqrt{1 + \gamma_0^2\sigma_0^2 + 2\gamma_0\sigma_0\rho_0} > 0.$$

Under the stated rank conditions, Block 2 identifies δ_0 and Block 3 identifies σ_0^2. Block 1 can at most identify

$$b_x := \frac{\beta_0}{s_0}, \qquad b_w := \frac{\delta_0\gamma_0}{s_0}.$$

If some component of δ_0 is nonzero, the known δ_0 and b_w identify only $a := \gamma_0/s_0$. They do not separately identify γ_0 and s_0.

For a direct rank/observational-equivalence proof, suppose $a \neq 0$. For every nearby candidate γ of the same sign, define

$$s(\gamma) = \frac{\gamma}{a}, \qquad \beta(\gamma) = s(\gamma)b_x, \qquad \rho(\gamma) = \frac{s(\gamma)^2 - 1 - \gamma^2\sigma_0^2}{2\gamma\sigma_0}.$$

Then

$$s(\gamma)^2 = 1 + \gamma^2\sigma_0^2 + 2\gamma\sigma_0\rho(\gamma), \qquad \frac{\beta(\gamma)}{s(\gamma)} = b_x, \qquad \frac{\delta_0\gamma}{s(\gamma)} = \delta_0 a.$$

Thus all these primitive triples give the same probit index and the same three blocks of population moments. If $|\rho_0| < 1$, continuity keeps $|\rho(\gamma)| < 1$ for all sufficiently nearby γ, so the continuum is admissible. If $a = 0$ (notably $\gamma_0 = 0$), ρ_0 drops out and identification again fails.

Although the raw count is $2(k+m)+1$ moments versus $k+m+3$ parameters, the moment Jacobian is rank deficient by at least one. Hence

$$\boxed{\text{Underidentification. Excess moment count is only an order condition, not a rank condition.}}$$

(b) 答. For a genuinely identified system with $q > p$, overidentification yields testable restrictions: Hansen's J statistic has a χ^2_{q-p} null limit and can assess model/moment validity. An exactly identified system sets its sample moments to zero by construction and provides no such specification test, even at equal asymptotic efficiency. For the system in (a), however, one must first add a restriction that removes the γ–ρ equivalence; naive "moments minus parameters" is not a valid J-test degree of freedom under rank failure.

(c) 答.

$$\boxed{\text{...is to } \textbf{GLS (generalized least squares).}}$$

类比：固定工具的恰好识别 GMM ↔ OLS（不加权）；最优工具 GMM ↔ GLS（按条件方差加权达有效）。见 7 最优工具。

5.9 Q11 · 乘性异方差下的识别矩与最优工具 / Optimal Instruments under Multiplicative Heteroskedasticity

归属：ECON 510（按内容归类） **重要度**：中高 **再考判断**：中
核心考点：乘性异方差 IV 与最优工具
历史复现：2020、2021；2017–2022 最优工具主线
掌握方式：从矩限制和条件方差现场推 D/V

原题 Q11

$y_i = x_i^\top\beta_0 + u_i$ with $x_i = \Pi_0^\top z_i + v_i$ and $u_i = \exp(z_i^\top\gamma_0)\, e_i$, where e_i is mean-zero unit-variance, $(e_i, v_i) \perp z_i$, all finite fourth moments.

(a) Write moment conditions that identify β_0 (and γ_0) under reasonable conditions; state the conditions.

(b) Derive the optimal instruments.

Solution.

(a) 中文思路. 关键观察：$e_i \perp z_i$ 且 $\mathbb{E}(e_i) = 0$，故

$$\mathbb{E}(u_i \mid z_i) = \exp(z_i^\top\gamma_0)\,\mathbb{E}(e_i \mid z_i) = \exp(z_i^\top\gamma_0)\cdot 0 = 0.$$

即 z_i（及其任意函数）是 u_i 的合法工具——这是一个标准条件矩模型（7）。

Moment conditions for β_0. For any measurable f,

$$\boxed{\mathbb{E}\Big(f(z_i)\,(y_i - x_i^\top\beta_0)\Big) = 0, \quad \text{in particular } \mathbb{E}\Big(z_i(y_i - x_i^\top\beta_0)\Big) = 0.}$$

Conditions for identification of β_0*:* the relevance/rank condition $\operatorname{rk}\big(\mathbb{E}\big(z_i x_i^\top\big)\big) = \dim(\beta_0)$, equivalently Π_0 has full column rank and z_i is correlated with x_i (so $\mathbb{E}\big(z_i x_i^\top\big) = \mathbb{E}\big(z_i z_i^\top\big)\,\Pi_0$ has full rank). 这保证 β_0 由上式唯一解出，是 IV 的秩条件。

For γ_0. Since $\mathbb{E}\big(u_i^2 \mid z_i\big) = \exp(2z_i^\top\gamma_0)\,\mathbb{E}\big(e_i^2 \mid z_i\big) = \exp(2z_i^\top\gamma_0)$, the conditional-variance equation

$$\boxed{\mathbb{E}\Big(g(z_i)\big[(y_i - x_i^\top\beta_0)^2 - \exp(2z_i^\top\gamma_0)\big]\Big) = 0}$$

identifies γ_0 given β_0, under a rank condition on $\mathbb{E}\big(z_i z_i^\top\big)$ (regressors z_i not collinear, so γ_0 is pinned by $\log \mathbb{E}\big(u^2 \mid z\big) = 2z^\top \gamma_0$).

For the nuisance reduced form Π_0. Add the reasonable normalization $\mathbb{E}(v_i) = 0$. Since $v_i \perp z_i$,

$$\boxed{\mathbb{E}\Big(z_i\big(x_i - \Pi_0^\top z_i\big)^\top\Big) = 0.}$$

This identifies Π_0 when $\mathbb{E}\big(z_i z_i^\top\big)$ is nonsingular. For the complete system, also require Π_0 to have full column rank and the Jacobian of the chosen γ-moments, $-2\mathbb{E}\big(g(z_i) z_i^\top \exp(2z_i^\top \gamma_0)\big)$, to have full column rank.

(b) 中文思路：Chamberlain 最优工具（7）. 对条件矩 $\mathbb{E}(u_i \mid z_i) = 0$，最优工具为

$$A^*(z_i) = D(z_i)^\top \, \Omega(z_i)^{-1}, \quad D(z_i) = \mathbb{E}(\partial u_i / \partial \beta \mid z_i) = -\mathbb{E}(x_i \mid z_i)^\top, \quad \Omega(z_i) = \operatorname{Var}(u_i \mid z_i)\,.$$

Formal. Here

$$\mathbb{E}(x_i \mid z_i) = \Pi_0^\top z_i + \mathbb{E}(v_i \mid z_i) = \Pi_0^\top z_i \quad (\text{since } v_i \perp z_i,\ \mathbb{E}(v_i) = 0),$$

so $D(z_i) = -(\Pi_0^\top z_i)^\top = -z_i^\top \Pi_0$, and

$$\Omega(z_i) = \operatorname{Var}(u_i \mid z_i) = \exp(2z_i^\top \gamma_0)\, \operatorname{Var}(e_i \mid z_i) = \exp(2z_i^\top \gamma_0).$$

Therefore the optimal instrument (the value that makes feasible-efficient GMM attain the semiparametric bound) is

$$\boxed{A^*(z_i) = D(z_i)^\top \Omega(z_i)^{-1} = -\,\Pi_0^\top z_i \cdot \exp\big(-2z_i^\top \gamma_0\big),}$$

i.e. instrument β_0's moment by $\mathbb{E}(x_i \mid z_i) = \Pi_0^\top z_i$ down-weighted by the inverse conditional variance $\exp(-2z_i^\top \gamma_0)$. The efficient estimating equation is

$$\mathbb{E}\Big(\big(\Pi_0^\top z_i\big) \exp(-2z_i^\top \gamma_0)\,(y_i - x_i^\top \beta_0)\Big) = 0.$$

直觉：把工具 $\mathbb{E}(x \mid z)$ 按条件方差的倒数 $\exp(-2z^\top \gamma_0)$ 加权，正是 GLS 式“给噪声小的观测更大权重”，与 Q10(c) 的 GLS 类比一致。
可行版：用第一步 $\hat{\beta}, \hat{\gamma}$ 估出 $\widehat{\Pi}_0, \hat{\gamma}_0$ 代入。

5.10 Q12 · 玩笑题 / Joke Item

归属：玩笑题（无课程知识归属） **重要度**：极低 **再考判断**：极低
核心考点：非数学陈述的真伪
历史复现：2020、2021
掌握方式：不投入复习时间

原题 Q12

True or false? Please explain. (a) “This exam is hard.”

Solution.

这是出题人印在卷尾的玩笑题，无计量内容。

Answer. **True.** A brief defensible explanation: the exam requires (i) recognizing characterizations of distributions from indirect data (hazard rate $\Rightarrow$ exponential in Q1; joint MGF $\Rightarrow$ bivariate normal in Q2), (ii) L^2-projection and orthogonality arguments (Q3), (iii) the subtle identification boundary between marginals and joint laws of potential outcomes (Q4), (iv) delta-method breakdowns requiring second-order/χ^2 limits (Q5b), (v) order-in-probability and uniform-integrability gaps (Q6), and (vi) GMM identification counting plus Chamberlain optimal instruments under multiplicative heteroskedasticity (Q10–Q11). *Therefore: true.* 当然，备考充分则未必——这正是本讲义存在的意义。

Chapter 6　2022 资格考精解（Worked Solutions）

导读 · Roadmap（先读这里，不含公式）

本章把 2022 年 8 月 16 日资格考（Guggenberger & Pinkse，四小时八题）逐题做透。前四题是 501 基础（马尔可夫链首达概率与吸收时间、乘积 σ 域上的可测性、OLS 比值估计量的 plim 加多元 CLT 加 delta 法、一组真伪判断），后四题是 510 进阶（M 估计量的一致性与渐近正态、识别概念分类、零膨胀 Poisson 的 score 检验、多项 logit 的条件矩 GMM 与最优工具）。每题先给*原题*（逐字），再给*中文思路*与 *English formal derivation*；用到的定理一律指回前两部分对应章节，不重复造轮子。

6.1　Q1（10%）马尔可夫链：弹劾的首达概率与期望时长

归属：ECON 501（按内容归类）　**重要度**：中　**再考判断**：中低
核心考点：有限 Markov 链首达概率与期望时长
历史复现：2017 Q1 有同类递推；此后未直接复现
掌握方式：会列边界递推并解线性方程；不背数值

原题 Q1（10%）

The legislatures of Utopia are trying to determine whether to impeach their President. The Congress meets first and can decide: (a) to impeach, (b) to pass the buck to the Senate, or (c) to not impeach. If decision (b) is made, then in the next round the Senate gets to decide on the same three options [(b) in this case is “pass the buck back to the Congress”]. The probabilities of each decision in each round are $p_C, q_C, 1 - p_C - q_C$ for the Congress and $p_S, q_S, 1 - p_S - q_S$ for the Senate, respectively.

(i) What is the probability p_I that the President is impeached?

(ii) If each decision (i.e. a decision in Congress and also a decision in the Senate) takes 1 month and $q := q_C = q_S$ what is the expected length of time before the process is terminated?

Solution.

思路。这是一个吸收马尔可夫链：两个暂态 C（轮到 Congress）、S（轮到 Senate），两个吸收态 I（弹劾）、N（不弹劾）。从 C 出发：以概率 p_C 进 I、以 $1-p_C-q_C$ 进 N、以 q_C 转到 S；从 S 对称。首达概率用一步条件（first-step analysis）列两条线性方程即可，比直接求几何级数更稳。「吸收链首达 = 解线性方程组」这套技术是自足的，不依赖某一章；若要类比，几何分布的首步递归思路可参 2。

(i) Let h_C (h_S) be the probability the President is *ever* impeached starting from a Congress (Senate) move. First-step analysis:

$$h_C = p_C + q_C\, h_S, \qquad h_S = p_S + q_S\, h_C.$$

Substituting the second into the first,

$$h_C = p_C + q_C\,(p_S + q_S h_C) \implies h_C\,(1-q_C q_S) = p_C + q_C p_S.$$

Congress moves first, so $p_I = h_C$:

$$\boxed{p_I = \frac{p_C + q_C\, p_S}{1-q_C\, q_S}} \qquad (0 < q_C q_S < 1).$$

校验：若 $p_S = q_S = 0$（参议院只会不弹劾），则 $p_I = p_C$，对；若 $p_C = p_S = p$，$q_C = q_S = q$，则 $p_I = \frac{p(1+q)}{1-q^2} = \frac{p}{1-q}$，正是「每一步以 q 续命、以 p 进 I」的几何首达概率，对。

(ii) 思路。T = 终止前发生的决策次数（每次决策 1 个月）。设 $q := q_C = q_S$。从任一暂态出发都恰好先走一步（+1），以概率 q 转到另一暂态继续，以概率 $1-q$ 当场吸收。于是不论起点，剩余步数分布相同，可用单一递归。

Let m be the expected number of remaining decisions starting from *either* body's turn (by symmetry in q the continuation probability is the same from C and from S). Each visit costs one decision and continues with probability q:

$$m = 1 + q\,m \implies \boxed{\mathbb{E}(T) = m = \frac{1}{1-q}\ \text{months}} \qquad (0 \le q < 1).$$

这里 T 服从参数 $1-q$ 的几何分布（支撑 $\{1,2,\dots\}$），均值 $1/(1-q)$。注意题目把「Congress 决策 + Senate 决策」并称每次 1 个月，所以单位是「决策步」即月，无需把 C、S 当两步合并。终止概率每步恒为 $1-q$，与具体落到 I 还是 N 无关，故 p_C, p_S 在 (ii) 中不出现。

Remark (Q1 易错点).

(ii) 常见错误是把一个「Congress 月 + Senate 月」当成一个周期 ⇒ 答 $\frac{2}{1-q^2}$。但原题明确「a decision in Congress and *also* a decision in the Senate）takes 1 month」指每个决策算一个时间单位（per-decision），且续命只在 q 上，故按单步递归得 $1/(1-q)$。若出题人本意是「每对决策 1 个月」，答案才会翻倍——考场上写清楚你对 timing 的读法即可拿分。

6.2　Q2（14%）测度论：乘积 σ 域由矩形生成元生成

归属：ECON 501（按内容归类）　**重要度**：高　**再考判断**：中高
核心考点：乘积 σ 域的生成
历史复现：测度论在 2017、2018、2022、2023、2024、2025 高频
掌握方式：理解生成元与单调类/π–λ 证明骨架

原题 Q2（14%）

a) Let $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, be measurable spaces. Assume $\mathcal{E}_2 \subset \mathcal{F}_2$ is such that $\sigma(\mathcal{E}_2) = \mathcal{F}_2$. Show $T : \Omega_1 \to \Omega_2$ is measurable (with respect to the sigma algebras $\mathcal{F}_1$ and $\mathcal{F}_2$) if and only if $T^{-1}(E_2) \in \mathcal{F}_1$ for all $E_2 \in \mathcal{E}_2$.

Let $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, be measurable spaces again, $\Omega := \Omega_1 \times \Omega_2$, and let $\rho_i : \Omega \to \Omega_i$ be the projection mapping onto the ith component, i.e. $\rho_i((w_1, w_2)) = w_i$, for $i = 1, 2$. Assume $\mathcal{E}_i \subset \mathcal{F}_i$ is such that i) $\sigma(\mathcal{E}_i) = \mathcal{F}_i$ for $i = 1, 2$ and ii) for each $i = 1, 2$ there are non-decreasing sequences $E_{mi} \in \mathcal{E}_i$ ($m \in \mathbb{N}$) such that $E_{mi} \uparrow \Omega_i$.

b) Let $\mathcal{A}$ be an arbitrary sigma-algebra in Ω. Show that each ρ_i is measurable (with respect to $\mathcal{A}$ and $\mathcal{F}_i$) if and only if $E_1 \times E_2 \in \mathcal{A}$ for any $E_i \in \mathcal{E}_i$, $i = 1, 2$. [Hint: use part a).]

c) Denote by $\otimes_{i=1}^2 \mathcal{F}_i$ the smallest sigma algebra on Ω such that both projection mappings ρ_i are measurable (with respect to $\mathcal{F}_i$) for $i = 1, 2$. Show that $\otimes_{i=1}^2 \mathcal{F}_i = \sigma(E_1 \times E_2;\ E_i \in \mathcal{E}_i,\ i = 1, 2)$.

d) Give an example that shows that the statement in c) no longer holds true if condition ii) is not assumed.

Solution.

贯穿全题的核心引理是「好集法 / good-sets」：判断一个映射是否可测，只需在生成元上验证原像落在源 σ 域里。这正是 1 的标准武器。(a) 把它证出来，(b)(c) 反复套用，(d) 找反例说明条件 ii) 不可省。

(a) (⇒) If T is $\mathcal{F}_1/\mathcal{F}_2$-measurable then $T^{-1}(E_2) \in \mathcal{F}_1$ for every $E_2 \in \mathcal{F}_2 \supseteq \mathcal{E}_2$, in particular for $E_2 \in \mathcal{E}_2$.

($\Leftarrow$) Suppose $T^{-1}(E_2) \in \mathcal{F}_1$ for all $E_2 \in \mathcal{E}_2$. Define the *good sets*

$$\mathcal{G} := \left\{B \subseteq \Omega_2 : \; T^{-1}(B) \in \mathcal{F}_1\right\}.$$

Because preimage commutes with all set operations, $T^{-1}(B^c) = \left(T^{-1}(B)\right)^c$ and $T^{-1}(\bigcup_k B_k) = \bigcup_k T^{-1}(B_k)$, and $T^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$, the collection $\mathcal{G}$ is a σ-algebra on Ω_2. By hypothesis $\mathcal{E}_2 \subseteq \mathcal{G}$, hence $\mathcal{F}_2 = \sigma(\mathcal{E}_2) \subseteq \mathcal{G}$. Thus $T^{-1}(B) \in \mathcal{F}_1$ for all $B \in \mathcal{F}_2$, i.e. T is measurable. □

(b) 把 (a) 用在 $T = \rho_i$，源 σ 域取 $\mathcal{A}$、目标 $\mathcal{F}_i$ 的生成元取 $\mathcal{E}_i$。关键是 $\rho_i^{-1}(E_i)$ 长什么样、以及怎么用条件 ii) 把它写成 $\mathcal{A}$ 里的可数并。

By (a) (with source σ-algebra $\mathcal{A}$, target generated by $\mathcal{E}_i$), ρ_i is $\mathcal{A}/\mathcal{F}_i$-measurable iff $\rho_i^{-1}(E_i) \in \mathcal{A}$ for all $E_i \in \mathcal{E}_i$. The preimages are cylinders:

$$\rho_1^{-1}(E_1) = E_1 \times \Omega_2, \qquad \rho_2^{-1}(E_2) = \Omega_1 \times E_2.$$

($\Leftarrow$) Assume $E_1 \times E_2 \in \mathcal{A}$ for all $E_i \in \mathcal{E}_i$. Using condition ii), pick $E_{m2} \uparrow \Omega_2$ with $E_{m2} \in \mathcal{E}_2$. Then for any fixed $E_1 \in \mathcal{E}_1$,

$$E_1 \times \Omega_2 = \bigcup_{m \in \mathbb{N}} (E_1 \times E_{m2}) \in \mathcal{A}$$

since each $E_1 \times E_{m2} \in \mathcal{A}$ and $\mathcal{A}$ is closed under countable unions; hence $\rho_1^{-1}(E_1) \in \mathcal{A}$. Symmetrically (using $E_{m1} \uparrow \Omega_1$), $\rho_2^{-1}(E_2) = \Omega_1 \times E_2 \in \mathcal{A}$. By (a), both ρ_i are measurable.

($\Rightarrow$) Assume both ρ_i are $\mathcal{A}/\mathcal{F}_i$-measurable. For any $E_i \in \mathcal{E}_i \subseteq \mathcal{F}_i$,

$$E_1 \times E_2 = (E_1 \times \Omega_2) \cap (\Omega_1 \times E_2) = \rho_1^{-1}(E_1) \cap \rho_2^{-1}(E_2) \in \mathcal{A},$$

an intersection of two $\mathcal{A}$-sets. □

(c) 两个包含。用 (b) 把「ρ_i 都可测」翻译成「所有矩形 $E_1 \times E_2$ 都在里面」。

Write $\mathcal{R} := \sigma\left(E_1 \times E_2 : \; E_i \in \mathcal{E}_i\right)$ and $\mathcal{D} := \otimes_{i=1}^2 \mathcal{F}_i$ (the smallest σ-algebra making both ρ_i measurable).

$\mathcal{D} \subseteq \mathcal{R}$: Apply (b) with $\mathcal{A} = \mathcal{R}$. Every generator $E_1 \times E_2$ lies in $\mathcal{R}$ by construction, so by (b) both ρ_i are $\mathcal{R}/\mathcal{F}_i$-measurable. Since $\mathcal{D}$ is the *smallest* such σ-algebra, $\mathcal{D} \subseteq \mathcal{R}$.

$\mathcal{R} \subseteq \mathcal{D}$: Apply (b) with $\mathcal{A} = \mathcal{D}$. By definition of $\mathcal{D}$ both ρ_i are $\mathcal{D}$-measurable, so the ($\Rightarrow$) direction of (b) gives $E_1 \times E_2 \in \mathcal{D}$ for all $E_i \in \mathcal{E}_i$. Hence $\mathcal{D}$ contains all generators of $\mathcal{R}$, so $\mathcal{R} = \sigma(\cdots) \subseteq \mathcal{D}$.

Therefore $\otimes_{i=1}^2 \mathcal{F}_i = \sigma(E_1 \times E_2 : \; E_i \in \mathcal{E}_i)$. □

$$\boxed{\otimes_{i=1}^2 \mathcal{F}_i = \sigma\left(E_1 \times E_2 : \; E_i \in \mathcal{E}_i, \; i = 1, 2\right)}$$

(d) **反例思路**：不能只找一个"不满足 ii)"的生成类；还必须真的证明它生成的矩形 σ 域严格小于乘积 σ 域。一个干净做法是用不可数集合上的 countable–

cocountable σ 域。

Let $\Omega_1 = \Omega_2 = \mathbb{R}$ and, for $i = 1, 2$, let

$$\mathcal{F}_i = \mathcal{C}(\mathbb{R}) := \{A \subseteq \mathbb{R} : \ A \text{ is countable or } A^c \text{ is countable}\}, \qquad \mathcal{E}_i = \{\{x\} : x \in \mathbb{R}\}.$$

Every countable set is a countable union of singletons, and complements then give every co-countable set. Conversely, $\mathcal{C}(\mathbb{R})$ is a σ-algebra containing all singletons. Hence

$$\sigma(\mathcal{E}_i) = \mathcal{C}(\mathbb{R}) = \mathcal{F}_i,$$

so condition i) holds. Condition ii) fails: if a sequence of singletons is non-decreasing under inclusion, all of its terms must be the same singleton, and therefore its union cannot equal the uncountable set $\mathbb{R}$.

The rectangles generated by $\mathcal{E}_1$ and $\mathcal{E}_2$ are precisely the singletons of $\mathbb{R}^2$:

$$\{x\} \times \{y\} = \{(x, y)\}.$$

Therefore

$$\mathcal{R} := \sigma\{E_1 \times E_2 : E_i \in \mathcal{E}_i\} = \mathcal{C}(\mathbb{R}^2),$$

the countable–cocountable σ-algebra on $\mathbb{R}^2$. By contrast, the product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ contains every measurable rectangle. In particular, for any fixed $x_0 \in \mathbb{R}$,

$$V := \{x_0\} \times \mathbb{R} \in \mathcal{F}_1 \otimes \mathcal{F}_2.$$

But V is uncountable and $V^c = (\mathbb{R} \setminus \{x_0\}) \times \mathbb{R}$ is also uncountable, so $V \notin \mathcal{C}(\mathbb{R}^2) = \mathcal{R}$. Thus

$$\boxed{\mathcal{R} \subsetneq \mathcal{F}_1 \otimes \mathcal{F}_2},$$

which is the required counterexample. □

Remark (Q2 为什么 ii) 不可省).

条件 ii) 的作用就在 (b) 的 ($\Leftarrow$) 那一步：把柱集 $E_1 \times \Omega_2$ 写成可数个矩形 $E_1 \times E_{m2}$ 的递增并。上面的反例把障碍看得最清楚：单点矩形只能生成 $\mathbb{R}^2$ 上的 countable–cocountable σ 域，却生成不了不可数且余集也不可数的整条竖线 $\{x_0\} \times \mathbb{R}$。注意：只证明某个生成类不满足 ii) 还不够；必须像这里一样展示结论确实失败。

6.3 Q3（14%）OLS 比值估计量的 plim、多元 CLT 与 delta 法

归属：ECON 501（按内容归类） **重要度**：高（必会） **再考判断**：高
核心考点：比值估计量的 plim、多元 CLT 与 delta 法
历史复现：2018–2025 几乎每年有渐近变换题
掌握方式：先写基础样本矩向量，再 Jacobian；不能直接猜方差

原题 Q3（14%）

Suppose that $Y_i = \beta X_i + \varepsilon_i$, where (X_i, ε_i) are i.i.d., X_i and ε_i are independent, $EX_i = \mu \neq 0$, and $E\varepsilon_i = 0$. Define

$$\widehat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}, \qquad \bar{\beta} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}, \qquad T_n = \exp(\widehat{\beta} + \bar{\beta}).$$

The objective is to work out the asymptotic distribution of T_n as $n \to \infty$ (after proper centering and normalization) following the steps (a)–(d) below. State any additional technical assumption (such as finiteness of moments) you require, precisely define any notation you introduce, precisely state any result from the lecture you use, and show all your work.

(a) Derive the probability limits of $\widehat{\beta}$, $\bar{\beta}$, and T_n.

(b) Define the vector $M_n = \left(n^{-1}\sum_{i=1}^n X_i\varepsilon_i,\ \ n^{-1}\sum_{i=1}^n \varepsilon_i\right)'$. Derive the asymptotic distribution of M_n (after proper centering and normalization).

(c) Consider $\widetilde{T}_n = \exp(2\beta)\exp\left(\dfrac{n^{-1}\sum X_i\varepsilon_i}{EX_i^2} + \dfrac{n^{-1}\sum \varepsilon_i}{\mu}\right)$. Derive the asymptotic distribution of $\widetilde{T}_n$ (after proper centering and normalization).

(d) Show that $\sqrt{n}(\widetilde{T}_n - T_n) = o_p(1)$. What is the asymptotic distribution of T_n?

Solution.

思路。这是一道把「LLN+ 连续映射 → plim」「多元 CLT」「delta 法」「$o_p(1)$ 等价化」串成一条线的标准题，方法全在 4。**额外矩条件**：设 $EX_i^2 < \infty$、$\mathrm{Var}(\varepsilon_i) = \sigma^2 < \infty$（且 $EX_i^2 > 0$，由 $\mu \neq 0$ 或方差非零保证分母 plim 非零）。记 $\sigma^2 := \mathrm{Var}(\varepsilon_i)$，$q := EX_i^2 = \mu^2 + \mathrm{Var}(X_i)$。

(a) Substitute $Y_i = \beta X_i + \varepsilon_i$:

$$\widehat{\beta} = \beta + \frac{n^{-1}\sum X_i\varepsilon_i}{n^{-1}\sum X_i^2}, \qquad \bar{\beta} = \beta + \frac{n^{-1}\sum \varepsilon_i}{n^{-1}\sum X_i}.$$

By the WLLN, $n^{-1}\sum X_i\varepsilon_i \xrightarrow{p} E[X_i\varepsilon_i] = E[X_i]E[\varepsilon_i] = \mu \cdot 0 = 0$ (independence), $n^{-1}\sum X_i^2 \xrightarrow{p} q > 0$, $n^{-1}\sum \varepsilon_i \xrightarrow{p} 0$, $n^{-1}\sum X_i \xrightarrow{p} \mu \neq 0$. By Slutsky / continuous mapping,

$$\widehat{\beta} \xrightarrow{p} \beta, \qquad \bar{\beta} \xrightarrow{p} \beta, \qquad \boxed{\operatorname{plim}\widehat{\beta} = \operatorname{plim}\bar{\beta} = \beta, \quad \operatorname{plim} T_n = \exp(2\beta)}$$

since $T_n = \exp(\widehat{\beta} + \bar{\beta})$ and exp is continuous.

(b) The summands $W_i := (X_i\varepsilon_i,\ \varepsilon_i)'$ are i.i.d. with mean 0 (just shown) and finite second moments. By the multivariate CLT (4),

$$\sqrt{n}\, M_n = \sqrt{n}\,(M_n - 0) \xrightarrow{d} N(0, \Sigma), \qquad \Sigma = \mathbb{E}\big(W_i W_i'\big).$$

Compute the entries using independence of X and ε and $E\varepsilon = 0$:

$$\begin{aligned}
\Sigma_{11} &= \mathbb{E}\big(X_i^2\varepsilon_i^2\big) = \mathbb{E}\big(X_i^2\big)\,\mathbb{E}\big(\varepsilon_i^2\big) = q\,\sigma^2,\\
\Sigma_{22} &= \mathbb{E}\big(\varepsilon_i^2\big) = \sigma^2,\\
\Sigma_{12} = \Sigma_{21} &= \mathbb{E}\big(X_i\varepsilon_i^2\big) = \mathbb{E}(X_i)\,\mathbb{E}\big(\varepsilon_i^2\big) = \mu\,\sigma^2.
\end{aligned}$$

$$\boxed{\sqrt{n}\, M_n \xrightarrow{d} N(0,\ \Sigma)\,, \qquad \Sigma = \sigma^2 \begin{pmatrix} q & \mu \\ \mu & 1 \end{pmatrix}, \quad q = EX_i^2}$$

(c) $\widetilde{T}_n = \exp(2\beta)\, g(M_n)$ 其中 $g(u_1, u_2) = \exp(u_1/q + u_2/\mu)$，在 $M_n \xrightarrow{p} 0$ 附近对 M_n 做一阶 delta 法。

Write $\widetilde{T}_n = \exp(2\beta)\,\phi(M_n)$ with $\phi(u) = \exp(u_1/q + u_2/\mu)$, smooth, $\phi(0) = 1$, gradient

$$g := \nabla\phi(0) = \left(\tfrac{1}{q},\ \tfrac{1}{\mu}\right)'.$$

Since $\widetilde{T}_n - \exp(2\beta) = \exp(2\beta)\,(\phi(M_n) - \phi(0))$ and $\sqrt{n}M_n \xrightarrow{d} N(0, \Sigma)$, the delta method (4) gives

$$\sqrt{n}\left(\widetilde{T}_n - \exp(2\beta)\right) \xrightarrow{d} N\big(0,\ \exp(4\beta)\, g'\Sigma g\big)\,.$$

Evaluate the asymptotic variance:

$$g'\Sigma g = \sigma^2 \left(\tfrac{1}{q}, \tfrac{1}{\mu}\right) \begin{pmatrix} q & \mu \\ \mu & 1 \end{pmatrix} \begin{pmatrix} 1/q \\ 1/\mu \end{pmatrix} = \sigma^2\left(\tfrac{1}{q} + \tfrac{2}{q} + \tfrac{1}{\mu^2}\right) = \sigma^2\left(\tfrac{3}{q} + \tfrac{1}{\mu^2}\right).$$

$$\boxed{\sqrt{n}\left(\widetilde{T}_n - \exp(2\beta)\right) \xrightarrow{d} N\Big(0,\ \exp(4\beta)\,\sigma^2\left(\tfrac{3}{q} + \tfrac{1}{\mu^2}\right)\Big)}$$

展开核验：$g'\Sigma g = \sigma^2\big[\frac{1}{q^2}\cdot q + 2\cdot\frac{1}{q}\frac{1}{\mu}\cdot\mu + \frac{1}{\mu^2}\cdot 1\big] = \sigma^2\big[\frac{1}{q} + \frac{2}{q} + \frac{1}{\mu^2}\big] = \sigma^2\big[\frac{3}{q} + \frac{1}{\mu^2}\big]$。交叉项里 μ 约掉了：$\frac{1}{q}\cdot\frac{1}{\mu}\cdot\mu = \frac{1}{q}$，故是 $\frac{2}{q}$ 不是 $\frac{2}{\mu}$。

(d) 要证 $\widetilde{T}_n$ 与 T_n 渐近等价。核心：$T_n = \exp(2\beta)\exp\big(R_{1n} + R_{2n}\big)$，其中两个比值 $R_{1n} = \widehat{\beta} - \beta = \frac{n^{-1}\sum X_i\varepsilon_i}{n^{-1}\sum X_i^2}$、$R_{2n} = \bar{\beta} - \beta = \frac{n^{-1}\sum \varepsilon_i}{n^{-1}\sum X_i}$；而 $\widetilde{T}_n$ 里的指数是把分母换成了它们的 plim（q 与 μ）。差别只来自分母的随机性，是 $o_p(1/\sqrt{n})$ 量级。

Let $a_n := \dfrac{n^{-1}\sum X_i\varepsilon_i}{q} + \dfrac{n^{-1}\sum \varepsilon_i}{\mu}$ (the exponent of $\widetilde{T}_n$, so $\widetilde{T}_n = \exp(2\beta)e^{a_n}$) and $b_n := (\widehat{\beta} - \beta) + (\bar{\beta} - \beta)$ (so $T_n = \exp(2\beta)e^{b_n}$). Both $a_n, b_n = O_p(n^{-1/2}) \xrightarrow{p} 0$. Their difference, writing $\hat{q} := n^{-1}\sum X_i^2$, $\hat{\mu} := n^{-1}\sum X_i$:

$$b_n - a_n = \underbrace{n^{-1}\sum X_i\varepsilon_i}_{=O_p(n^{-1/2})}\left(\tfrac{1}{\hat{q}} - \tfrac{1}{q}\right) + \underbrace{n^{-1}\sum \varepsilon_i}_{=O_p(n^{-1/2})}\left(\tfrac{1}{\hat{\mu}} - \tfrac{1}{\mu}\right).$$

Since $\hat{q} \xrightarrow{p} q$ and $\hat{\mu} \xrightarrow{p} \mu$, continuous mapping gives $\frac{1}{\hat{q}} - \frac{1}{q} = o_p(1)$ and $\frac{1}{\hat{\mu}} - \frac{1}{\mu} = o_p(1)$. Hence each product is $O_p(n^{-1/2})o_p(1) = o_p(n^{-1/2})$, so

$$\sqrt{n}\,(b_n - a_n) = o_p(1).$$

Now $\sqrt{n}(T_n - \widetilde{T}_n) = \exp(2\beta)\sqrt{n}\left(e^{b_n} - e^{a_n}\right)$. By the mean value theorem $e^{b_n} - e^{a_n} = e^{\xi_n}(b_n - a_n)$ for ξ_n between a_n, b_n; since $a_n, b_n \xrightarrow{p} 0$, $e^{\xi_n} \xrightarrow{p} 1 = O_p(1)$, so

$$\sqrt{n}(T_n - \widetilde{T}_n) = \exp(2\beta)\, e^{\xi_n}\sqrt{n}(b_n - a_n) = O_p(1)\cdot o_p(1) = o_p(1).$$

于是 $\sqrt{n}(T_n - \exp(2\beta)) = \sqrt{n}(\widetilde{T}_n - \exp(2\beta)) + o_p(1)$，由 Slutsky 两者同分布。

$$\boxed{\sqrt{n}\,(T_n - \exp(2\beta)) \xrightarrow{d} N\Big(0,\ \exp(4\beta)\,\sigma^2\left(\tfrac{1}{EX^2} + \tfrac{2}{EX^2} + \tfrac{1}{\mu^2}\right)\Big)}$$

Remark (Q3 方法地图).

(a) LLN+CMT 给 plim；(b) 多元 CLT，协方差靠 $X \perp \varepsilon$ 把交叉矩拆开；(c) 一阶 delta 法（梯度在 plim 处取）；(d) 把真估计量 T_n 的随机分母换成常数分母，差是 $o_p(1/\sqrt{n})$，于是真估计量与线性化 $\widetilde{T}_n$ 共享极限律。这四步是 4 的招牌流程，几乎每年都考其变体。

6.4 Q4（12%）真伪判断（六小题，重在解释）

归属：ECON 501（按内容归类） **重要度**：高（必会） **再考判断**：高
核心考点：收敛模式与期望/概率极限真伪
历史复现：2020、2021、2022、2024、2025
掌握方式：为每条命题准备定理或反例；条件写全

原题 Q4（12%）True or false? Explain! (The explanation is what counts!)

(a) Let $X_n \in \mathbb{R}$ be a sequence of integrable random variables. Then, if $X_n \to_p b \in \mathbb{R}$ then $EX_n \to b$.

(b) Assume we test $H_0 : \mu_1 = \mu_2 = 0$ vs H_1 : not H_0 based on a sample of iid bivariate random vectors $X_i \sim N(\mu, I_2)$, $i = 1, \ldots, n$, where $\mu = (\mu_1, \mu_2)'$. Assume for some $0 < \alpha < 1/2$ the null is rejected if $|t_{nj}| > z_{1-\alpha/2}$ for $j = 1$ or $j = 2$, where $t_{nj} = n^{1/2}(\overline{X}_{nj} - \mu_j)$, $\overline{X}_{nj} = n^{-1} \sum_{i=1}^{n} X_{ij}$ for $j = 1, 2$, and $X_i = (X_{i1}, X_{i2})'$. Then the asymptotic null rejection probability of the test is α.

(c) By the Hölder inequality, $E(1/X) \leq 1/E(X)$ for any positive random variable $X > 0$.

(d) A sufficient statistic is one whose marginal distribution does not depend on any unknown parameters.

(e) If $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$, and $m > n$, then $X - Y \sim \chi^2(m - n)$.

(f) Let X and Y be two mean zero random variables. Then, "$E(X|Y) = 0$ almost surely Y and $E(Y|X) = 0$ almost surely X" if and only if "X and Y are uncorrelated."

Solution.

(a) False. 依概率收敛不蕴含期望收敛，缺的是一致可积（UI）。Convergence in probability does not control the tails. Counterexample: $X_n = n$ with probability $1/n$ and $X_n = 0$ otherwise. Then $\mathrm{P}(|X_n| > \varepsilon) = 1/n \to 0$ so $X_n \xrightarrow{p} 0 = b$, yet $EX_n = n \cdot \frac{1}{n} = 1 \not\to 0$. The implication holds only under uniform integrability (or L^1-domination). See 4.

$$\boxed{\text{False: need uniform integrability for } EX_n \to b.}$$

(b) False. 这是「对两个坐标各做一次 level-α 检验、任一拒则拒」的并集检验，多重比较把 size 抬高。 Under H_0 (indeed under the stated centering, for any μ) the two statistics t_{n1}, t_{n2} are exactly $N(0, 1)$ and *independent* (since I_2 covariance), for every n. The test does *not* reject iff both $|t_{nj}| \leq z_{1-\alpha/2}$:

$$\mathrm{P}(\text{reject} \mid H_0) = 1 - (1 - \alpha)^2 = 2\alpha - \alpha^2 > \alpha.$$

The asymptotic (here exact) null rejection probability is $2\alpha - \alpha^2$, not α.

$$\boxed{\text{False: size } = 2\alpha - \alpha^2 > \alpha \text{ (union of two } \alpha\text{-tests).}}$$

(c) False. 方向反了，而且用的是 Jensen 不是 Hölder。 The map $x \mapsto 1/x$ is *convex* on $(0, \infty)$, so *Jensen's* inequality gives

$$E(1/X) \;\geq\; 1/E(X),$$

the opposite inequality, with equality iff X is degenerate. The cited tool (Hölder) and the direction are both wrong.

$$\boxed{\text{False: Jensen gives } E(1/X) \geq 1/E(X).}$$

(d) False. 这描述的是辅助统计量（ancillary），不是充分统计量。 A statistic whose *marginal* distribution is parameter-free is *ancillary*. A statistic T is *sufficient* if the *conditional* distribution of the data given T does not depend on the parameter (equivalently, by the factorization theorem, the likelihood factors as $g(T;\theta)h(x)$). See 5.

$$\boxed{\text{False: that is an \textit{ancillary} statistic; sufficiency is about the \textit{conditional} law.}}$$

(e) False. 一般不成立：$X-Y$ 甚至可能取负值，连支撑都不对。For arbitrary chi-squares the difference $X-Y$ can be negative (positive probability), so it cannot be χ^2. The statement holds *only* under a specific nested/independence structure, e.g. if $X = Y + Z$ with $Z \sim \chi^2(m-n)$ independent of Y —then $X - Y = Z \sim \chi^2(m-n)$. Mere marginal degrees of freedom $m > n$ with independent X, Y do *not* give it.

$$\boxed{\text{False in general (needs } X = Y + Z,\ Z \perp Y,\ Z \sim \chi^2(m-n)).}$$

(f) False. 一个方向成立，反方向不成立，所以「当且仅当」是假的。

($\Rightarrow$) If $E(X|Y) = 0$ a.s., then by the law of iterated expectations $E(XY) = E[Y\,E(X|Y)] = 0$; with $EX = EY = 0$ this gives $\mathrm{Cov}(X,Y) = 0$, so X, Y uncorrelated. (One conditional-mean condition already suffices for this direction.)

($\Leftarrow$) Uncorrelatedness does *not* imply conditional-mean-zero. Counterexample: $Y \sim N(0,1)$, $X = Y^2 - 1$. Then $EX = 0$, $EY = 0$, and $\mathrm{Cov}(X,Y) = E\left[(Y^2-1)Y\right] = E[Y^3] - E[Y] = 0$, so they are uncorrelated; but $E(X|Y) = Y^2 - 1 \not\equiv 0$. Hence the "iff" fails.

$$\boxed{\text{False: conditional-mean-zero } \Rightarrow \text{ uncorrelated, but not conversely.}}$$

6.5 Q5（15%）M 估计量：一致性与渐近正态

归属：ECON 510（按内容归类） **重要度**：高（必会） **再考判断**：高
核心考点：极值估计量的一致性与渐近正态
历史复现：2022、2023 Pinkse Q4、2024 510 Q1
掌握方式：背 argmin consistency 与 FOC 四步展开

原题 Q5（15%）

For the remaining questions, bold face means random variable. Consider

$$\theta_0 = \arg\min_{\theta\in\mathbb{R}} \mathbb{E}(3\log\{1+\exp(\theta-\mathbf{x}_1)\} - (\theta-\mathbf{x}_1)), \tag{1}$$

where $\mathbf{x}_1$ is integrable and drawn from a nondegenerate distribution. Suppose that you have an i.i.d. sample $\{\mathbf{x}_i\}$ at your disposal. Propose an estimator $\hat{\theta}$ of θ_0, prove its consistency, prove its asymptotic normality, and derive its asymptotic distribution. [15%]

Solution.

思路。这是教科书式的 M 估计量，照搬 8 的「极值估计量 → 一致性 → 渐近正态」三段模板即可。把准则函数记 $m(\theta,x) = 3\log(1+e^{\theta-x}) - (\theta-x)$，目标 $Q_0(\theta) = \mathbb{E}(m(\theta,\mathbf{x}_1))$。估计量取样本类比 $\hat{\theta} = \arg\min_\theta \widehat{Q}_n(\theta)$，$\widehat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^n m(\theta,\mathbf{x}_i)$。

Estimator. Let

$$\hat{\theta} = \arg\min_{\theta\in\mathbb{R}} \widehat{Q}_n(\theta), \qquad \widehat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \left[3\log\left(1+e^{\theta-\mathbf{x}_i}\right) - (\theta-\mathbf{x}_i)\right].$$

Score and Hessian (population). Let $\Lambda(z) = \dfrac{e^z}{1+e^z}$ be the logistic cdf, $\Lambda'(z) = \Lambda(z)(1-\Lambda(z))$. With $z = \theta - x$,

$$\frac{\partial}{\partial\theta} m(\theta,x) = 3\Lambda(\theta-x) - 1 =: s(\theta,x), \qquad \frac{\partial^2}{\partial\theta^2} m(\theta,x) = 3\Lambda'(\theta-x) > 0.$$

m 对 θ 严格凸（二阶导恒正），故 Q_0 严格凸，θ_0 是唯一最小元 ⇒ 识别成立。The first-order condition defining θ_0 is

$$Q_0'(\theta_0) = \mathbb{E}(3\Lambda(\theta_0-\mathbf{x}_1) - 1) = 0 \iff \mathbb{E}(\Lambda(\theta_0-\mathbf{x}_1)) = \tfrac{1}{3}.$$

(I) Consistency. 用 8 的标准条件（Newey–McFadden）：(i) 识别——θ_0 唯一最小元，上面凸性已给；(ii) 准则一致收敛 ULLN。 Conditions:

(i) *Identification.* Q_0 is strictly convex (since $\mathbb{E}(3\Lambda'(\theta-\mathbf{x}_1)) > 0$ for all θ, $\mathbf{x}_1$ nondegenerate) and finite (integrability of $\mathbf{x}_1$ controls the linear term; $0 \le \log(1+e^z) \le |z| + \log 2$), so θ_0 is the unique minimizer.

(ii) *Uniform convergence.* $m(\theta,x)$ is continuous in θ and dominated on compacts: for θ in a compact Θ, $|m(\theta,x)| \le 3(|\theta|+|x|+\log 2) + |\theta| + |x|$, integrable since $E|\mathbf{x}_1| < \infty$. By the uniform LLN, $\sup_{\theta\in\Theta} |\widehat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$.

By the consistency theorem for extremum estimators (8), $\hat{\theta} \xrightarrow{p} \theta_0$. (Strict convexity plus an initial compactification argument handles $\Theta = \mathbb{R}$; alternatively note $\widehat{Q}_n$ is convex, and convex-in-probability convergence at each point upgrades to uniform on compacts, giving $\hat\theta \xrightarrow{p} \theta_0$ by the convexity/argmax lemma.)

(II) Asymptotic normality. 标准 sandwich：把 FOC $\widehat{Q}_n'(\hat\theta) = 0$ 在 θ_0 处展开。 The FOC is $\widehat{Q}_n'(\hat\theta) = \frac{1}{n}\sum_{i=1}^n s(\hat\theta, \mathbf{x}_i) = 0$. Mean-value expansion about θ_0:

$$0 = \frac{1}{n}\sum_{i=1}^n s(\theta_0, \mathbf{x}_i) + \left[\frac{1}{n}\sum_{i=1}^n s'(\bar\theta, \mathbf{x}_i)\right](\hat\theta - \theta_0),$$

with $\bar\theta$ between $\hat\theta$ and θ_0. Define

$$H := Q_0''(\theta_0) = \mathbb{E}\big(3\Lambda'(\theta_0 - \mathbf{x}_1)\big) > 0, \qquad V := \mathrm{Var}(s(\theta_0, \mathbf{x}_1)) = \mathrm{Var}(3\Lambda(\theta_0 - \mathbf{x}_1) - 1) = 9\,\mathrm{Var}(\Lambda(\theta_0 - \mathbf{x}_1$$

By the CLT, $\sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^n s(\theta_0, \mathbf{x}_i) \xrightarrow{d} N(0, V)$ (mean zero by the FOC for θ_0); by the ULLN and $\bar\theta \xrightarrow{p} \theta_0$, $\frac{1}{n}\sum_{i=1}^n s'(\bar\theta, \mathbf{x}_i) \xrightarrow{p} H$. Solving,

$$\sqrt{n}\left(\hat\theta - \theta_0\right) = -\left[\frac{1}{n}\sum_{i=1}^n s'(\bar\theta, \mathbf{x}_i)\right]^{-1}\sqrt{n}\,\frac{1}{n}\sum_{i=1}^n s(\theta_0, \mathbf{x}_i) \xrightarrow{d} N\big(0,\ H^{-1}VH^{-1}\big).$$

$$\boxed{\begin{aligned}&\sqrt{n}\left(\hat\theta - \theta_0\right) \xrightarrow{d} N\left(0,\ \frac{V}{H^2}\right),\\ &V = 9\,\mathrm{Var}(\Lambda(\theta_0 - \mathbf{x}_1)),\quad H = 3\,\mathbb{E}(\Lambda(\theta_0 - \mathbf{x}_1)(1 - \Lambda(\theta_0 - \mathbf{x}_1)))\end{aligned}}$$

这是 M 估计量的通用 $H^{-1}VH^{-1}$ 夹心方差（misspecification-robust），不是 MLE 的 $-H^{-1}$，因为这个准则*不是*某个对数似然。一致估计 $\widehat{V}, \widehat{H}$ 用样本类比（在 $\hat\theta$ 处取）即可做推断。

Remark (Q5 这是不是某个似然?).

准则 $3\log(1 + e^{\theta - x}) - (\theta - x)$ 的 score 是 $3\Lambda(\theta - x) - 1$，要它是对数似然的 score 得能写成某分布的对数密度对参数的导数；这里 θ 进入的是「位置」而 $\mathbf{x}_1$ 是数据，结构上对应「分位/位置型 M 估计」而非该数据的 MLE。所以渐近方差走**夹心式**最稳妥，考场上别写成 H^{-1} 信息矩阵形式。

6.6 Q6（5%）识别概念之间的逻辑蕴含

归属：ECON 510（按内容归类） **重要度**：高 **再考判断**：高
核心考点：点、部分与弱识别的逻辑关系
历史复现：2022、2023、2025
掌握方式：定义必须逐字准确；不要写不存在的蕴含链

原题 Q6（5%）

For each pair (x, y) of the following concepts, indicate if (necessarily) $x \Rightarrow y$, $x \Rightarrow \neg y$, or $x \Leftrightarrow y$ (omit if there is no implication): (a) identified (b) weakly identified (c) exactly identified (d) overidentified (e) underidentified. No points without a correct explanation. *Correctly identified relationships earn points, incorrectly identified ones lose points.* [5%]

Solution.

思路。这题真正的陷阱不是画箭头，而是术语有两套常见约定。若不先声明定义，“exactly/overidentified” 有时表示已经满足秩条件的识别状态，有时只表示 moment counting；“underidentified” 也可能表示不点识别，或仅表示 $k < p$。两套约定给出的箭头不同。下面先给本讲义采用、也最适合 identification 题的实质性约定，再说明纯计数约定下应如何改。定义见 7 与 9.2。

Convention I（推荐作答：标签包含 rank/识别含义）. 令参数维数为 p、矩条件数为 k，人口矩 Jacobian 为 $G \in \mathbb{R}^{k\times p}$：

(a) *identified*: 观测分布（或人口矩系统）唯一钉住 θ_0；

(b) *weakly identified*: θ_0 仍点识别，但识别强度很小，或在三角阵列中趋于零；

(c) *exactly identified*: $k = p$ 且 G 满秩（局部恰好识别）；

(d) *overidentified*: $k > p$ 且 G 满列秩（局部过度识别）；

(e) *underidentified*: 参数不点识别（可能是 $k < p$，也可能是 $k \geq p$ 但 rank 失败）。

在这套定义下，所有必然的 pairwise 关系是

$$\boxed{\begin{gathered}\textbf{(e)} \Leftrightarrow \neg\textbf{(a)},\\ \textbf{(b)} \Rightarrow \textbf{(a)}, \qquad \textbf{(c)} \Rightarrow \textbf{(a)}, \qquad \textbf{(d)} \Rightarrow \textbf{(a)},\\ \textbf{(a)}, \textbf{(b)}, \textbf{(c)}, \textbf{(d)} \Rightarrow \neg\textbf{(e)}, \qquad \textbf{(e)} \Rightarrow \neg\textbf{(a)}, \neg\textbf{(b)}, \neg\textbf{(c)}, \neg\textbf{(d)},\\ \textbf{(c)} \Rightarrow \neg\textbf{(d)}, \qquad \textbf{(d)} \Rightarrow \neg\textbf{(c)}.\end{gathered}}$$

The repeated arrows in the third line are just contrapositives or consequences of the definitions; on an exam one need not list the same logical fact twice. There is no pairwise implication from (a) to either (c) or (d): an identified model may be exactly identified or overidentified. Nor is there a necessary implication between weakness (b) and the count labels (c)/(d): both an exactly and an overidentified model can be weak.

Convention II（若老师把 exact/over/under 只当作 order labels）. Define

$$\textbf{(c)}_o: \; k = p, \qquad \textbf{(d)}_o: \; k > p, \qquad \textbf{(e)}_o: \; k < p.$$

Then these three labels are mutually exclusive, so every one implies the negation of the other two. Under the regular smooth finite-dimensional moment setup, local identification requires $\mathrm{rk}(G) = p$, hence $k \geq p$; consequently

$$\textbf{(a)} \Rightarrow \neg\textbf{(e)}_o, \qquad \textbf{(b)} \Rightarrow \textbf{(a)} \Rightarrow \neg\textbf{(e)}_o, \qquad \textbf{(e)}_o \Rightarrow \neg\textbf{(a)}.$$

But $k = p$ or $k > p$ alone does *not* imply identification: for example, $G = 0$ has deficient rank regardless of how many rows it has. Thus under the order-only convention, omit $\textbf{(c)}_o \Rightarrow \textbf{(a)}$ and $\textbf{(d)}_o \Rightarrow \textbf{(a)}$.

两套约定不能混用。原答案一面把 (c)(d) 定义成纯计数，一面把 (e) 定义成「不识别」，于是会让一个 $k = p$ 但 G 降秩的模型同时被叫作 exact 与 underidentified；这并非逻辑矛盾，而是标签来自不同分类轴。修订后的答案把两个轴分开。若考场没有给定义，第一句应写："I use exact/overidentified to include the full-rank condition and underidentified to mean not point identified." 然后只列 Convention I 的箭头。另一个边界：若某文献把 weak-identification class 连同识别强度恰为零的点也一起称作 weakly identified，则 (b)⇒(a) 在该扩展用法下也应 omit；本题通常采用「weak but nonzero」的用法。

Remark (Q6 倒扣分策略).

先写定义，再写最少的铁律。推荐的最短安全答案是：weak/exact/over ⇒ identified；underidentified ⇔ not identified；exact 与 over 互斥；identified 本身不分别推出 exact 或 over，weakness 与 exact/over 没有必然关系。若老师明确说 exact/over 仅按 k 与 p 计数，则删除 exact/over ⇒ identified，并补上 order 三分类的互斥关系。

6.7 Q7（10%）零膨胀 Poisson 的 score 检验

归属：ECON 510（按内容归类） **重要度**：中 **再考判断**：中
核心考点：零膨胀 Poisson 的 score/LM 检验
历史复现：检验三件套见 2022、2024、2025
掌握方式：写受限 MLE、score、信息矩阵和 χ^2 自由度

原题 Q7（10%）

Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are iid draws from the distribution with probability mass function

$$p(x) = \gamma_0\, I(x = 0) + \lambda_0^x \exp(-\lambda_0)(1 - \gamma_0)/x!, \qquad x = 0, 1, \ldots,$$

where $-1/(e^{\lambda_0} - 1) \leq \gamma_0 \leq 1$ and $\lambda_0 > 0$. [10%]

(a) Derive the score test statistic for $\gamma_0 = 0$.

(b) Suppose you conducted both the score test and the likelihood ratio test for $\gamma_0 = 0$ and found that the likelihood ratio test produced a p-value equal to 0.03

and the score test produced a p-value equal to 0.45. How would you explain that contrast in this case?

Solution.

思路。score（LM）检验的好处是*只需在约束 MLE 处求值*——这里 $\gamma = 0$ 时模型退化成普通 Poisson，约束 MLE 就是 $\hat{\lambda} = \bar{x}$。方法模板见 6 的五检验。注意 γ_0 有下界 $-1/(e^{\lambda_0} - 1) < 0$，所以 $\gamma_0 = 0$ *不在边界上*（内点），常规 score 检验 χ_1^2 适用——这点在 (b) 的解释里很关键。

(a) Setup. Write the log-likelihood contribution for observation x, with parameters (γ, λ):

$$p(x; \gamma, \lambda) = \gamma\, I(x = 0) + (1 - \gamma)\, f(x; \lambda), \qquad f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

Score in γ: $\frac{\partial}{\partial \gamma} \log p = \dfrac{I(x = 0) - f(x; \lambda)}{p(x; \gamma, \lambda)}$. Evaluate *under* $H_0 : \gamma = 0$, where $p(x; 0, \lambda) = f(x; \lambda)$:

$$\frac{\partial}{\partial \gamma} \log p\bigg|_{\gamma=0} = \frac{I(x = 0) - f(x; \lambda)}{f(x; \lambda)} = \frac{I(x = 0)}{f(x; \lambda)} - 1 = \frac{I(x = 0)}{e^{-\lambda}} - 1 \quad (\text{since } f(0; \lambda) = e^{-\lambda}).$$

即 $u_\gamma(x; \lambda) = I(x = 0)e^{\lambda} - 1$。当 $x = 0$ 取 $e^{\lambda} - 1 > 0$，当 $x > 0$ 取 -1；它在度量「0 是否比 Poisson 预测的多」。

Restricted MLE. Under $\gamma = 0$ the model is plain Poisson, so $\hat{\lambda} = \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$.

Score statistic. Let the (efficient) score for γ at $(\gamma, \lambda) = (0, \hat{\lambda})$, summed over the sample, be

$$S_n = \sum_{i=1}^{n} \left(I(\mathbf{x}_i = 0)e^{\hat{\lambda}} - 1 \right) = e^{\hat{\lambda}}\, n_0 - n, \qquad n_0 := \#\{i : \mathbf{x}_i = 0\},$$

where the λ-MLE first-order condition $\sum_{i=1}^{n} (\mathbf{x}_i / \hat{\lambda} - 1) = 0$ at $\hat{\lambda} = \bar{x}$ leaves the numerator unchanged. The γ-score is *not* orthogonal to the λ-score, however, so the variance in the denominator must be the *efficient*-score variance (i.e. the λ-block partialled out). The score test statistic is

$$\mathrm{LM}_n = \frac{S_n^2}{n\, \widehat{\mathrm{Var}}_0 \left[u_\gamma(\mathbf{x}_1; \hat{\lambda}) \right]},$$

where, under the Poisson(λ), $u_\gamma(x; \lambda) = I(x = 0)e^{\lambda} - 1$ has mean 0 and raw variance

$$\mathrm{Var}(u_\gamma) = e^{2\lambda}\, \mathrm{Var}(I(x = 0)) = e^{2\lambda} \cdot e^{-\lambda}(1 - e^{-\lambda}) = e^{\lambda} - 1.$$

But u_γ is correlated with the λ-score $u_\lambda(x;\lambda) = x/\lambda - 1$:

$$\mathrm{Cov}(u_\gamma, u_\lambda) = E\Big[(I(x=0)e^\lambda - 1)(x/\lambda - 1)\Big] = e^\lambda E[I(x=0)(x/\lambda-1)] = e^\lambda \cdot e^{-\lambda}(-1) = -1,$$

since on $\{x=0\}$ we have $x/\lambda - 1 = -1$ and $P(x=0) = e^{-\lambda}$ (the -1 part has mean 0 as u_λ is a score). With $\mathrm{Var}(u_\lambda) = 1/\lambda$ (Poisson Fisher information), the efficient-score variance is

$$\mathrm{Var}(u_\gamma) - \frac{\mathrm{Cov}(u_\gamma, u_\lambda)^2}{\mathrm{Var}(u_\lambda)} = (e^\lambda - 1) - \frac{(-1)^2}{1/\lambda} = e^\lambda - 1 - \lambda.$$

这是标准 ZIP score 检验结果：γ-score 与 λ-score 协方差 $= -1 \neq 0$，必须扣掉 λ-投影 $\mathrm{Cov}^2/\mathrm{Var}(u_\lambda) = \lambda$，效率分母是 $e^\lambda - 1 - \lambda$（不是裸方差 $e^\lambda - 1$）。

Plugging in,

$$\boxed{\mathrm{LM}_n = \frac{\left(e^{\hat\lambda} n_0 - n\right)^2}{n\left(e^{\hat\lambda} - 1 - \hat\lambda\right)} \xrightarrow{d} \chi^2_1 \text{ under } H_0, \qquad \hat\lambda = \bar{\mathbf{x}}}$$

Reject at level α if $\mathrm{LM}_n > \chi^2_{1,1-\alpha}$. 直观：$e^{\hat\lambda} n_0 - n = n\big(e^{\hat\lambda}\hat p_0 - 1\big)$，其中 $\hat p_0 = n_0/n$ 是样本中 0 的比例；当观测到的 0 多于 Poisson 预测的 $e^{-\hat\lambda}$ 时 $S_n > 0$，检验侦测到「零膨胀」。

(b) LR p= 0.03（拒）与 score p= 0.45（不拒）严重背离。要从「score 在约束 MLE 处线性化、LR 看全局对数似然差」的本质讲。

The score test linearizes the log-likelihood at the *restricted* (Poisson) MLE and uses local curvature; the LR test compares the restricted and unrestricted maxima globally. Under the regular null here they are asymptotically equivalent, but they need not have similar finite-sample values. A gap as large as 0.03 versus 0.45 is evidence that the local quadratic approximation is poor for this sample (or that the two procedures were implemented differently), not a logical contradiction. Concretely:

- *Asymmetry / non-quadratic log-likelihood in γ.* The zero-inflation parameter enters very nonlinearly; the slope at $\gamma = 0$ (what the score sees) can be modest while the likelihood keeps rising substantially as γ moves away from 0 (what LR captures). The restricted-point gradient understates the global gain.
- *Near-boundary behavior.* Although $\gamma_0 = 0$ is interior (the lower bound is $-1/(e^{\lambda_0} - 1) < 0$), the unconstrained MLE $\hat\gamma$ may sit far out where curvature differs markedly from that at 0; the score's local Fisher information at 0 then mis-scales the statistic.
- *Little efficient information when λ is small.* After nuisance adjustment the information for γ is $I_{\gamma|\lambda} = e^\lambda - 1 - \lambda = \lambda^2/2 + O(\lambda^3)$, which can be tiny. Local normal/χ^2 calibration can then be inaccurate in moderate samples.

Bottom line: the discrepancy is a finite-sample warning. It does *not* follow from the two p-values alone that LR is correct and score is wrong. Check that both tests use the same two-sided null, the admissible parameter space, the efficient denominator $e^{\hat\lambda}-1-\hat\lambda$, and the same nuisance treatment; if the decision matters, calibrate the finite-sample distributions by a parametric bootstrap or Monte Carlo experiment under the fitted Poisson null.

Remark (Q7 score vs LR 背离的真正原因).

三检验（Wald/LR/LM）的等价是渐近结论；有限样本里，score 只看约束点的局部斜率，LR 看两个极值的全局高度。这里真正的有效信息是 $e^{\lambda}-1-\lambda\sim\lambda^2/2$，不是 $e^{\lambda}-1$。因此应把巨大背离解释为「局部二次近似或实现细节需要检查」，不能仅凭两个 p 值武断宣布 LR 更可信。

6.8 Q8（20%）多项 logit 的条件矩 GMM 与最优工具

归属：ECON 510（按内容归类） **重要度**：中高 **再考判断**：中
核心考点：多项 logit 条件矩 GMM 与最优工具
历史复现：2017–2022 最优工具；离散选择见 2020/2021
掌握方式：先写 cell probabilities 和 score，再识别条件矩

原题 Q8（20%）

Suppose that $\mathbf{y}_i$ denotes the product purchased by individual $i=1,\dots,n$, $\mathbf{x}_{ij}$ some observed product characteristics, and $\boldsymbol{\xi}_j$ an unobserved product characteristic, where $j=1,\dots,J$. There is an 'outside good' (no purchase), product 0. Suppose further that we know that

$$\mathrm{P}(\mathbf{y}_i=j\mid \mathbf{x}_{i\cdot},\boldsymbol{\xi})=\frac{\exp(\mathbf{x}_{ij}^T\theta_0+\boldsymbol{\xi}_j)}{1+\sum_{t=1}^J\exp(\mathbf{x}_{it}^T\theta_0+\boldsymbol{\xi}_t)}.$$

Finally, suppose that the object of estimation is θ_0, that J is fixed, and that $n\to\infty$. Use GMM for the questions below, even though it is not the most natural estimation method here. [20%]

(a) What conditional moments conditions would you use?

(b) Derive the optimal instruments.

(c) What would have been a more natural estimation method here?

Solution.

思路。把「选了产品 j」写成示性向量，与模型预测概率作差就是均值为 0 的条件矩残差。关键是不能把题面称为 unobserved 的 $\boldsymbol{\xi}$ 偷偷当成已知：它只有 product 下标，且 J fixed，所以应把 realized $\boldsymbol{\xi}=(\xi_1,\dots,\xi_J)'$ 当 finite-dimensional fixed unknown nuisance，与 θ 联合估计。最优工具仍用 Chamberlain 公式，但 derivative

要对 joint parameter 求。

Notation. For each i let $d_{ij} := I(\mathbf{y}_i = j)$, $j = 1, \ldots, J$, stacked into $d_i = (d_{i1}, \ldots, d_{iJ})'$. Let X_i be the $J \times p$ matrix whose jth row is $\mathbf{x}'_{ij}$, define

$$\eta = (\theta', \xi')', \qquad B_i = [\, X_i \;\; I_J \,],$$

and stack the inside-good probabilities into $P_i(\eta) = (P_1, \ldots, P_J)' \in \mathbb{R}^J$. The denominator's 1 normalizes the outside-good effect to zero.

(a) Conditional moment conditions. Since $\mathbb{E}(d_{ij} \mid X_i, \xi_0) = P_j(X_i; \eta_0)$, the choice-indicator residuals have conditional mean zero:

$$\boxed{\mathbb{E}(\, g_i(\eta_0) \mid X_i, \xi_0) = 0, \qquad g_i(\eta) := d_i - P_i(\eta) \in \mathbb{R}^J.}$$

这是对 inside goods 的 J 维条件矩。把它乘上任意 X_i-可测 matrix 就得到 unconditional moments。去掉 outside-good indicator 后，只要 outside probability 与各 inside probability 均为正，conditional covariance 在这 J 维上满秩。可实施估计必须同时解出 ξ；把 ξ 写入 conditioning set 不等于 econometrician 已经知道它。

(b) Optimal instruments. Chamberlain (1987) 最优工具：对 joint parameter η，在所有 X_i-可测工具里使 GMM 渐近方差最小的是

$$\begin{aligned} A^*_\eta(X_i) &= D_\eta(X_i)'\Omega_i^{-1}, \\ D_\eta(X_i) &:= \mathbb{E}\left(\frac{\partial g_i(\eta_0)}{\partial \eta'} \,\middle|\, X_i, \xi_0 \right), \\ \Omega_i &:= \operatorname{Var}(g_i(\eta_0) \mid X_i, \xi_0)\,. \end{aligned}$$

The conditional covariance of the J inside-good indicators is

$$\Omega_i = \operatorname{Diag}(P_i) - P_i P_i', \qquad (\Omega_i)_{jk} = P_j(\delta_{jk} - P_k).$$

Because $P_{i0} > 0$ and all $P_{ij} > 0$, Ω_i is positive definite. The multinomial-logit derivative with respect to both parameter blocks is

$$\frac{\partial P_i(\eta_0)}{\partial \eta'} = \Omega_i B_i, \qquad D_\eta(X_i) = -\Omega_i B_i.$$

Consequently,

$$A^*_\eta(X_i) = D_\eta(X_i)'\Omega_i^{-1} = -B_i'.$$

Multiplying every estimating equation by -1 changes neither its root nor its efficiency, so use the sign-equivalent optimal instrument

$$\boxed{\widetilde{A}^*_\eta(X_i) = B_i'.}$$

Its joint moment is

$$B_i'\{d_i - P_i(\eta)\} = \begin{pmatrix} X_i'\{d_i - P_i(\eta)\} \\ d_i - P_i(\eta) \end{pmatrix}.$$

This is exactly the joint multinomial-logit likelihood score. Identification requires

$$\boxed{I_\eta := \mathbb{E}\left(B_i'\Omega_i B_i\right) \ \text{nonsingular.}}$$

Equivalently, there must be no nonzero (a, b) such that $X_i a + b = 0$ almost surely; regressor columns collinear with product fixed effects must be removed or normalized. 若 ξ 已知，theta-only score 才简化为 $X_i'(d_i - P_i)$。题面里的 ξ 是 unobserved，所以该 oracle 公式不能替代 joint answer。

(c) More natural method. 模型把条件选择概率完全指定了，自然该用 MLE。

$$\boxed{\begin{array}{l}\text{Joint (or profile) conditional multinomial-logit MLE:} \\ (\hat\theta', \hat\xi')' = \arg\max\limits_{\theta,\xi} \displaystyle\sum_{i=1}^n \left[d_i'(X_i\theta + \xi) - \log\left\{1 + \sum_{j=1}^J e^{x_{ij}'\theta + \xi_j}\right\}\right].\end{array}}$$

Profile out ξ and retain the θ block. If one wants the efficient score for θ alone, partition the information as

$$\begin{aligned} I_{\theta\theta} &= \mathbb{E}\left(X_i'\Omega_i X_i\right), & I_{\theta\xi} &= \mathbb{E}\left(X_i'\Omega_i\right), \\ I_{\xi\xi} &= \mathbb{E}(\Omega_i), & I_{\xi\theta} &= I_{\theta\xi}'. \end{aligned}$$

Then

$$s_{\theta i}^{\text{eff}} = \{X_i' - I_{\theta\xi}I_{\xi\xi}^{-1}\}\{d_i - P_i\}, \qquad I_{\theta\cdot\xi} = I_{\theta\theta} - I_{\theta\xi}I_{\xi\xi}^{-1}I_{\xi\theta},$$

and $\text{Avar}\{\sqrt{n}(\hat\theta - \theta_0)\} = I_{\theta\cdot\xi}^{-1}$. The joint MLE performs this nuisance adjustment automatically. Thus optimal GMM and MLE use the same joint score, while ML is the more direct implementation.

Remark (Q8 最优工具 = joint MLE score).

这题的「啊哈」是：把 unobserved ξ 纳入 joint parameter 后，$\partial P/\partial \eta' = \Omega[X\ I_J]$，所以 Chamberlain 最优工具化成 $-[X\ I_J]'$；相应 moments 正是 joint logit score。只写 $-X'$ 等于做了「ξ 已知」的 oracle 题，不是原题。

Remark (本章小结 · 2022 卷方法清单).

Q1 吸收链首达 = 一步条件解线性方程，期望吸收时间 = 几何均值 $1/(1-q)$；Q2 好集法 + 条件 ii) 把柱集拆成矩形可数并，反例用 singletons 与 countable–cocountable σ 域；Q3 LLN/CMT→plim、多元 CLT、delta 法、$o_p(1/\sqrt{n})$ 等价化四

连击；Q4 六个真伪全为 False，逐一点名 UI / 多重比较 / Jensen 方向 / ancillary / χ^2 差 / 条件均值零 $\neq$ 不相关；Q5 M 估计量夹心方差 $H^{-1}VH^{-1}$；Q6 只写定义上必然成立的蕴含，遇到含混术语先声明口径；Q7 score 在 Poisson MLE 处求值得 $\mathrm{LM} = (e^{\hat{\lambda}}n_0 - n)^2/[n(e^{\hat{\lambda}} - 1 - \hat{\lambda})]$，巨大背离提示有限样本二次近似或实现细节需检查；Q8 把 fixed-J product effects 纳入 $\eta = (\theta', \xi')'$，$\partial P/\partial\eta' = \Omega[\,X\ I_J\,]$，optimal GMM 正是 joint logit score，natural method 是 joint/profile MLE。

Chapter 7　2023 资格考精解（Worked Solutions）

本章逐题给出 2023 年 8 月资格考的完整解答。试卷由两位老师署名：前半（Q1–Q4，510 进阶）为 Patrik Guggenberger，考渐近 size ／ UMP、非光滑 GMM 与分位回归、线性 IV 点识别与 Das–Newey–Vella 选择模型、ridge ／ LASSO；后半（Pinkse Q1–Q8，501 基础）为 Joris Pinkse 的 8 道短题，他明确要求每题十分钟内、答案不超过几行。中文负责思路与直觉，英文负责形式推导与方框最终答案。许多题正是前面章节里已经做过的范例——本章只点章节、套公式、独立重推，不重复讲理论。

Part A ·Guggenberger（510 进阶）

7.1 Q1(a) · 渐近 size 与渐近原假设拒绝概率 / Asymptotic Size vs. Null Rejection Probability

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：点态 NRP 与 uniform asymptotic size
历史复现：2023、2024、2025 连续出现
掌握方式：定义中的 $\limsup_n \sup_\gamma$ 顺序必须背准；例子要会解释

原题 Q1(a)

Formally define what it means for a test to control (i) the (asymptotic) null rejection probability and (ii) the (asymptotic) size. Provide an example of a test where for a certain parameter space the asymptotic null rejection probability is controlled but not the asymptotic size. What is the issue for applied researchers when employing such a test?

Solution.

中文思路. 两个概念的差别全在**逐点极限**与**对参数空间取上确界**之分。原假设拒绝概率控制是“对每个固定的零分布，$n \to \infty$ 的拒绝概率极限 $\le \alpha$”；渐近 size 控制要求“先对零参数空间取 sup、再取极限上确界 $\le \alpha$”，即拒绝概率**一致地**被压住。两者可以背离：当极限在参数空间内不一致（存在“坏点序列”，如弱工具、近边界、近单位根），逐点都达标但 sup 飘到 α 之上。整套框架见 13 的渐近 size。

Formal. Let the data have law P_γ indexed by $\gamma \in \Gamma$, and let $\Gamma_0 \subseteq \Gamma$ be the null. A test ϕ_n (rejection region R_n) of nominal level α:

(i) controls the *asymptotic null rejection probability* if

$$\limsup_{n\to\infty} \mathrm{P}_\gamma(R_n) \le \alpha \qquad \text{for each fixed } \gamma \in \Gamma_0,$$

a *pointwise* (in γ) statement.

(ii) controls the *asymptotic size* if

$$\mathrm{AsySz} := \limsup_{n\to\infty} \sup_{\gamma\in\Gamma_0} \mathrm{P}_\gamma(R_n) \le \alpha,$$

where the sup over Γ_0 is taken *inside*, *before* the limit —a *uniform* statement. Always $\sup_\gamma \limsup_n \le \limsup_n \sup_\gamma$, so (ii)⇒(i) but not conversely.

Example (controlled NRP, uncontrolled size). Take the just-identified linear IV t-test of $H_0 : \beta = \beta_0$, with parameter $\gamma = (\beta, \pi, \dots)$ including the first-stage strength π, and $\Gamma_0 = \{\gamma : \beta = \beta_0\}$ allowing π arbitrarily close to 0 (weak instruments). For each *fixed* $\pi \neq 0$ the usual t-statistic is $\xrightarrow{d} N(0,1)$ under H_0, so $\lim_n \mathrm{P}_\gamma(|t_n| > z_{1-\alpha/2}) = \alpha$: the asymptotic NRP is controlled at every fixed point. But along the drifting sequence $\pi_n = c/\sqrt{n}$ (weak-IV embedding) the t-statistic has a non-standard limit whose tail mass exceeds α, so

$$\sup_{\gamma\in\Gamma_0} \mathrm{P}_\gamma\big(|t_n| > z_{1-\alpha/2}\big) \not\to \alpha, \qquad \mathrm{AsySz} > \alpha.$$

机理：$\pi = 0$（或 π 任意小）是参数空间里的不连续点，逐点极限在它附近不一致收敛。漂移序列 $\pi_n = c/\sqrt{n}$ 把这个坏点“拉进”每一个有限 n，于是 worst case 超额。其它同型例子：近单位根 AR、参数在边界的检验、pretest 后推断。

对应用研究者的问题（the issue）. 渐近 NRP 受控只是**逐点**保证：对任何**固定**真值，样本足够大就接近名义水平。但研究者手里只有**一个有限样本**，且不知道自己离“坏点”（弱工具、近边界）有多近。size 失控意味着：存在数据生成过程，使得**无论** n **多大**该检验的实际 size 都严格高于 α——

有限样本里真实第一类错误率可被任意拉高到 α 之上，置信区间名不副实（欠覆盖），“大样本”救不了。

解药是用一致有效的方法（如 weak-IV 稳健的 AR ／ CLR 检验，见 14、14）。

7.2 Q1(b) · UMP 存在的情形与无 UMP 时的甄选准则 / When a UMP Exists, and Criteria When It Does Not

归属：ECON 510（Guggenberger） **重要度**：高 **再考判断**：高
核心考点：UMP 存在、无 UMP 时的甄选准则
历史复现：2017、2018、2019、2023、2025
掌握方式：背 NP/MLR；无 UMP 时会写 invariance、similarity、local power

原题 Q1(b)

In a scenario where there are several tests that control the size and one of them is uniformly most powerful (UMP), applied researchers should use that test. However, often UMP tests do not exist. (a) Give a (nontrivial) example where there exists a UMP test. Prove your claim. (b) In a situation where no UMP test exists (in the class of tests that control size) describe general criteria according to which one could still potentially single out a recommended test. Provide an example of that strategy (with details) that we dealt with in class.

Solution.

(a) 中文思路：单调似然比 + Karlin–Rubin（6）. 单参数指数族对单边假设有 UMP 检验，根因是 MLR 结构：似然比随充分统计量单调，使得 N–P 最优拒绝域对**所有**备择参数同时是同一个域，从而“最优”可以一致达成。

Formal claim & proof. Let $X \sim p_\theta$ form a one-parameter family with density $p_\theta(x) = h(x)\exp\{\eta(\theta)T(x) - A(\theta)\}$, η strictly increasing (one-parameter exponential family ⇒ MLR in $T(x)$). Test $H_0:\theta\le\theta_0$ vs. $H_1:\theta>\theta_0$.

Concrete instance: $X_1,\dots,X_n \overset{i.i.d.}{\sim} N(\theta,1)$, $T=\sum_i X_i$, test $H_0:\theta\le\theta_0$ vs. $\theta>\theta_0$. Claim: $\phi^*=\mathbb{1}\{\bar X>\theta_0+z_{1-\alpha}/\sqrt{n}\}$ is UMP of size α.

Proof for Theorem

> *Step 1 (size).* Under $\theta=\theta_0$, $\bar X\sim N(\theta_0,1/n)$, so $\mathrm{P}_{\theta_0}(\phi^*=1)=\alpha$; since the power $\theta\mapsto \mathrm{P}_\theta(\phi^*=1)=1-\Phi(z_{1-\alpha}-\sqrt{n}(\theta-\theta_0))$ is increasing in θ, $\sup_{\theta\le\theta_0}\mathrm{P}_\theta(\phi^*=1)=\alpha$. Size controlled.
>
> *Step 2 (UMP via Neyman–Pearson + MLR).* Fix any $\theta_1>\theta_0$. The likelihood ratio
> $$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}=\exp\Big\{(\theta_1-\theta_0)\textstyle\sum_i x_i-\frac{n}{2}(\theta_1^2-\theta_0^2)\Big\}$$
> is strictly increasing in $T=\sum_i x_i$ (since $\theta_1-\theta_0>0$). By Neyman–Pearson, the most powerful size-α test of the *simple* θ_0 vs. θ_1 rejects for large T, i.e. exactly $\{\bar X>\theta_0+z_{1-\alpha}/\sqrt{n}\}=\phi^*$. The critical region does *not depend on the particular* $\theta_1>\theta_0$, so ϕ^* is simultaneously MP against every $\theta_1>\theta_0$. By Karlin–Rubin this MLR test is also valid against the composite null $\theta\le\theta_0$. Hence ϕ^* is UMP. ■

One-sided test in an MLR (e.g. one-parameter exponential) family has a UMP test (Karlin–Rubin).

“非平凡”关键词：UMP 在两边假设 $H_1:\theta\ne\theta_0$ 一般**不**存在，因为左右两侧的 N–P 最优拒绝域方向相反，无法用同一个域一致最优——这正是过渡到 (b) 的动机。

(b) 中文思路：无 UMP 时，缩小竞争类或换最优性准则. 既然“对所有备择一致最优”做不到，就有两条路：(1) 加一条合理的约束把检验类缩小，使 UMP 在**子类**里复活；(2) 把“一致最优”换成一个标量化的最优性目标（加权平均功效／极大极小）。课上讲过的标准策略：

- **无偏性 ⇒ UMPU.** 要求 $\mathbb{E}_\theta\phi\ge\alpha$ 对所有 $\theta\in H_1$（功效不低于水平）。在此子类里两边检验的 UMPU 存在。
- **不变性 ⇒ UMP invariant.** 在群作用下要求检验不变，把问题约化到极大不变量，常使 UMP 在不变类里复活（14）。
- **加权平均功效（WAP）/ 极大极小 / 最不利分布.** 选一个先验权重 w 把功效曲面标量化，最大化 $\int\beta_\phi(\theta)\,w(\mathrm{d}\theta)$；或在 H_1 上最大化最小功效。

Detailed in-class example: the two-sided z-test as UMPU. $X_1,\dots,X_n \overset{i.i.d.}{\sim} N(\theta,1)$, test $H_0:\theta=\theta_0$ vs. $H_1:\theta\ne\theta_0$. No UMP test exists (one-sided rejection regions beat the two-sided one on their own side). Restrict to *unbiased* size-α tests. An unbiased test must have power $\ge\alpha$ everywhere and, being differentiable here, must satisfy the first-order (similar-on-the-boundary) condition

$\partial_\theta \mathbb{E}_\theta \phi|_{\theta_0} = 0$. Within this class the generalized Neyman–Pearson lemma yields the symmetric two-sided region

$$\boxed{\phi^{\text{UMPU}} = \mathbb{1}\{\sqrt{n}\,|\bar{X} - \theta_0| > z_{1-\alpha/2}\},}$$

which is UMP *among unbiased* (UMPU) size-α tests. *(Alternative in-class strategy: invariance + CLR for weak IV, 14; or weighted-average-power optimal tests.)* 要点链条：UMP 不存在 → 加"无偏"这一自然约束 → 子类里 UMPU 复活 → 落到对称双边 z 检验。这条"缩类换最优"的思路在弱 IV（CLR= 某 WAP 意义下近最优的不变检验）里反复出现。

7.3 Q2 · 非光滑 GMM 的 V_0, Γ 估计与分位回归 / Nonsmooth GMM: Estimating V_0, Γ, and Quantile Regression

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：非光滑 GMM 的 V_0, Γ 与分位回归
历史复现：2023、2024、2025 连续同构出现
掌握方式：通用估计式与 QR 特化必须逐行背熟并理解符号

原题 Q2

Under Assumptions EE3 and CF-NS we derived $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\big(0, (\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1}\big)$ for the GMM estimator with nonsmooth criterion $Q_n(\theta) = \|\bar{g}_n(\theta)\|$, $\bar{g}_n(\theta) = n^{-1}\sum_{i=1}^n g(W_i, \theta)$, $g(\theta) = \mathbb{E}(g(W_i, \theta))$. **(a)** Provide estimators for V_0 and Γ and discuss their consistency under appropriate conditions. For the quantile-regression example $(Y_i, X_i')'$ i.i.d., $Y_i = X_i'\theta_0 + U_i$ with τ-quantile of $U_i \mid X_i$ equal to 0: **(b)** derive V_0 and Γ; **(c)** find a simple expression for $(\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1}$. [CF-NS: θ_0 interior; g differentiable at θ_0 with $\Gamma = \partial g(\theta_0)/\partial\theta'$ of full rank $d \le k$; $g(\theta_0) = 0$; $\sqrt{n}\,\bar{g}_n(\theta_0) \xrightarrow{d} N(0, V_0)$; stochastic equicontinuity. EE3: $Q_n(\hat{\theta}_n) = \inf_\theta Q_n + o_p(n^{-1/2})$, $\hat{\theta}_n \xrightarrow{p} \theta_0$.]

Solution.

本题正是 11 的非光滑 GMM 模板的直接应用——只点公式、套估计量、做分位回归算例。

(a) 中文思路. V_0 是矩函数在真值处的渐近方差，i.i.d. 下就是矩外积的总体均值，用样本外积加 plug-in $\widehat{\theta}$ 即可。$\Gamma = \partial g(\theta_0)/\partial\theta'$ 是**期望**矩 $g(\theta) = \mathbb{E}(g(W, \theta))$ 的雅可比——注意 $g(\theta)$ 光滑（期望把示性函数的跳变抹平了），但样本矩 $\bar{g}_n(\theta)$ 不光滑，故不能逐样本求导，要用**数值差分**（或核平滑）来估 Γ。

Estimating V_0. Under i.i.d. and $g(\theta_0)=0$,

$$V_0=\operatorname{Var}(g(W_i,\theta_0))=\mathbb{E}\big(g(W_i,\theta_0)g(W_i,\theta_0)'\big).$$

A consistent estimator is the sample outer product at $\widehat{\theta}$:

$$\boxed{\widehat{V}_0=\frac{1}{n}\sum_{i=1}^{n}g(W_i,\widehat{\theta})\,g(W_i,\widehat{\theta})'.}$$

Consistency: by a uniform LLN over a neighborhood of θ_0 plus $\widehat{\theta}\xrightarrow{p}\theta_0$ and continuity of $\theta\mapsto\mathbb{E}(gg')$ at θ_0, $\widehat{V}_0\xrightarrow{p}V_0$ (finite second moments of g required). 时间序列相依时改用 Newey–West（HAC）核加权和以纳入自协方差。

Estimating Γ. Since $g(W,\theta)$ is nonsmooth in θ (e.g. contains $\mathbb{1}\{\cdot\}$), use a *numerical derivative* of the sample analogue $\bar{g}_n$ to estimate the derivative of the smooth population moment $g(\theta)$: with step $\varepsilon_n\to 0$ and unit vectors e_j,

$$\boxed{\widehat{\Gamma}_{\cdot j}=\frac{\bar{g}_n(\widehat{\theta}+\varepsilon_n e_j)-\bar{g}_n(\widehat{\theta}-\varepsilon_n e_j)}{2\varepsilon_n},\qquad j=1,\dots,d.}$$

Consistency: Lecture 16 assumes $\varepsilon_n\to 0$ and $1/(\sqrt{n}\,\varepsilon_n)=O_p(1)$; CF-NS(v) then makes the amplified local empirical remainder $o_p(1)$. The stronger, easy-to-remember sufficient choice $\sqrt{n}\,\varepsilon_n\to\infty$ also works. Together with differentiability of $g(\theta)$ at θ_0 and $\widehat{\theta}\xrightarrow{p}\theta_0$, this gives $\widehat{\Gamma}\xrightarrow{p}\Gamma$. 两难（bandwidth dilemma）：$\varepsilon_n$ 太大则差分不再瞄准 local derivative；太小则会放大 local empirical remainder。可背 $\varepsilon_n=n^{-a}$、$0<a<1/2$ 作为安全选择。

(b) 中文思路：分位回归算例. 分位回归的"一阶条件"矩函数 $g(W_i,\theta)=X_i(\tau-\mathbb{1}\{Y_i\le X_i'\theta\})$（来自 check 函数次梯度）。在 θ_0 处 $\{Y_i\le X_i'\theta_0\}=\{U_i\le 0\}$，由条件 τ-分位为 0 得 $\mathrm{P}(U_i\le 0\mid X_i)=\tau$。

Compute V_0. At θ_0, $g(W_i,\theta_0)=X_i(\tau-\mathbb{1}\{U_i\le 0\})$. The scalar $\tau-\mathbb{1}\{U_i\le 0\}$ is mean-zero given X_i (since $\mathbb{E}(\mathbb{1}\{U_i\le 0\}\mid X_i)=\tau$) and has conditional variance $\operatorname{Var}(\mathbb{1}\{U_i\le 0\}\mid X_i)=\tau(1-\tau)$. Hence

$$V_0=\mathbb{E}\big(gg'\big)=\mathbb{E}\big(X_iX_i'(\tau-\mathbb{1}\{U_i\le 0\})^2\big)=\mathbb{E}\big(X_iX_i'\tau(1-\tau)\big)=\boxed{\tau(1-\tau)\,\mathbb{E}\big(X_iX_i'\big).}$$

Compute Γ. The smoothed moment is $g(\theta)=\mathbb{E}(X_i(\tau-\mathbb{1}\{Y_i\le X_i'\theta\}))=\mathbb{E}\big(X_i(\tau-F_{U|X}(X_i'(\theta-\theta_0)))\big)$. Differentiate at θ_0 (chain rule, $\partial_\theta F_{U|X}(X_i'(\theta-\theta_0))=f_{U|X}(0)X_i'$):

$$\Gamma=\frac{\partial g(\theta_0)}{\partial\theta'}=-\mathbb{E}\big(X_i\,f_{U|X_i}(0)\,X_i'\big)=\boxed{-\,\mathbb{E}\big(f_{U|X_i}(0)\,X_iX_i'\big).}$$

符号说明：Γ 带负号（$\mathbb{1}\{Y\le X'\theta\}$ 随 θ 增大），但下面三明治里 Γ 成对出现、负号平方抵消，不影响方差。文献里常吸收负号写 $\Gamma=\mathbb{E}\big(f_{U|X}(0)XX'\big)$。这里 $d=k$

（恰好识别，Γ 方阵可逆）。

(c) 中文思路：方阵情形三明治塌缩. 此处 Γ 是 $k\times k$ 满秩方阵，$(\Gamma'\Gamma)^{-1}\Gamma' = \Gamma^{-1}$，三明治退化为 $\Gamma^{-1}V_0\Gamma^{-1\prime}$。

Formal. For square invertible Γ, $(\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1} = \Gamma^{-1}V_0(\Gamma')^{-1}$. Writing $J := \mathbb{E}\big(f_{U|X}(0)XX'\big)$ (so $\Gamma = -J$, symmetric) and $V_0 = \tau(1-\tau)\mathbb{E}(XX')$,

$$\boxed{\mathrm{AVar} = \tau(1-\tau)\left[\,\mathbb{E}\big(f_{U|X}(0)XX'\big)\,\right]^{-1}\mathbb{E}\big(XX'\big)\,\left[\,\mathbb{E}\big(f_{U|X}(0)XX'\big)\,\right]^{-1}.}$$

This is exactly Koenker's classical quantile-regression sandwich covariance. 条件密度同方差特例 $f_{U|X}(0) \equiv f_U(0)$ 时进一步塌缩为 $\dfrac{\tau(1-\tau)}{f_U(0)^2}\big(\mathbb{E}(XX')\big)^{-1}$。

7.4　Q3(a) · 线性 IV 模型的点识别（秩条件）/ Point Identification of Linear IV via the Rank Condition

归属：ECON 510（Guggenberger）　**重要度**：高　**再考判断**：高
核心考点：线性 IV 的点识别与秩条件
历史复现：2017–2022 IV/GMM；2023、2025 识别概念
掌握方式：从 reduced form 写可观测矩阵并证明满列秩

原题 Q3(a)

Consider $y = x'\beta + z_1'\gamma + \varepsilon$, $x = \Pi z + \eta$, where $z = (z_1', z_2')' \in \mathbb{R}^{d_z}$, $z_j \in \mathbb{R}^{d_{z_j}}$ $(j = 1, 2)$, $\mathbb{E}(z\varepsilon) = 0$, $\mathbb{E}(z\eta') = 0$, and $\mathbb{E}(zz')$ nonsingular. Show that $\theta = (\beta', \gamma')'$ is point identified under the rank condition $\mathrm{rk}(\Pi_2) = d_x$, where $\Pi = (\Pi_1 : \Pi_2) \in \mathbb{R}^{d_x\times d_z}$, $\Pi_j \in \mathbb{R}^{d_x\times d_{z_j}}$.

Solution.

中文思路. 这是 7 线性 IV 识别的标准计数 + 秩条件。回归元 $w = (x', z_1')'$（共 $d_x + d_{z_1}$ 维），工具 z（共 $d_z = d_{z_1} + d_{z_2}$ 维）。识别 = 总体矩方程 $\mathbb{E}(z(y - w'\theta)) = 0$ 有唯一解。由 $\mathbb{E}(z\varepsilon) = 0$，真值满足该方程；唯一性 $\iff$ 系数矩阵 $\mathbb{E}(zw')$ **列满秩**。把 $\mathbb{E}(zw')$ 拆块，用 $x = \Pi z + \eta$、$\mathbb{E}(z\eta') = 0$ 算出它等于 $\mathbb{E}(zz')$ 乘一个由 Π 和选择矩阵搭成的块阵，列满秩 $\iff$ $\mathrm{rk}(\Pi_2) = d_x$。

Formal. Stack regressors $w = \begin{pmatrix} x \\ z_1 \end{pmatrix} \in \mathbb{R}^{d_x + d_{z_1}}$. The population moment condition is

$$\mathbb{E}\big(z(y - w'\theta)\big) = 0.$$

Because $\mathbb{E}(z\varepsilon) = 0$, θ_0 solves it. The solution is *unique* (point identification) iff the $d_z \times (d_x + d_{z_1})$ matrix $M := \mathbb{E}(zw')$ has full column rank $d_x + d_{z_1}$. Compute the two blocks using $x = \Pi z + \eta$ and $\mathbb{E}(z\eta') = 0$:

$$\mathbb{E}\big(zx'\big) = \mathbb{E}\big(z(\Pi z + \eta)'\big) = \mathbb{E}\big(zz'\big)\,\Pi', \qquad \mathbb{E}\big(zz_1'\big) = \mathbb{E}\big(zz'\big)\ S_1',$$

where $S_1 = (I_{d_{z_1}} : 0)$ selects z_1 from z (so $z_1 = S_1 z$). Hence

$$M = \mathbb{E}\big(zw'\big) = \mathbb{E}\big(zz'\big)\,\big[\,\Pi' \;:\; S_1'\,\big].$$

Since $\mathbb{E}(zz')$ is nonsingular, $\mathrm{rk}(M) = \mathrm{rk}\big([\Pi' : S_1']\big)$, an $d_z \times (d_x + d_{z_1})$ matrix. Partition rows by $z = (z_1', z_2')'$:

$$[\Pi' : S_1'] = \begin{pmatrix} \Pi_1' & I_{d_{z_1}} \\ \Pi_2' & 0 \end{pmatrix} \quad \text{(rows } z_1 \text{ on top, } z_2 \text{ below).}$$

Full column rank: suppose $[\Pi' : S_1']\binom{a}{b} = 0$ with $a \in \mathbb{R}^{d_x}$, $b \in \mathbb{R}^{d_{z_1}}$. The bottom block gives $\Pi_2' a = 0$; under $\mathrm{rk}(\Pi_2) = d_x$ (i.e. Π_2' has full column rank d_x) this forces $a = 0$. The top block then gives $b = 0$. So the only null combination is trivial:

$$\boxed{\mathrm{rk}(\Pi_2) = d_x \implies \mathrm{rk}(M) = d_x + d_{z_1} \implies \theta = (\beta', \gamma')' \text{ is point identified.}}$$

解释：z_2 是排除性工具（只进第一阶段、不进结果方程），$\mathrm{rk}(\Pi_2) = d_x$ 就是“工具相关性／秩条件”——z_2 对内生 x 的 d_x 个分量有 d_x 维的独立解释力，恰好把 β 识别出来；z_1 既当回归元又当自己的工具，识别 γ。这是 2SLS 可行的充要秩条件。

7.5 Q3(b) · Das–Newey–Vella 非参选择模型与导数识别 / Das–Newey–Vella Selection: Identifying the Derivative

归属：ECON 510（Guggenberger） **重要度**：中 **再考判断**：中低
核心考点：Das–Newey–Vella 非参选择识别
历史复现：仅 2023 直接点名；选择识别在 2024/2025 以 Heckit 再现
掌握方式：记模型、support/exclusion 条件和导数识别结论

原题 Q3(b)

Define the nonparametric selection model of Das, Newey, and Vella (2003) discussed in class. Without proof, specify conditions that are sufficient for identification of the derivative of the unknown function in the main equation.

Solution.

中文思路（7 的选择／控制函数）. DNV 把 Heckit 的线性主方程与联合正态假设同时放松。被选择样本中的条件均值可写成未知结构函数 $m(x)$ 加一个只依赖 propensity score 的未知选择修正 $\lambda(\pi(x,z))$。关键不是口头说“固定 propensity score”，而是利用排除变量 z 的导数先识别 $\lambda'(\pi)$，再从 x 方向的总导数里减掉选择项，得到 $\partial m(x)/\partial x$。

The DNV model. The observables are (Y, X, Z, D), with latent outcome and selection rule

$$Y^* = m(X) + \varepsilon, \qquad Y = DY^*, \qquad D = \mathbb{1}\{\pi(X, Z) \geq \eta\}, \qquad (\varepsilon, \eta) \perp (X, Z).$$

The excluded variable Z enters the selection function $\pi(X, Z)$ but does not enter the structural outcome function $m(X)$. Assume η has a continuous distribution. By the probability integral transform we may normalize $\eta \sim \mathrm{Unif}(0, 1)$ without loss of generality. Then

$$\pi(x, z) = \mathrm{P}(D = 1 \mid X = x, Z = z) = \mathbb{E}(D \mid X = x, Z = z)$$

is identified from the observable selection probability.

On the selected sample,

$$\begin{aligned} r(x, z) &:= \mathbb{E}(Y \mid X = x, Z = z, D = 1) \\ &= m(x) + \mathbb{E}(\varepsilon \mid \eta \leq \pi(x, z)) \\ &= m(x) + \lambda(\pi(x, z)), \end{aligned}$$

where the unknown selection-correction function is

$$\lambda(s) := \frac{1}{s} \int_0^s \mathbb{E}(\varepsilon \mid \eta = t)\, dt, \qquad 0 < s < 1.$$

Thus $r(x, z)$ and $\pi(x, z)$ are identified from observables, while both m and λ are unknown.

Sufficient conditions. At an interior support point (x, z), assume:

(i) the model, exclusion, independence, and continuous-error normalization above hold;

(ii) $r(x, z)$, $m(x)$, $\lambda(s)$, and $\pi(x, z)$ are continuously differentiable in the relevant arguments;

(iii) the exclusion is operative through the rank condition

$$\frac{\partial \pi(x, z)}{\partial z} \neq 0_{d_z},$$

so changing z changes selection while leaving $m(x)$ fixed;

(iv) the support of (X, Z) contains a neighborhood needed for the derivatives (and $0 < \pi(x, z) < 1$).

Identification calculation. Differentiate the observed selected-sample regression $r(x, z) = m(x) + \lambda(\pi(x, z))$ with respect to the excluded z:

$$\frac{\partial r(x, z)}{\partial z} = \lambda'(\pi(x, z)) \frac{\partial \pi(x, z)}{\partial z}.$$

Let $a(x,z) := \partial\pi(x,z)/\partial z$ be a nonzero $d_z \times 1$ vector. Then the scalar derivative of the unknown correction is identified by projection onto a:

$$\boxed{\lambda'(\pi(x,z)) = \frac{a(x,z)'\,\partial r(x,z)/\partial z}{a(x,z)'a(x,z)}.}$$

Now differentiate in the outcome regressor x:

$$\frac{\partial r(x,z)}{\partial x} = \frac{\partial m(x)}{\partial x} + \lambda'(\pi(x,z))\frac{\partial \pi(x,z)}{\partial x}.$$

Every term on the right except $\partial m(x)/\partial x$ has just been identified, so

$$\boxed{pdm(x)x = \frac{\partial r(x,z)}{\partial x} - \lambda'(\pi(x,z))\frac{\partial \pi(x,z)}{\partial x}.}$$

This is the requested identification result. The level of m may still be normalized only up to an additive constant shared with λ, but its derivative is unaffected and is point identified. 考场最短合格链：写出 $r(x,z) = m(x) + \lambda(\pi(x,z))$；列 rank condition $\partial_z\pi \neq 0$；由 $\partial_z r = \lambda'(\pi)\partial_z\pi$ 识别 $\lambda'(\pi)$；再用 $\partial_x m = \partial_x r - \lambda'(\pi)\partial_x\pi$。缺少中间那条 z 导数，就没有真正说明如何把未知选择修正从 x 导数里剥离。

7.6 Q4(a) · 岭估计的偏差与协方差 / Bias and Covariance of the Ridge Estimator

归属：ECON 510（Guggenberger） **重要度**：高 **再考判断**：高
核心考点：Ridge 条件偏差与协方差
历史复现：2023、2025
掌握方式：矩阵代入现场推；偏差符号和 sandwich 顺序不能错

原题 Q4(a)

Model $Y = X'\beta + e$, $\beta \in \mathbb{R}^p$, $\mathbb{E}(e \mid X) = 0$, i.i.d. observations. Show that $\hat\beta_{\text{ridge}} = (X'X + \lambda I_p)^{-1}X'Y$ with fixed $\lambda > 0$ has bias$(\hat\beta_{\text{ridge}} \mid X) = -\lambda(X'X + \lambda I_p)^{-1}\beta$ and $\text{Var}\left(\hat\beta_{\text{ridge}} \mid X\right) = (X'X + \lambda I_p)^{-1}(X'DX)(X'X + \lambda I_p)^{-1}$, $D := \mathbb{E}(ee' \mid X)$.

Solution.

中文思路. 纯线性代数 + 条件期望／方差（15）。代入 $Y = X\beta + e$，条件在 X 上把 $A := (X'X + \lambda I_p)^{-1}$ 当常数，$\mathbb{E}(e \mid X) = 0$ 拿掉随机部分得偏差；方差用 $\text{Var}(Ae \mid X) = A(X'\cdot \text{Var}(e \mid X) \cdot X)A'$。

Formal. Write $A := (X'X + \lambda I_p)^{-1}$ (symmetric). Plug $Y = X\beta + e$:

$$\hat\beta_{\text{ridge}} = AX'(X\beta + e) = AX'X\,\beta + AX'e.$$

Bias. Conditional on X, $\mathbb{E}(AX'e \mid X) = AX'\mathbb{E}(e \mid X) = 0$, so

$$\mathbb{E}\left(\hat{\beta}_{\text{ridge}} \mid X\right) = AX'X\,\beta = A(X'X + \lambda I_p - \lambda I_p)\beta = AA^{-1}\beta - \lambda A\beta = \beta - \lambda A\beta.$$

Hence

$$\boxed{\text{bias}(\hat{\beta}_{\text{ridge}} \mid X) = \mathbb{E}\left(\hat{\beta}_{\text{ridge}} \mid X\right) - \beta = -\lambda(X'X + \lambda I_p)^{-1}\beta.}$$

Covariance. Only $AX'e$ is random given X:

$$\begin{aligned}\text{Var}\left(\hat{\beta}_{\text{ridge}} \mid X\right) &= \text{Var}\big(AX'e \mid X\big) = AX'\,\mathbb{E}\big(ee' \mid X\big)\;XA' \\ &= \boxed{(X'X + \lambda I_p)^{-1}(X'DX)(X'X + \lambda I_p)^{-1}.}\end{aligned}$$

$\lambda \to 0$ 偏差 $\to 0$、方差 $\to (X'X)^{-1}X'DX(X'X)^{-1}$（白噪声异方差稳健 OLS 三明治），与 OLS 衔接。$D = \sigma^2 I$ 同方差时方差 $= \sigma^2 AX'XA$。

7.7 Q4(b) ·岭 vs. OLS 的 MSE 比较 / Ridge Beats OLS in MSE

归属：ECON 510（Guggenberger） **重要度**：中高 **再考判断**：中
核心考点：Ridge 与 OLS 的 MSE 比较
历史复现：2023；Ridge 构造 2025 再现
掌握方式：会拆 bias squared + variance，并用特征值界

原题 Q4(b)

Show that $\text{mse}(\hat{\beta}_{\text{ridge}} \mid X) < \text{mse}(\hat{\beta}_{\text{OLS}} \mid X)$ if $2\lambda_{\min}(D) > \lambda\,\beta'\beta$, where $\lambda_{\min}(D)$ is the smallest eigenvalue of $D = \mathbb{E}(ee' \mid X)$ and λ is the ridge penalty.

Solution.

中文思路. MSE = trace(方差 + 偏差外积)。关键不靠对小 λ 做 Taylor，而用一条**精确恒等式**（Theobald 1974）：记 $S := X'X$、$A := (S + \lambda I)^{-1}$、$V := X'DX$，则 $S^{-1} = A + \lambda AS^{-1}$，把 OLS 的三明治 $S^{-1}VS^{-1}$ 精确拆成「岭的方差 AVA」加上一项**严格正**的修正。于是 MSE 之差可写成闭式，对**任意**满足条件的固定 λ（不限小 λ）都成立。再把方差降幅用 $\lambda_{\min}(D)$ 从下界压住、把平方偏差从上界压住，得到充分条件 $2\lambda_{\min}(D) > \lambda\beta'\beta$。

Setup. Let $S := X'X$ (symmetric, $\succ 0$), $A := (S + \lambda I_p)^{-1}$, $V := X'DX \succeq 0$. Scalar (trace) MSE conditional on X:

$$\text{mse} = \text{tr}\big(\text{Var}(\cdot \mid X)\big) + \|\text{bias}\|^2.$$

OLS ($\lambda = 0$): bias 0, $\text{mse}_{\text{OLS}} = \text{tr}\big(S^{-1}VS^{-1}\big)$.

Ridge: $\mathrm{mse}_{\mathrm{ridge}} = \mathrm{tr}\big(AVA\big) + \lambda^2\beta' A^2\beta$ (from Q4(a)).

Exact decomposition (Theobald 1974, no small-λ approximation). From $S + \lambda I = A^{-1}$, multiply $A^{-1} = S + \lambda I$ on the left by A and on the right by S^{-1}:

$$S^{-1} = A\,(S+\lambda I)\,S^{-1} = A + \lambda A S^{-1}.$$

This identity holds for *every* $\lambda > 0$. Substitute it into both factors of the OLS sandwich:

$$S^{-1}VS^{-1} = (A+\lambda AS^{-1})\,V\,(A+\lambda S^{-1}A) = AVA+\lambda\big(AS^{-1}VA+AVS^{-1}A\big)+\lambda^2 AS^{-1}VS^{-1}A,$$

using symmetry of A, S. Taking traces and subtracting $\mathrm{tr}(AVA)$ gives the exact variance reduction

$$\Delta_{\mathrm{var}} := \mathrm{tr}(S^{-1}VS^{-1}) - \mathrm{tr}(AVA) = 2\lambda\,\mathrm{tr}\big(AVS^{-1}A\big) + \lambda^2\,\mathrm{tr}\big(AS^{-1}VS^{-1}A\big) \;\geq\; 2\lambda\,\mathrm{tr}\big(S^{-1}AVA\big),$$

where the last step drops the nonnegative λ^2 term ($S^{-1}VS^{-1} \succeq 0$ sandwiched by A) and uses cyclicity $\mathrm{tr}(AVS^{-1}A) = \mathrm{tr}(S^{-1}AVA)$.

The exact MSE difference. Therefore, with no approximation,

$$\Delta := \mathrm{mse}_{\mathrm{ridge}} - \mathrm{mse}_{\mathrm{OLS}} = \underbrace{\lambda^2\beta' A^2\beta}_{\text{squared-bias cost}} - \underbrace{\Delta_{\mathrm{var}}}_{\text{variance gain}}.$$

这里没有丢任何 $O(\lambda^2)$ 项：Δ 是 $\mathrm{mse}_{\mathrm{ridge}} - \mathrm{mse}_{\mathrm{OLS}}$ 的精确闭式，偏差代价与方差收益都原样保留，下面只用单调的矩阵不等式去界两边，故结论对任意满足条件的固定 λ 成立，而非仅小 λ。

Bounding the two sides by eigenvalues. For the variance gain, $V = X'DX \succeq \lambda_{\min}(D)\,X'X = \lambda_{\min}(D)\,S$ (since $D \succeq \lambda_{\min}(D)I$ and congruence preserves $\succeq$), and $A^2 \succeq 0$, so $AVA \succeq \lambda_{\min}(D)\,ASA$. Because $S^{-1} \succ 0$,

$$\Delta_{\mathrm{var}} \;\geq\; 2\lambda\,\mathrm{tr}\big(S^{-1}AVA\big) \;\geq\; 2\lambda\,\lambda_{\min}(D)\,\mathrm{tr}\big(S^{-1}ASA\big) = 2\lambda\,\lambda_{\min}(D)\,\mathrm{tr}\big(A^2\big),$$

the last equality by cyclicity $\mathrm{tr}(S^{-1}ASA) = \mathrm{tr}(AS^{-1}SA) = \mathrm{tr}(A^2)$ (all factors symmetric and A, S commute, being functions of the same S). For the squared-bias cost,

$$\lambda^2\beta' A^2\beta \;\leq\; \lambda^2\,\lambda_{\max}(A^2)\,\beta'\beta \;\leq\; \lambda^2\,\mathrm{tr}(A^2)\,\beta'\beta.$$

Hence

$$\Delta \;\leq\; \lambda^2\,\mathrm{tr}(A^2)\,\beta'\beta - 2\lambda\,\lambda_{\min}(D)\,\mathrm{tr}(A^2) = \lambda\,\mathrm{tr}(A^2)\,\big[\lambda\beta'\beta - 2\lambda_{\min}(D)\big],$$

and since $\lambda > 0$ and $\mathrm{tr}(A^2) > 0$, this is < 0 exactly when $\lambda\beta'\beta < 2\lambda_{\min}(D)$:

$$\boxed{2\lambda_{\min}(D) > \lambda\,\beta'\beta \implies \mathrm{mse}(\hat\beta_{\mathrm{ridge}} \mid X) < \mathrm{mse}(\hat\beta_{\mathrm{OLS}} \mid X) \quad \text{(for every fixed } \lambda > 0).}$$

要点：方差收益 $\propto \lambda_{\min}(D)$（噪声越大，收缩越值），代价 $\propto \lambda\beta'\beta$（真信号越大、罚越重，偏差代价越大）。证明全程是精确不等式（Theobald 恒等式 + Loewner 序 + 迹的循环性），对任意固定 λ（含大 λ）都给出严格 $\Delta < 0$，不再是小 λ 的领头项论证。这正是 Hoerl–Kennard 岭存在性定理的一个充分版本。

7.8 Q4(c) · LASSO 的 oracle 性质及其假设 / The Oracle Property of LASSO and Its Assumptions

归属：ECON 510（Guggenberger） **重要度**：高 **再考判断**：高
核心考点：LASSO oracle 性质及强假设批判
历史复现：2023、2025
掌握方式：背 oracle 两部分、selection consistency 条件及非一致性批判

原题 Q4(c)

Discuss what is meant by the “oracle property” of the Lasso estimator. Critically discuss the underlying assumptions needed to obtain that result.

Solution.

中文思路（15）. “Oracle 性质”是 Fan–Li 提出的两条合一：(1) **选择相合**——以概率趋 1 选对真支撑（把零系数估成零、非零系数判为非零）；(2) **渐近有效**——非零系数的估计渐近分布与“事先知道真模型、只对那几个非零变量做 OLS ／ MLE 的 oracle”完全一样（同样 $\sqrt{n}$ 速率、同样有效协方差）。合起来：估计量表现得像“神谕预先告诉了你哪些变量重要”。

Statement. An estimator $\hat{\beta}$ of sparse β_0 (support $\mathcal{S} = \{j : \beta_{0,j} \neq 0\}$) has the oracle property if, with probability $\to 1$,

$$\text{(i)}\quad \operatorname{supp}(\hat{\beta}) = \mathcal{S} \quad \text{(selection consistency)}, \qquad \text{(ii)}\quad \sqrt{n}(\hat{\beta}_{\mathcal{S}} - \beta_{0,\mathcal{S}}) \xrightarrow{d} N(0, \Sigma_{\mathcal{S}}^{\text{oracle}}),$$

the same limit as the OLS that knew $\mathcal{S}$ in advance.

Critical discussion of the assumptions.

1. **稀疏性**（sparsity）：真参数只有少数非零（$|\mathcal{S}|$ 小），否则“选对支撑”无从谈起。
2. β-min / 信号强度: 最小非零系数 $\min_{j\in\mathcal{S}} |\beta_{0,j}|$ 不能太小（须 $\gg$ 噪声尺度），否则弱信号被罚没、漏选。
3. **不可表示／不相干条件**（irrepresentable/incoherence）：相关变量对无关变量的“可表示性”受控，
   否则相关的零变量被误选——这是普通 LASSO 选择相合的几乎必要条件。
4. **限制特征值／可识别**（restricted eigenvalue）：设计阵在稀疏方向上不退化，保证非零块可估。
5. **调参**λ_n: selection 通常要求惩罚足够大（在本归一化下常见$\sqrt{n}\lambda_n \to \infty$），
   而 oracle 正态性要求活跃坐标上的有效惩罚为$o(n^{-1/2})$；plain LASSO 的两要求冲突。

Critical caveat. 普通 LASSO 一般**不**同时拥有 oracle 性质：要么 λ_n 大到选对支撑、却给非零系数引入不可忽略的**收缩偏差**（破坏 (ii) 的有效性）；要么 λ_n 小到活跃坐标近似无偏、却压不掉零坐标。因此不能把"$\lambda_n \to 0$ 且 $\sqrt{n}\lambda_n \to \infty$"误说成同时兼顾两端——它服务于选择，却与 plain-LASSO 的 oracle 正态性冲突。不相干条件 (3) 在实际相关设计下也常被违反。**所以严格的 oracle 性质通常要靠 Zou(2006) 的** *adaptive LASSO*（按初始一致估计自适应加权惩罚，使活跃坐标的有效权重趋小、零坐标的权重趋大），或 SCAD ／ MCP 等非凸惩罚——它们放松这组冲突并消去对非零系数的一阶偏差。

Oracle = 选择相合 + 非零系数渐近有效；标准 LASSO 一般达不到，需 adaptive LASSO / SCAD，并依赖 sparsity、β-min、不相干、调参条件。

考试要点：能说清"oracle= 两条"、并**批判性**指出标准 LASSO 的偏差—选择两难、不相干条件易破，从而需要自适应版本——这正是题目"critically discuss"的落点。

Part B ·Pinkse（501 基础，每题十分钟内）

Pinkse 明确说每题十分钟内、答案几行即可——写长就是浪费时间。以下解答按此节奏，点到关键步骤与最终答案。

7.9 Pinkse Q1 ·两集合的容斥 / Two-set Inclusion-Exclusion

归属：ECON 501（Pinkse） **重要度**：低 **再考判断**：低
核心考点：概率测度的二集合容斥
历史复现：基础恒等式；未作为主问题重复
掌握方式：现场由不交并可列可加推出

原题 P-Q1

Using the definition of a probability measure, show that $\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B)$.

Solution.

中文思路. 把 $A \cup B$ 拆成三块互不相交，用有限可加性（概率测度公理）凑等式 (1)。

Formal. Decompose into disjoint pieces:

$$A = (A \setminus B) \sqcup (A \cap B), \quad B = (B \setminus A) \sqcup (A \cap B), \quad A \cup B = (A \setminus B) \sqcup (A \cap B) \sqcup (B \setminus A).$$

Finite additivity gives $\mathrm{P}(A) = \mathrm{P}(A \setminus B) + \mathrm{P}(A \cap B)$, $\mathrm{P}(B) = \mathrm{P}(B \setminus A) + \mathrm{P}(A \cap B)$, and $\mathrm{P}(A \cup B) = \mathrm{P}(A \setminus B) + \mathrm{P}(A \cap B) + \mathrm{P}(B \setminus A)$. Adding the first two and subtracting the third:

$$\begin{aligned} \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B) = &\big[\mathrm{P}(A \setminus B) + \mathrm{P}(B \setminus A) + 2\,\mathrm{P}(A \cap B)\big] \\ &- \big[\mathrm{P}(A \setminus B) + \mathrm{P}(B \setminus A) + \mathrm{P}(A \cap B)\big] = \mathrm{P}(A \cap B). \end{aligned}$$

$$\boxed{\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B).}$$

7.10 Pinkse Q2 ·不相关／条件均值独立／独立（正态的平方）/ Uncorrelated vs. CMI vs. Independent

归属：ECON 501（Pinkse） **重要度**：高 **再考判断**：高
核心考点：不相关、条件均值独立与独立
历史复现：2020、2021、2023、2024
掌握方式：掌握蕴含关系与 $Y = X^2$ 反例

原题 P-Q2

Let $y = x^2$, where $x \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. For what values of μ, σ^2 are: (a) x, y uncorrelated; (b) x conditionally mean independent of y; (c) y conditionally mean independent of x; (d) x, y independent?

Solution.

中文思路. 四个概念有严格的强弱方向：独立 $\Rightarrow$ 双向条件均值独立 $\Rightarrow$ 不相关；用集合语言是 {独立分布} $\subset$ {双向 CMI} $\subset$ {不相关}。逐个用正态矩与对称性判定（3, 2）。

(a) Uncorrelated. $\mathrm{Cov}(x, y) = \mathbb{E}(x^3) - \mathbb{E}(x)\,\mathbb{E}(x^2)$. For $x \sim N(\mu, \sigma^2)$, $\mathbb{E}(x^3) = \mu^3 + 3\mu\sigma^2$, $\mathbb{E}(x)\,\mathbb{E}(x^2) = \mu(\mu^2 + \sigma^2)$, so $\mathrm{Cov}(x, y) = 2\mu\sigma^2$. Since $\sigma > 0$,

$$\boxed{x, y \text{ uncorrelated} \iff \mu = 0 \text{ (any } \sigma^2 > 0).}$$

(b) x CMI of y ($\mathbb{E}(x \mid y)$ constant). Given $y = x^2 = c$, $x \in \{+\sqrt{c}, -\sqrt{c}\}$ with conditional probabilities $\propto \phi((\pm\sqrt{c} - \mu)/\sigma)$. For $\mu = 0$ these are equal by symmetry, so $\mathbb{E}(x \mid y) = 0$ (constant) —mean-independent. For $\mu \neq 0$ the sign is tilted and $\mathbb{E}(x \mid y)$ varies with y. Thus

$$\boxed{x \text{ CMI of } y \iff \mu = 0.}$$

(c) y CMI of x ($\mathbb{E}(y \mid x)$ constant in x). But $\mathbb{E}(y \mid x) = \mathbb{E}(x^2 \mid x) = x^2$, a nonconstant function of x for every $\sigma > 0$. Hence

$$\boxed{y \text{ is } \textit{never} \text{ CMI of } x \text{ (for any } \mu,\ \sigma > 0).}$$

(d) Independent. $y = x^2$ is a (nonconstant, since $\sigma > 0$) deterministic function of x, so x and y can never be independent:

$$\boxed{x, y \text{ are } \textit{never} \text{ independent.}}$$

阶梯一目了然：$\mu = 0$ 时只有 (a)(b) 成立；(c)(d) 对任何参数都失败。这正是“不相关 $\subsetneq$ 条件均值独立 $\subsetneq$ 独立”、且条件均值独立**不**对称的教科书反例。

7.11 Pinkse Q3 · 矩阵均值的光滑函数的极限分布（delta 法）/ Delta Method for a Smooth Functional of a Matrix Mean

归属：ECON 501（Pinkse） **重要度**：高 **再考判断**：高
核心考点：矩阵样本均值光滑函数的 delta 法
历史复现：2018–2025 渐近变换题高频
掌握方式：向量化或 directional derivative 两种写法都要会

原题 P-Q3

Let $\{X_i\}$ be matrix-valued i.i.d. variables with bounded support and mean M_0, with sample mean $\bar{X}$. Derive the limit distribution of $r_n\{\exp(u^\top \bar{X} v) - \exp(u^\top M_0 v)\}$ for given vectors $u, v \neq 0$, with r_n such that the limit exists and is non-degenerate.

Solution.

中文思路. $u^\top \bar{X} v$ 是 $\bar{X}$ 的**线性**泛函，等于标量随机变量 $\xi_i := u^\top X_i v$ 的样本均值，对它直接上 CLT；外层 $\exp(\cdot)$ 是光滑函数，用一阶 delta 法（4），导数就是 $\exp(u^\top M_0 v)$。$r_n = \sqrt{n}$。

Formal. Let $\xi_i := u^\top X_i v$ (scalar), i.i.d. with mean $u^\top M_0 v$ and finite variance $s^2 := \operatorname{Var}\big(u^\top X_1 v\big)$ (finite by bounded support). Then $u^\top \bar{X} v = \frac{1}{n}\sum_{i=1}^n \xi_i$, and CLT gives

$$\sqrt{n}\big(u^\top \bar{X} v - u^\top M_0 v\big) \xrightarrow{d} N(0, s^2).$$

Apply the delta method to $h(t) = e^t$, $h'(u^\top M_0 v) = \exp(u^\top M_0 v)$:

$$\boxed{\sqrt{n}\big\{\exp(u^\top \bar{X} v) - \exp(u^\top M_0 v)\big\} \xrightarrow{d} N\Big(0,\ \exp\big(2\,u^\top M_0 v\big)\cdot \operatorname{Var}\Big(u^\top X_1 v\Big)\Big), \qquad r_n = \sqrt{n}.}$$

非退化前提：$s^2 = \operatorname{Var}\big(u^\top X_1 v\big) > 0$（$u, v \neq 0$ 且 X_1 在那个方向上有随机性）。导数 $e^{u^\top M_0 v} > 0$ 恒不为零，无须像 Q5b 那样升阶。

7.12 Pinkse Q4 · 极值估计量的相合性 / Consistency of an Extremum Estimator

归属：ECON 501（Pinkse） **重要度**：高 **再考判断**：高
核心考点：极值估计量的 argmin consistency
历史复现：2022 Q5、2023、2024 510 Q1
掌握方式：背 separation + uniform convergence 的标准反证

原题 P-Q4

Suppose $\theta_0 = \arg\min_{\theta\in\Theta}\Omega(\theta)$ (unique), Ω continuous, Θ compact, and $\sup_{\theta\in\Theta}|\hat\Omega(\theta)-\Omega(\theta)| = o_p(1)$. Prove $\hat\theta = \arg\min_{\theta\in\Theta}\hat\Omega(\theta) \xrightarrow{p} \theta_0$.

Solution.

中文思路. 标准三件套(5):唯一最小 + 连续 + 紧 ⇒ 可识别唯一性(well-separated minimum)；再叠加一致收敛，挤出 $\hat\theta \xrightarrow{p} \theta_0$。

Proof for Theorem

Fix $\varepsilon > 0$. By compactness of Θ, continuity of Ω, and uniqueness of θ_0, the minimum is well separated:

$$\delta := \inf_{\theta\in\Theta:\,\|\theta-\theta_0\|\ge\varepsilon} \Omega(\theta) - \Omega(\theta_0) > 0$$

(the inf is attained on the compact set $\{\|\theta-\theta_0\|\ge\varepsilon\}$ and exceeds $\Omega(\theta_0)$ by uniqueness). Let $\hat\theta$ minimize $\hat\Omega$. Then

$$\Omega(\hat\theta)-\Omega(\theta_0) = \underbrace{[\Omega(\hat\theta)-\hat\Omega(\hat\theta)]}_{\le\sup|\hat\Omega-\Omega|} + \underbrace{[\hat\Omega(\hat\theta)-\hat\Omega(\theta_0)]}_{\le 0\ (\hat\theta\text{ minimizes }\hat\Omega)} + \underbrace{[\hat\Omega(\theta_0)-\Omega(\theta_0)]}_{\le\sup|\hat\Omega-\Omega|} \le 2\sup_{\theta\in\Theta}|\hat\Omega(\theta)-\Omega(\theta)|.$$

On the event $\{\sup_\theta|\hat\Omega-\Omega| < \delta/2\}$ we get $\Omega(\hat\theta)-\Omega(\theta_0) < \delta$, which by definition of δ forces $\|\hat\theta-\theta_0\| < \varepsilon$. Since $\sup_\theta|\hat\Omega-\Omega| = o_p(1)$,

$$\mathrm{P}\Big(\|\hat\theta-\theta_0\|\ge\varepsilon\Big) \le \mathrm{P}\Big(\sup_\theta|\hat\Omega-\Omega|\ge\delta/2\Big) \to 0.$$

∎

$$\boxed{\hat\theta \xrightarrow{p} \theta_0.}$$

这是 Newey–McFadden 相合性定理的最小证法：uniform convergence + identifiable uniqueness。三明治不等式的中间项 ≤ 0 是“$\hat\theta$ 最小化 $\hat\Omega$”的唯一用处。

7.13 Pinkse Q5 · 期望极限的存在（控制收敛）/ Existence of $\lim\mathbb{E}(y_n)$ via DCT

归属：ECON 501（Pinkse） **重要度**：中高 **再考判断**：中高
核心考点：DCT 下的期望极限
历史复现：2019 截断、2023、2024/2025 收敛陷阱
掌握方式：逐项验证点态极限、可测性和可积 dominator

原题 P-Q5

Let x be continuously distributed with support $[1,\infty)$ and density f. Let $y_n = (1+x/n)^{-n}$. Prove that $\lim_{n\to\infty}\mathbb{E}(y_n)$ exists and determine its limit.

Solution.

中文思路. 逐点 $(1+x/n)^{-n}\to e^{-x}$；且 $0\le y_n\le 1$ 被常数 1（关于 f 可积）控制，DCT 直接换序（4, 1）。

Formal. For each fixed $x\ge 1$, $(1+x/n)^{-n}\to e^{-x}$ as $n\to\infty$. Also $0\le y_n = (1+x/n)^{-n}\le 1$, and the constant 1 is integrable w.r.t. the probability density f. By the Dominated Convergence Theorem,

$$\boxed{\lim_{n\to\infty}\mathbb{E}(y_n) = \mathbb{E}\big(e^{-x}\big) = \int_1^\infty e^{-x} f(x)\,\mathrm{d}x.}$$

支撑 $[1,\infty)$ 让一切良定，但极限式对一般 f 就停在积分形式（除非给定 f）。控制函数取常数 1 是因为 $y_n\in[0,1]$。

7.14 Pinkse Q6 · 用 Hoeffding 不等式构造有限样本 size-α 检验 / A Finite-sample Test via Hoeffding

归属：ECON 501（Pinkse） **重要度**：中高 **再考判断**：中
核心考点：Hoeffding 构造有限样本检验
历史复现：2023；不等式在 2024/2025 仍重要
掌握方式：记代入范围与解 critical value 的代数

原题 P-Q6

Hoeffding: if $\{x_i\}$ independent with $\mathrm{P}(a_i\le x_i\le b_i)=1$, then $\forall z>0$, $\mathrm{P}(|\sum_{i=1}^n(x_i-\mathbb{E}(x_i))|>z)\le 2\exp(-2z^2/\sum_{i=1}^n(b_i-a_i)^2)$. Suppose i.i.d. $\{x_i\}$ has support $[-1/\sqrt{2},1/\sqrt{2}]$. Construct a test of $H_0:\mu_0=0$ vs. $H_1:\mu_0\ne 0$ ($\mu_0=\mathbb{E}(x_1)$) with rejection probability $\le\alpha$ under H_0, of the form "reject if $|\bar{x}|>C_\alpha$".

Solution.

中文思路. 直接把 Hoeffding 套到 $\bar{x}$ 上，令尾界等于 α 解出 C_α。支撑宽 $b_i-a_i=\sqrt{2}$，故 $\sum(b_i-a_i)^2=2n$。这是**有限样本精确**的非渐近水平检验（6）。

Formal. Under H_0, $\mu_0=0$ so $\sum_{i=1}^n(x_i-\mathbb{E}(x_i))=\sum_{i=1}^n x_i = n\bar{x}$. With $b_i-a_i=\frac{1}{\sqrt{2}}-(-\frac{1}{\sqrt{2}})=\sqrt{2}$, $\sum_{i=1}^n(b_i-a_i)^2=2n$. Hoeffding with $z=nC$:

$$\mathrm{P}_{H_0}(|\bar{x}|>C)=\mathrm{P}_{H_0}\left(\Big|\sum_{i=1}^n(x_i-\mathbb{E}(x_i))\Big|>nC\right)\le 2\exp\Big(\frac{-2(nC)^2}{2n}\Big)=2\exp(-nC^2).$$

Set $2\exp(-nC_\alpha^2) = \alpha$:

$$\boxed{\text{reject } H_0 \text{ if } |\bar{x}| > C_\alpha := \sqrt{\tfrac{1}{n}\ln(2/\alpha)}, \qquad \mathrm{P}_{H_0}(\text{reject}) \le \alpha\ \forall n.}$$

亮点：这是**非渐近**（finite-sample）水平 α 检验——不靠 CLT，靠有界支撑 + Hoeffding，对任意 n 都成立（偏保守，因为 Hoeffding 是上界）。

7.15 Pinkse Q7 · 标准正态二次型的精确分布 / Exact Law of a Quadratic Form in Normals

归属：ECON 501（Pinkse） **重要度**：中高 **再考判断**：中
核心考点：标准正态二次型的精确分布
历史复现：正态二次型/检验在 2018、2023、2025 出现
掌握方式：谱分解要会；一般是加权独立 χ_1^2 而非总是 χ^2

原题 P-Q7

Consider $x = z^\top W z$ for a positive semidefinite W, where z is a vector of i.i.d. $N(0,1)$. Derive the exact distribution of x.

Solution.

中文思路. 对称 PSD 阵谱分解 $W = P\Lambda P'$，正态在正交变换下不变 $\Rightarrow$ x 是独立 χ_1^2 的特征值加权和（2, 3）。

Formal. Diagonalize $W = P\Lambda P'$ with P orthogonal, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_j \ge 0$ (PSD). Let $w := P'z$; since P is orthogonal, $w \sim N(0, P'I_nP) = N(0, I_n)$, so $w_1, \ldots, w_n \overset{i.i.d.}{\sim} N(0,1)$. Then

$$x = z^\top W z = z^\top P\Lambda P' z = w^\top \Lambda w = \sum_{j=1}^n \lambda_j w_j^2.$$

Each $w_j^2 \sim \chi_1^2$ independently, so

$$\boxed{x \overset{d}{=} \sum_{j=1}^n \lambda_j\, \chi_{1,j}^2 \quad (\text{independent } \chi_1^2),\quad \lambda_j = \text{eigenvalues of } W.}$$

特例：W 幂等（$W^2 = W$）秩 r 时特征值是 r 个 1 与其余 0，$x \sim \chi_r^2$——这正是 OLS 残差平方和、F 检验背后的结构。一般情形是“广义卡方”(Imhof 可算 CDF)，MGF $= \prod_j (1 - 2t\lambda_j)^{-1/2}$，$t < 1/(2\max_j \lambda_j)$。

7.16 Pinkse Q8 ·重尾（Cauchy 型）分布下方差的相合渐近检验 / Consistent Test for the Scale of a Cauchy-type Law

归属： ECON 501（Pinkse） **重要度：** 中 **再考判断：** 中低
核心考点： 重尾分布尺度参数的稳健检验
历史复现： 仅 2023 该具体分布；稳健变换思想可复现
掌握方式： 识别可积变换后用 CLT；不硬算不存在的矩

原题 P-Q8

Suppose $\{x_i\}$ i.i.d. with density $(6\sigma_0\sqrt{3})\pi^{-1}\{3\sigma_0^2+(x-\mu_0)^2\}^{-1}$. Propose a consistent and asymptotically valid test for $H_0:\sigma_0^2=1$ vs. $H_1:\sigma_0^2\neq 1$. Be precise (no need to argue consistency/validity).

Solution.

中文思路（这是陷阱题）．这个密度是 **Cauchy 型**：$\dfrac{1}{\pi}\dfrac{s}{s^2+(x-\mu_0)^2}$ 配上尺度 $s=\sqrt{3}\,\sigma_0$（前面的常数因子是出题人手误，结构是 Cauchy）。Cauchy **没有有限均值与方差**——所以**绝不能**用样本方差去估 σ_0^2（样本方差不收敛！）。正确做法：用一个**稳健的分位数泛函**（IQR / MAD）来 $\sqrt{n}$-一致地估 σ_0，再做 Wald 检验（6, 4）。

Construction. The law is Cauchy with location μ_0 and scale $s=\sqrt{3}\,\sigma_0$. Its quartiles are $\mu_0\pm s$, so the population interquartile range is $\mathrm{IQR}=2s=2\sqrt{3}\,\sigma_0$, giving

$$\sigma_0=\frac{\mathrm{IQR}}{2\sqrt{3}},\qquad \sigma_0^2=\frac{\mathrm{IQR}^2}{12}.$$

Estimate σ_0 by the sample interquartile range $\widehat{\mathrm{IQR}}=\hat{q}_{.75}-\hat{q}_{.25}$ (sample quantiles), and set $\hat{\sigma}_0^2=\widehat{\mathrm{IQR}}^2/12$. Here the asymptotic variance can be derived explicitly, so no unspecified nuisance constant is needed. For sample quantiles,

$$\mathrm{ACov}(\hat{q}_p,\hat{q}_r)=\frac{\min(p,r)-pr}{f(q_p)f(q_r)}.$$

At the Cauchy quartiles $q_{.25}=\mu_0-s$ and $q_{.75}=\mu_0+s$, $f(q_{.25})=f(q_{.75})=1/(2\pi s)$. Consequently,

$$\mathrm{AVar}(\hat{q}_{.25})=\mathrm{AVar}(\hat{q}_{.75})=\frac{(1/4)(3/4)}{(1/(2\pi s))^2}=\frac{3}{4}\pi^2 s^2,$$
$$\mathrm{ACov}(\hat{q}_{.25},\hat{q}_{.75})=\frac{1/4-(1/4)(3/4)}{(1/(2\pi s))^2}=\frac{1}{4}\pi^2 s^2.$$

Therefore

$$\sqrt{n}\{\widehat{\mathrm{IQR}} - 2s\} \xrightarrow{d} N(0, \pi^2 s^2).$$

The derivative of $h(q) = q^2/12$ at $q = 2s$ is $h'(2s) = s/3$. Since $s^2 = 3\sigma_0^2$, the delta method gives

$$\sqrt{n}(\hat{\sigma}_0^2 - \sigma_0^2) \xrightarrow{d} N\left(0, \frac{s^2}{9}\pi^2 s^2\right) = N(0, \pi^2 \sigma_0^4).$$

Thus a fully explicit Wald test is

$$\boxed{W_n = \frac{n\,(\hat{\sigma}_0^2 - 1)^2}{\pi^2 \hat{\sigma}_0^4} \overset{H_0}{\underset{}{\xrightarrow{d}}} \chi_1^2, \qquad \text{reject } H_0 : \sigma_0^2 = 1 \text{ if } W_n > \chi^2_{1,1-\alpha}, \quad \hat{\sigma}_0^2 = \frac{\widehat{\mathrm{IQR}}^2}{12}.}$$

Under H_0 one may equivalently use the null variance π^2 in the denominator, because $\sigma_0^4 = 1$. 先把出题人手误点破（也是评分点）：照抄题面，$\int_{-\infty}^{\infty}(6\sigma_0\sqrt{3})\pi^{-1}\{3\sigma_0^2 + (x - \mu_0)^2\}^{-1}\mathrm{d}x = 6 \neq 1$，所以打印出来的根本**不是合法密度**（前面的常数因子多了 6 倍）。把常数改成 $\sqrt{3}\,\sigma_0/\pi$ 才积分为 1，结构就是尺度 $s = \sqrt{3}\,\sigma_0$ 的 Cauchy——考试时应当明写“题面密度不归一、按 Cauchy 结构作答”，阅卷会给分。 核心要点（评分点）：识别出 Cauchy/重尾 $\Rightarrow$ 方差不存在 $\Rightarrow$ 不能用经典 s^2 的方差检验，必须用 IQR/MAD 这类**基于分位数**的稳健尺度估计搭 Wald。等价地可用 MLE：Cauchy 位置-尺度 MLE 渐近正态有效，对 σ_0^2 做 Wald/LR 亦可（5）。

Chapter 8　2024 资格考精解（Worked Solutions）

导读 · Roadmap（先读这里，不含公式）

2024 年 8 月资格考分两卷。**第一卷 ECON 501**（主考 Sung Jae Jun，六题）全是测度论与概率基础：Q1 推前测度（pushforward）的多种刻画与换元密度、Q2 联合密度上的非线性区域概率/方差/嵌套条件期望/残差正交、Q3 中位数极小化 $\mathbb{E}(|X-y|)$ 与「均值–中位数」距离上界、Q4 用联合 MGF 定双正态再做条件密度与单边截断条件均值（逆 Mills 比）、Q5 振荡密度下独立事件概率之和的近似（带证明）、Q6 四道真伪概念题。**第二卷 Comprehensive (510)**（主考 Patrik Guggenberger，五题）：Q1 OLS 渐近 + 异方差稳健夹心方差及一致性证明、Q2 非光滑 GMM 准则下 V_0, Γ 的估计并落到分位回归、Q3 Heckit 三问（逆 Mills 选择偏误项、probit 识别 π、排除限制识别 θ 与 $\sigma_{\varepsilon\eta}$）、Q4 矩不等式检验的漂移序列渐近（h_1 置信域、h_2 一致估计、数据依赖临界值的 asymptotic size、Bonferroni size 修正 $\delta=\alpha-\beta$ 与功效权衡）、Q5 非参 bootstrap 与子抽样（$s=\sqrt{n}$）对均值检验的 NRP 与一致性比较。

每题先给原题（逐字），再给中文思路与 *English formal derivation + boxed answer*。所有数值（$\mathrm{Var}(X)=1/20$、Z 的两个原子 $21/32, 27/32$、$\mathbb{E}(Y \mid 2X+Y<0)=-\sqrt{2/(5\pi)}$、Q5 之和 ≈ 2）我都独立重推并数值核验过。需要的定理一律指回前两部分章节，本章只指向并套用。

8.1　Q1（501）推前测度与换元密度

归属：ECON 501（Sung Jae Jun）　**重要度**：高（必会）　**再考判断**：高
核心考点：推前测度等价刻画与换元密度
历史复现：2017、2018、2019、2024
掌握方式：四步换元机器和密度 Jacobian 都要会

原题 Q1 (verbatim)

Consider the probability space $\big((0,1), \mathcal{B}(0,1), \lambda\big)$, where $\mathcal{B}(0,1)$ is the Borel σ algebra over the interval $(0,1)$, and λ is the Lebesgue measure. Let $X : (0,1) \to \mathbb{R}$ be a random variable defined by $X(\omega) = \exp(-\omega)$.

(a) Provide at least two ways of completely characterizing the probability law $P_X := \lambda \circ X^{-1}$ on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ algebra over $\mathbb{R}$.

(b) Find a nonnegative function f on $\mathbb{R}$ such that $\int_{(0,1)} g\big(X(\omega)\big)\,\lambda(d\omega) = \int_{\mathbb{R}} g(x) f(x)\,\mu(dx)$ for any nonnegative function g, where μ is the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$.

Solution.

思路。$X(\omega) = e^{-\omega}$ 在 $(0,1)$ 上严格递减，把 $\omega \in (0,1)$ 一一映到 $x \in (1/e, 1)$。推前测度 $P_X = \lambda \circ X^{-1}$ 就是「Uniform(0,1) 取 $e^{-\cdot}$」的分布。完整刻画一个 $\mathbb{R}$ 上的概率律有几条等价路：CDF、密度、MGF/特征函数——给两条即可（1 的推前测度 + 2 的分布刻画）。(b) 要找的 f 恰是 P_X 关于 Lebesgue 的 Radon–Nikodym 密度，单调换元 $x = e^{-\omega} \Rightarrow \omega = -\ln x$，$d\omega = -dx/x$ 立得。

(a) Two characterizations of $P_X = \lambda \circ X^{-1}$. Since $\omega \mapsto e^{-\omega}$ is a strictly decreasing bijection of $(0,1)$ onto $(1/e, 1)$, P_X is supported on $(1/e, 1)$.

(i) CDF. For $x \in \mathbb{R}$, using $X(\omega) \le x \iff e^{-\omega} \le x \iff \omega \ge -\ln x$,

$$F_X(x) = \mathrm{P}(X \le x) = \lambda\big(\{\omega \in (0,1) : \ \omega \ge -\ln x\}\big) = \begin{cases} 0, & x \le 1/e, \\ 1 + \ln x, & 1/e < x < 1, \\ 1, & x \ge 1. \end{cases}$$

(For $1/e < x < 1$, $-\ln x \in (0,1)$ and $\lambda([-\ln x, 1)) = 1 - (-\ln x) = 1 + \ln x$.)

(ii) Density w.r.t. Lebesgue. Differentiating F_X on $(1/e, 1)$,

$$f_X(x) = F_X'(x) = \frac{1}{x}\,\mathbb{1}\big(\tfrac{1}{e} < x < 1\big).$$

(iii) MGF (a third, optional). $M_X(t) = \mathbb{E}\big(e^{tX}\big) = \int_0^1 e^{te^{-\omega}} \mathrm{d}\omega = \int_{1/e}^1 e^{tx} \frac{dx}{x}$, finite for all t; this too determines P_X. Any one of (i)/(ii)/(iii) plus another constitutes "two ways."

$$\boxed{\begin{aligned} &F_X(x) = (1 + \ln x)\,\mathbb{1}(1/e < x < 1) + \mathbb{1}(x \ge 1), \\ &f_X(x) = \tfrac{1}{x}\,\mathbb{1}(1/e < x < 1). \end{aligned}}$$

(b) The density f via change of variables / unconscious statistician. By the change-of-variables (image-measure) theorem (1), for any nonnegative measurable g, $\int_{(0,1)} g\big(X(\omega)\big)\,\lambda(d\omega) = \int_{\mathbb{R}} g(x)\,P_X(dx)$. Compute the left side directly with $x = e^{-\omega}$, $\omega = -\ln x$, $d\omega = -\,dx/x$; as $\omega : 0 \to 1$, $x : 1 \to 1/e$, so

$$\int_0^1 g(e^{-\omega})\,d\omega = \int_{x=1}^{x=1/e} g(x)\left(-\frac{dx}{x}\right) = \int_{1/e}^1 g(x)\,\frac{1}{x}\,dx = \int_{\mathbb{R}} g(x)\,\underbrace{\frac{1}{x}\,\mathbb{1}(1/e < x < 1)}_{=:f(x)}\,\mu(dx).$$

Hence

$$\boxed{f(x) = \frac{1}{x}\,\mathbb{1}\big(\tfrac{1}{e} < x < 1\big)}$$

which is exactly $dP_X/d\mu = f_X$ from (a). □

Remark (Q1 要点).

f 必须只在像集 $(1/e, 1)$ 上非零——这是「写对支撑」最容易丢分的地方。$\int_{1/e}^1 x^{-1}dx = \ln 1 - \ln(1/e) = 1$，密度积分为 1，自洽。

8.2 Q2（501）联合密度：区域概率、方差、嵌套条件期望与残差正交

归属：ECON 501（Sung Jae Jun） **重要度**：高 **再考判断**：高
核心考点：联合密度、区域积分、嵌套条件期望与正交
历史复现：2019、2020、2021、2024
掌握方式：先画支持集；tower 与 projection 用 σ 域包含关系

原题 Q2 (verbatim)

Suppose that the joint probability density function of (X, Y) is given by $f(x, y) = c \cdot x \cdot \mathbb{1}(0 < x < y < 1)$ for some constant $c > 0$.

(a) Compute $\mathrm{P}\big(Y < X^2 + 0.5\big)$. (b) Compute $\mathrm{Var}(X)$.

(c) Characterize the probability distribution of $Z := \mathbb{E}\big(\mathbb{E}\big(\mathbb{E}(Y \mid \mathbb{1}(X \le 0.5)) \mid \mathbb{1}(X \le 0.5), X^2\big) \mid X\big)$.

(d) What is the correlation coefficient between $Y - Z$ and $\mathbb{1}(X > 0.5)$, where Z is defined in part (c)?

Solution.

先定常数。在三角形 $0 < x < y < 1$ 上 $\int_0^1\int_0^y cx\,dx\,dy = c\int_0^1 \frac{y^2}{2}dy = \frac{c}{6} = 1$，故 $c = 6$。下面 (a)(b) 直接积分；(c)(d) 是 3 的 tower 性质与条件期望正交性的概念题，硬算反而绕远。

(a) The region is $\{0 < x < y < 1,\ y < x^2 + 0.5\}$. For each $x \in (0, 1)$ the bounds on y are x and $\min(1, x^2 + 0.5)$ (note $x^2 + 0.5 > x$ always, since $x^2 - x + 0.5$ has negative discriminant); and $x^2 + 0.5 < 1 \iff x < 1/\sqrt{2}$. With

$\int_x^U 6x\,dy = 6x(U-x)$,

$$\mathrm{P}\big(Y < X^2 + \tfrac{1}{2}\big) = \int_0^{1/\sqrt{2}} 6x\left(x^2 + \tfrac{1}{2} - x\right)dx + \int_{1/\sqrt{2}}^{1} 6x(1-x)\,dx.$$

Evaluating, the first integral is $\left[\frac{6}{4}x^4 + \frac{3}{2}x^2 - 2x^3\right]_0^{1/\sqrt{2}} = \frac{3}{8} + \frac{3}{4} - \frac{1}{\sqrt{2}} = \frac{9}{8} - \frac{1}{\sqrt{2}}$, and the second is $\left[3x^2 - 2x^3\right]_{1/\sqrt{2}}^{1} = 1 - \left(\frac{3}{2} - \frac{1}{\sqrt{2}}\right) = -\frac{1}{2} + \frac{1}{\sqrt{2}}$. Summing, the $1/\sqrt{2}$ terms cancel:

$$\boxed{\mathrm{P}\big(Y < X^2 + 0.5\big) = \frac{9}{8} - \frac{1}{2} = \frac{5}{8}} \qquad (= 0.625).$$

(b) Marginal of X: $f_X(x) = \int_x^1 6x\,dy = 6x(1-x)$ on $(0,1)$ (a Beta$(2,2)$ density). Hence $\mathbb{E}(X) = \int_0^1 x\cdot 6x(1-x)\,dx = \frac{1}{2}$ and $\mathbb{E}\big(X^2\big) = \int_0^1 x^2\cdot 6x(1-x)\,dx = \frac{3}{10}$, so

$$\boxed{\mathrm{Var}(X) = \tfrac{3}{10} - \tfrac{1}{4} = \tfrac{1}{20} = 0.05\,.}$$

(c) 思路（tower 坍缩）。最内层 $W := \mathbb{E}(Y\,|\,\mathbb{1}(X \le 0.5))$ 只是 $\mathbb{1}(X \le 0.5)$ 的函数（两值随机变量）。再对 $(\mathbb{1}(X \le 0.5), X^2)$ 取条件期望：W 已是该 σ 域可测，内层那次什么都不改，仍是 W。最外层对 X：W 是 $\mathbb{1}(X \le 0.5)$ 的函数、从而 $\sigma(X)$ 可测，再取一次还是 W。所以 $Z = W = \mathbb{E}(Y\,|\,\mathbb{1}(X \le 0.5))$。

By the tower property (3): $W := \mathbb{E}(Y\,|\,\mathbb{1}(X \le 0.5))$ is $\sigma(\mathbb{1}(X \le 0.5))$-measurable; since $\sigma(\mathbb{1}(X \le 0.5)) \subseteq \sigma(\mathbb{1}(X \le 0.5), X^2) \subseteq \sigma(X)$, both outer conditional expectations leave W unchanged. Thus $Z = W$, a *two-valued* random variable. Compute the two atoms. With $\mathbb{E}(Y\,|\,X = x) = \int_x^1 y\frac{6x}{6x(1-x)}dy = \frac{1+x}{2}$ and $\mathrm{P}(X \le 0.5) = \int_0^{1/2} 6x(1-x)dx = \frac{1}{2}$,

$$a_1 := \mathbb{E}(Y\,|\,X \le 0.5) = \frac{\int_0^{1/2}\int_x^1 6xy\,dy\,dx}{\mathrm{P}(X \le 0.5)} = \frac{21/64}{1/2} = \frac{21}{32}, \qquad a_2 := \mathbb{E}(Y\,|\,X > 0.5) = \frac{27/64}{1/2} = \frac{27}{32}.$$

Therefore Z takes value a_1 on $\{X \le 0.5\}$ and a_2 on $\{X > 0.5\}$:

$$\boxed{Z = \begin{cases} \dfrac{21}{32}, & \text{w.p. } \mathrm{P}(X \le 0.5) = \frac{1}{2}, \\[4pt] \dfrac{27}{32}, & \text{w.p. } \mathrm{P}(X > 0.5) = \frac{1}{2}. \end{cases}}$$

校验 $\mathbb{E}(Z) = \frac{1}{2}\left(\frac{21}{32} + \frac{27}{32}\right) = \frac{48}{64} = \frac{3}{4} = \mathbb{E}(Y)$，与 tower 性质一致（$\mathbb{E}(Y) = \mathbb{E}\left(\frac{1+X}{2}\right) = \frac{1+1/2}{2} = \frac{3}{4}$）。

(d) $Z = \mathbb{E}(Y\,|\,\mathbb{1}(X \le 0.5))$ 是 Y 在 $\sigma(\mathbb{1}(X \le 0.5))$ 上的 L^2 投影，故残差 $Y - Z$ 与该 σ 域里任何变量正交。By the orthogonality of conditional expectation,

$Y - Z = Y - \mathbb{E}(Y \mid \mathbb{1}(X \le 0.5))$ satisfies $\mathbb{E}(Y - Z \mid \mathbb{1}(X \le 0.5)) = 0$, hence

$$\mathbb{E}(Y - Z) = 0, \qquad \mathbb{E}((Y - Z)\,\mathbb{1}(X > 0.5)) = \mathbb{E}(\mathbb{1}(X > 0.5)\,\mathbb{E}(Y - Z \mid \mathbb{1}(X \le 0.5))) = 0,$$

where the second equality uses that $\mathbb{1}(X > 0.5) = 1 - \mathbb{1}(X \le 0.5)$ is $\sigma(\mathbb{1}(X \le 0.5))$-measurable and $\mathbb{E}(Y - Z \mid \mathbb{1}(X \le 0.5)) = 0$. Thus $\mathrm{Cov}(Y - Z, \mathbb{1}(X > 0.5)) = 0$ and

$$\boxed{\mathrm{Corr}\left(Y - Z,\ \mathbb{1}(X > 0.5)\right) = 0\,.}$$

Remark (Q2 易错点).

(c) 别被三层嵌套唬住——核心只是「内层结果对外层 σ 域可测则坍缩」。(d) 不必算任何积分：投影残差与被投影的 σ 域正交是定义级事实（3）。

8.3 Q3（501）中位数极小化 $\mathbb{E}(|X - y|)$ 与均值–中位数距离上界

归属：ECON 501（Sung Jae Jun） **重要度**：中高 **再考判断**：中
核心考点：中位数最小化绝对损失与均值–中位数界
历史复现：2024；凸损失与 QR 连接当前 510
掌握方式：理解左右导数/次梯度；界用 Jensen/三角不等式

原题 Q3 (verbatim)

Suppose that X has a probability density function f on $\mathbb{R}$ such that $f(x) > 0$ for all $x \in \mathbb{R}$, and $\int_{-\infty}^{\infty} x^2 f(x)\,dx < \infty$. For $y \in \mathbb{R}$, define $g(y) := \int_{-\infty}^{\infty} |x - y|\,f(x)\,dx$.

(a) Derive a characterization of the value y^* that minimizes the function g by using the probability density function f.

(b) Let $x^* := \int_{-\infty}^{\infty} x f(x)\,dx$. What is the largest possible distance between x^* and y^*?

Solution.

思路。(a) $\mathbb{E}(|X - y|)$ 的极小点是中位数——把 $|x - y|$ 对 y 求导，$\partial_y |x - y| = \mathrm{sgn}(y - x)$（$x \ne y$ 处），积分得 $g'(y) = 2F(y) - 1$，令零得 $F(y^*) = 1/2$。$f > 0$ 处处保证 F 严格增、y^* 唯一。(b) 经典「均值–中位数」不等式：$|x^* - y^*| \le \sigma$，由 g 在 y^* 取最小 + Jensen 给出。

(a) FOC $\Rightarrow$ median. Write $g(y) = \int_{-\infty}^{y}(y - x)f(x)dx + \int_{y}^{\infty}(x - y)f(x)dx$. The integrand $|x - y|$ is Lipschitz in y (slope ± 1) and $\int(1 + |x|)f < \infty$, so dominated convergence justifies differentiating under the integral (2); the boundary terms vanish by continuity at $x = y$. Hence

$$g'(y) = \int_{-\infty}^{y} f(x)\,dx - \int_{y}^{\infty} f(x)\,dx = F(y) - \big(1 - F(y)\big) = 2F(y) - 1.$$

g is convex ($g''(y) = 2f(y) > 0$), so the unique minimizer solves $g'(y^*) = 0$:

$$\boxed{F(y^*) = \tfrac{1}{2}, \quad \text{i.e.} \int_{-\infty}^{y^*} f(x)\,dx = \tfrac{1}{2} \quad (y^* = \text{the median}).}$$

Since $f > 0$ everywhere, F is strictly increasing and y^* is unique.

(b) $|x^* - y^*| \le \sigma$**.** 两步：先用 y^* 是 $\mathbb{E}(|X-y|)$ 的极小点，把 $\mathbb{E}(|X-y^*|)$ 压到 $\mathbb{E}(|X-x^*|)$ 之下；再用 Jensen/Cauchy–Schwarz 把 $\mathbb{E}(|X-x^*|)$ 顶到 σ。 Let $\sigma^2 = \mathrm{Var}(X) = \mathbb{E}\big((X-x^*)^2\big) < \infty$. First, by Jensen's inequality applied to the convex map $t \mapsto |t|$,

$$|x^* - y^*| = |\mathbb{E}(X - y^*)| \le \mathbb{E}(|X - y^*|)\,.$$

Since y^* minimizes $g(y) = \mathbb{E}(|X-y|)$, $\mathbb{E}(|X-y^*|) \le \mathbb{E}(|X-x^*|)$. Finally, by Cauchy–Schwarz (Jensen on $t \mapsto t^2$), $\mathbb{E}(|X-x^*|) \le \Big(\mathbb{E}\big((X-x^*)^2\big)\Big)^{1/2} = \sigma$. Chaining,

$$\boxed{|x^* - y^*| \le \sigma = \sqrt{\mathrm{Var}(X)}\,.}$$

The coefficient 1 is sharp (it can be approached by increasingly skewed distributions, and the strict-positive-density requirement can be met by an arbitrarily small smooth perturbation). If "largest possible distance" is read as asking for one numerical bound uniform over *all* admissible densities, then there is no finite bound: rescaling X by a constant rescales both $|x^* - y^*|$ and σ without violating the assumptions. Thus the meaningful sharp answer is the scale-normalized inequality above; across the unrestricted class of f, the absolute supremum is $+\infty$.

Remark (Q3 注).

(a) 答案要写成「$F(y^*) = 1/2$」即中位数的刻画，这正是题目要的「用 f 来刻画 y^*」。(b) 标准且 sharp 的尺度化结论是 $|\text{mean} - \text{median}| \le \sigma$；评分关心你能否把三步（Jensen $\to$ y^* 最优 $\to$ Cauchy–Schwarz）串起来。若问的是不固定尺度时的绝对数值 supremum，则因可任意缩放而是 $+\infty$；把这句补上可消除题意歧义。

8.4　Q4（501）联合 MGF 定双正态 + 单边截断条件均值

归属：ECON 501（Sung Jae Jun）　**重要度**：高（必会）　**再考判断**：高
核心考点：联合 MGF、二元正态与截断均值
历史复现：2020、2021、2024、2025

掌握方式：MGF 识别、条件正态、逆 Mills 三步模板

原题 Q4 (verbatim)

Suppose that X and Y are random variables with $\mathbb{E}(\exp(tX+sY)) = \exp\{(t^2+s^2)/2\}$ for all $t, s \in \mathbb{R}$.

(a) Obtain the conditional density function of Y given $2X+Y$.

(b) Obtain $\mathbb{E}(Y \mid 2X+Y<0)$.

Solution.

思路。联合 MGF $\exp\{(t^2+s^2)/2\} = e^{t^2/2}e^{s^2/2}$ 可分解，故 $X \perp Y$ 且各自 $N(0,1)$（MGF 唯一决定分布，2/3）。(a) 令 $W := 2X+Y$，(Y,W) 联合正态：$\mathrm{Var}(W) = 4+1=5$，$\mathrm{Cov}(Y,W) = \mathrm{Var}(Y) = 1$。条件正态 $Y \mid W = w \sim N\big(\frac{1}{5}w,\, 4/5\big)$。(b) $\mathbb{E}(Y \mid W<0) = \mathbb{E}(\mathbb{E}(Y \mid W) \mid W<0) = \frac{1}{5}\mathbb{E}(W \mid W<0)$，再用截断正态均值/逆 Mills 比。

(a) Since the joint MGF factors as $e^{t^2/2}\,e^{s^2/2}$, X and Y are independent $N(0,1)$. Let $W = 2X+Y$. Then (Y,W) is bivariate normal (linear in the Gaussian (X,Y)) with

$$\mathbb{E}(Y) = \mathbb{E}(W) = 0, \quad \mathrm{Var}(Y) = 1, \quad \mathrm{Var}(W) = 2^2\cdot 1 + 1^2 \cdot 1 = 5,$$
$$\mathrm{Cov}(Y,W) = \mathrm{Cov}(Y, 2X+Y) = \mathrm{Var}(Y) = 1.$$

By the Gaussian conditioning formula (3), $\mathbb{E}(Y \mid W=w) = \mathbb{E}(Y) + \frac{\mathrm{Cov}(Y,W)}{\mathrm{Var}(W)}(w - \mathbb{E}(W))$ and $\mathrm{Var}(Y \mid W) = \mathrm{Var}(Y) - \frac{\mathrm{Cov}(Y,W)^2}{\mathrm{Var}(W)}$, i.e.

$$\mathbb{E}(Y \mid W=w) = \tfrac{1}{5}w, \qquad \mathrm{Var}(Y \mid W=w) = 1 - \tfrac{1}{5} = \tfrac{4}{5}.$$

$$\boxed{\begin{aligned} &f_{Y|2X+Y}(y \mid w) = \frac{1}{\sqrt{2\pi\cdot\frac{4}{5}}}\exp\Big(-\frac{(y-\frac{1}{5}w)^2}{2\cdot\frac{4}{5}}\Big), \\ &\text{i.e. } Y \mid 2X+Y = w \sim N\big(\tfrac{w}{5},\, \tfrac{4}{5}\big).\end{aligned}}$$

(b) By the tower property and part (a), $\mathbb{E}(Y \mid W<0) = \mathbb{E}(\mathbb{E}(Y \mid W) \mid W<0) = \frac{1}{5}\,\mathbb{E}(W \mid W<0)$. For $W \sim N(0,5)$, write $W = \sqrt{5}\,Z$, $Z \sim N(0,1)$; the truncated mean of a standard normal below 0 is $\mathbb{E}(Z \mid Z<0) = -\phi(0)/\Phi(0) = -\frac{1/\sqrt{2\pi}}{1/2} = -\sqrt{2/\pi}$ (negative inverse Mills ratio at the boundary; truncated-normal conditioning, 3; inverse Mills/Heckit, 7). Hence $\mathbb{E}(W \mid W<0) = \sqrt{5}\,\mathbb{E}(Z \mid Z<0) = -\sqrt{5}\sqrt{2/\pi} = -\sqrt{10/\pi}$, and

$$\mathbb{E}(Y \mid 2X+Y<0) = \tfrac{1}{5}\big(-\sqrt{10/\pi}\big) = -\frac{\sqrt{10/\pi}}{5} = -\sqrt{\frac{10}{25\pi}}.$$

$$\boxed{\mathbb{E}(Y \mid 2X+Y<0) = -\sqrt{\frac{2}{5\pi}} \approx -0.3568\,.}$$

Remark (Q4 要点).

(a) 关键是认出 $X, Y \overset{i.i.d.}{\sim} N(0,1)$（MGF 可分解 + 各自 $e^{t^2/2}$）。(b) 不要直接对 Y 做截断——Y 与事件 $\{W<0\}$ 不独立；正确路径是先投到 W（$\mathbb{E}(Y \mid W) = W/5$）再对 W 截断，把二维截断化成一维逆 Mills 比。$\phi(0) = 1/\sqrt{2\pi}$，$\Phi(0) = 1/2$。

8.5 Q5（501）振荡密度下独立事件概率之和的近似

归属：ECON 501（Sung Jae Jun） **重要度**：中 **再考判断**：中
核心考点：振荡密度、独立事件和概率和的近似
历史复现：2024 直接出现；不等式/近似为常见能力题
掌握方式：拆主项与误差项并给统一界，不背答案

原题 Q5 (verbatim)

Suppose that $X_1, X_2, \ldots$ are independent random variables, and that the probability density function of X_n is given by $f_n(x) = \{1+\cos(2\pi n x)\}\,\mathbb{1}(0 \le x \le 1)$. Compute

$$\sum_{n=501}^{600} \mathrm{P}(0.5 \le X_n < 0.7 \text{ and } 0.2 < X_{n+3} \le 0.3)\,.$$

You may provide an approximation with a proof that justifies it.

Solution.

思路。X_n 与 X_{n+3} 下标不同 ⇒ 独立，联合概率拆成两个边缘概率之积。每个区间概率 = 区间长 + 一个 $O(1/n)$ 的余弦余项。求和后主项 $= 100 \times 0.2 \times 0.1 = 2$，余项被 $\sum O(1/n)$ 控制到 $\sim 10^{-6}$，故答案 ≈ 2，且能给严格上界。用 4 的「振荡项渐近消失」思路。

Independence + interval probabilities. For $m \ne n$, $X_m \perp X_n$, so the joint probability factors. For an interval $[a,b] \subseteq [0,1]$,

$$\mathrm{P}(a \le X_n \le b) = \int_a^b \big(1+\cos(2\pi n x)\big)dx = (b-a) + \frac{\sin(2\pi n b) - \sin(2\pi n a)}{2\pi n} = (b-a) + R_n(a,b),$$

with $|R_n(a,b)| \le \dfrac{1}{\pi n}$. Therefore, with $p_n := \mathrm{P}(0.5 \le X_n < 0.7) = 0.2 + R_n(0.5, 0.7)$

and $q_n := \mathrm{P}(0.2 < X_{n+3} \le 0.3) = 0.1 + R_{n+3}(0.2, 0.3)$,

$$S := \sum_{n=501}^{600} p_n\, q_n = \sum_{n=501}^{600} \big(0.2 + R_n\big)\big(0.1 + R_{n+3}\big).$$

Leading term and error bound. The leading term is $\sum_{n=501}^{600} 0.2 \cdot 0.1 = 100 \cdot 0.02 = 2$. The remainder is $\sum_{n=501}^{600} \big(0.2\, R_{n+3} + 0.1\, R_n + R_n R_{n+3}\big)$, with

$$|0.2\, R_{n+3} + 0.1\, R_n + R_n R_{n+3}| \le \frac{0.2}{\pi(n+3)} + \frac{0.1}{\pi n} + \frac{1}{\pi^2 n(n+3)}.$$

Over $n \in [501, 600]$ each summand is $\le \frac{0.3}{\pi \cdot 501} + \frac{1}{\pi^2 \cdot 501^2} \approx 1.9 \times 10^{-4}$, so the total remainder is at most $\approx 2 \times 10^{-2}$ in this crude bound; the exact value (the sin terms nearly cancel) is $\approx 3.6 \times 10^{-6}$. Hence

$$\boxed{\begin{array}{l} \displaystyle\sum_{n=501}^{600} \mathrm{P}(0.5 \le X_n < 0.7,\ 0.2 < X_{n+3} \le 0.3) = 2 + r, \\ |r| \le 2 \times 10^{-2}\ \text{(crude)}, \quad \text{actual} \approx 3.6 \times 10^{-6}, \quad \text{so} \ \approx 2. \end{array}}$$

Remark (Q5 要点).

两件事拿满分：(i) 由下标不同断独立，把联合概率拆开（不是「同一个 X」）；(ii) 给出 $|R_n| \le 1/(\pi n)$ 这一可证上界来 justify「$\approx$ 区间长」。主项 $100 \cdot 0.2 \cdot 0.1 = 2$。题目允许近似但要求证明，余项界就是那个证明。

8.6　Q6（501）四道真伪概念题

归属： ECON 501（Sung Jae Jun）　**重要度：** 高　**再考判断：** 高
核心考点： 测度概率与收敛概念真伪
历史复现： 2020、2021、2022、2024、2025
掌握方式： 每项准备最短证明或反例

原题 Q6 (verbatim)

State whether each of the following statements is true or false, and explain why.

(a) If $X_n \xrightarrow{p} 0$, then X_n is approximately equal to 0 when n is large, and therefore we must have $\mathrm{P}(X_n \le 0) \to 1$.

(b) The p-value, significance level, size of a test are all interchangeable terms, which indicate the chance of making a Type I error.

(c) Suppose that we are interested in some parameter θ_0, for which an asymptotically normal estimator $\widehat{\theta}$ is available. Suppose that an estimate $\widehat{\theta} = 0.10$ is reported with the standard error 0.025. Based on this information, it is correct to say that θ_0 is approximately between 0.05 and 0.15 with probability 95%.

(d) Suppose that we have an i.i.d. sample $\{X_1, \ldots, X_n\}$, where $X_1 \sim N(\mu_0, \sigma_0^2)$. If we knew σ_0, then we would be able to estimate μ_0 more accurately than we could without knowing σ_0.

Solution.

(a) FALSE. $\xrightarrow{p} 0$ 只说 $\mathrm{P}(|X_n| > \varepsilon) \to 0$，对 X_n 落在 0 哪一侧毫无约束。Convergence in probability means $\mathrm{P}(|X_n| > \varepsilon) \to 0$ for every $\varepsilon > 0$; it says nothing about the *sign*. Counterexample: $X_n \sim N(0, 1/n)$. Then $X_n \xrightarrow{p} 0$, yet by symmetry $\mathrm{P}(X_n \le 0) = \frac{1}{2} \not\to 1$. The conclusion fails.

$$\boxed{\text{(a) False: } X_n \sim N(0,1/n) \Rightarrow X_n \xrightarrow{p} 0 \text{ but } \mathrm{P}(X_n \le 0) = \tfrac{1}{2}.}$$

(b) FALSE. 三者是三回事。The *p-value* is a data-dependent random variable (under H_0 with a continuous test statistic it is Uniform$(0,1)$); the *significance level* α is a pre-chosen threshold the researcher fixes; the *size* is $\sup_{\theta\in\Theta_0} \mathrm{P}_{H_0}(\text{reject})$, the test's worst-case Type I error rate (6, 13). A level-α test rejects when p-value $< \alpha$, but the three are not interchangeable.

$$\boxed{\text{(b) False: p-value (random)} \ne \text{ level (chosen)} \ne \text{ size (sup Type I error).}}$$

(c) FALSE (as a probability statement about θ_0). θ_0 是固定常数，不是随机的；95% 指随机区间的长期覆盖率，不是「固定 θ_0 落在这个已实现区间」的概率。θ_0 is a fixed (nonrandom) parameter; the realized interval $[0.05, 0.15] = [\widehat{\theta} \pm 1.96 \cdot \mathrm{se}]$ is one realization of the random interval $[\widehat{\theta} \pm 1.96\,\widehat{\mathrm{se}}]$, which covers θ_0 with *asymptotic frequency* 95% over repeated samples (8). It is not correct to assign probability 95% to the event that the fixed θ_0 lies in this particular realized interval.

$$\boxed{\text{(c) False: 95\% is long-run coverage of the random CI, not } \mathrm{P}(\theta_0 \in [0.05, 0.15]).}$$

(d) FALSE. 正态的均值与方差参数正交：信息矩阵块对角，已知 σ_0 对估 μ_0 无效率增益。For $X_i \overset{i.i.d.}{\sim} N(\mu_0, \sigma_0^2)$, the score for μ is $\sum(X_i - \mu)/\sigma^2$ and for σ^2 is $\frac{1}{2}\sum[(X_i-\mu)^2/\sigma^4 - 1/\sigma^2]$; the Fisher information is block-diagonal,

$$I(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma_0^2 & 0 \\ 0 & 1/(2\sigma_0^4) \end{pmatrix},$$

so the μ-block of I^{-1} is σ_0^2 whether or not σ_0 is known (5). The MLE $\bar{X}_n$ has variance σ_0^2/n in both cases; knowing σ_0 gives *no* efficiency gain for μ_0 (parameter orthogonality / ancillarity of the scale for the location).

$$\boxed{\text{(d) False: } \mu, \sigma^2 \text{ are info-orthogonal; } \operatorname{Var}\left(\bar{X}_n\right) = \sigma_0^2/n \text{ regardless.}}$$

8.7 Q1（510）OLS 渐近分布与异方差稳健方差的一致性

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：OLS/极值估计渐近、稳健方差及一致性
历史复现：2017、2022、2023、2024；历年课程核心
掌握方式：FOC 四步、sandwich 和残差替换必须逐行重现

原题 Q1 (verbatim)

Suppose $Y_i = X_i'\beta + U_i$ for $i = 1, 2, \ldots, n$, $\{(X_i, U_i) : i \geq 1\}$ are iid with $E(U_i \mid X_i) = 0$ a.s. and $E(U_i^2 \mid X_i) < \infty$ a.s.

(a) Derive the asymptotic distribution of the least squares estimator.

(b) Provide a consistent estimator for the asymptotic variance of the least squares estimator and prove consistency.

Precisely state any additional assumptions you may need.

Solution.

思路。标准 OLS 渐近：$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1})$，$Q = \mathbb{E}(X_i X_i')$，$\Omega = \mathbb{E}\left(X_i X_i' U_i^2\right)$。证明走 LLN + CLT + Slutsky/CMT（8 极值估计 → 推断模板，OLS 是其线性特例）。(b) White/Eicker–Huber 夹心估计，一致性靠 LLN + 残差替换 + CMT，要补 $\mathbb{E}\left(\|X_i\|^4\right) < \infty$ 一类矩条件。

(a) *Additional assumptions:* (A1) $Q := \mathbb{E}(X_i X_i')$ exists and is nonsingular (positive definite); (A2) $\Omega := \mathbb{E}\left(X_i X_i' U_i^2\right)$ exists and is finite (e.g. $\mathbb{E}\left(\|X_i\|^2 U_i^2\right) < \infty$). From $\widehat{\beta} = \left(\frac{1}{n}\sum X_i X_i'\right)^{-1} \frac{1}{n}\sum X_i Y_i$ and $Y_i = X_i'\beta + U_i$,

$$\sqrt{n}(\widehat{\beta} - \beta) = \left(\tfrac{1}{n}\textstyle\sum_i X_i X_i'\right)^{-1} \left(\tfrac{1}{\sqrt{n}}\textstyle\sum_i X_i U_i\right).$$

By the WLLN, $\frac{1}{n}\sum X_i X_i' \xrightarrow{p} Q$ (nonsingular by (A1)), so by CMT the inverse $\xrightarrow{p} Q^{-1}$. The summands $X_i U_i$ are iid with mean $\mathbb{E}(X_i U_i) = \mathbb{E}(X_i \mathbb{E}(U_i \mid X_i)) = 0$ and finite variance Ω (by (A2)), so by the Lindeberg–Lévy CLT $\frac{1}{\sqrt{n}}\sum X_i U_i \xrightarrow{d} N(0, \Omega)$. By Slutsky,

$$\boxed{\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0,\ V), \qquad V = Q^{-1}\Omega Q^{-1},\ \ Q = \mathbb{E}\left(X_i X_i'\right),\ \Omega = \mathbb{E}\left(X_i X_i' U_i^2\right).}$$

(Under conditional homoskedasticity $\mathbb{E}\left(U_i^2 \mid X_i\right) = \sigma^2$, $\Omega = \sigma^2 Q$ and $V = \sigma^2 Q^{-1}$.)

(b) *Additional assumption:* (A3) $\mathbb{E}\left(\|X_i\|^4\right) < \infty$ and $\mathbb{E}\left(\|X_i\|^2 U_i^2\right) < \infty$ (so the LLN applies to the relevant fourth-order averages). With residuals $\hat{U}_i = Y_i - X_i'\widehat{\beta}$, the Eicker–Huber–White heteroskedasticity-robust sandwich estimator is

$$\hat{V} = \hat{Q}^{-1}\hat{\Omega}\,\hat{Q}^{-1}, \qquad \hat{Q} = \tfrac{1}{n}\sum_i X_i X_i', \qquad \hat{\Omega} = \tfrac{1}{n}\sum_i X_i X_i' \hat{U}_i^2.$$

Proof of consistency. $\hat{Q} \xrightarrow{p} Q$ by the WLLN and $\hat{Q}^{-1} \xrightarrow{p} Q^{-1}$ by CMT (using (A1)). For $\hat{\Omega}$, expand $\hat{U}_i = U_i - X_i'(\widehat{\beta} - \beta)$, so $\hat{U}_i^2 = U_i^2 - 2U_i X_i'(\widehat{\beta} - \beta) + \left(X_i'(\widehat{\beta} - \beta)\right)^2$ and

$$\hat{\Omega} = \underbrace{\tfrac{1}{n}\sum X_i X_i' U_i^2}_{\xrightarrow{p}\Omega\ \text{(WLLN, (A2))}} - 2\left(\tfrac{1}{n}\sum X_i X_i' U_i X_i'\right)(\widehat{\beta} - \beta) + \left(\tfrac{1}{n}\sum X_i X_i' \left(X_i'(\widehat{\beta} - \beta)\right)^2\right).$$

By (A3) the two correction averages are $O_p(1)$ (bounded in probability via the LLN/Markov), while $\widehat{\beta} - \beta \xrightarrow{p} 0$ from (a); hence both correction terms are $o_p(1)$. Therefore $\hat{\Omega} \xrightarrow{p} \Omega$, and by CMT

$$\boxed{\hat{V} = \hat{Q}^{-1}\hat{\Omega}\,\hat{Q}^{-1} \xrightarrow{p} Q^{-1}\Omega Q^{-1} = V\,.}$$

Remark (Q1（510）要点).

(a) 误差只需均值独立 $\mathbb{E}(U \mid X) = 0$(不必同方差),故渐近方差是夹心型 $Q^{-1}\Omega Q^{-1}$,稳健型才对。(b) 一致性证明的核心是把 $\hat{U}_i^2$ 展开、用 $\widehat{\beta} \xrightarrow{p} \beta$ 把交叉项压成 $o_p(1)$;要 *precisely state* 的额外条件就是 Q 满秩 (A1)、Ω 有限 (A2)、四阶矩 (A3)。

8.8 Q2（510）非光滑 GMM：V_0, Γ 的估计与分位回归

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：非光滑 GMM 与分位回归
历史复现：2023、2024、2025 连续出现
掌握方式：同 2023/2025 Q2，必须形成条件反射

原题 Q2 (verbatim)

Under Assumptions EE3 and CF-NS we derived $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, (\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1}\right)$ for the GMM estimator with nonsmooth stochastic criterion $Q_n(\theta) = \|\bar{g}_n(\theta)\|$, $\bar{g}_n(\theta) = n^{-1}\sum_i g(W_i, \theta)$, $g(\theta) = \mathbb{E}(g(W_i, \theta))$. [CF-NS: (i) θ_0 interior; (ii) $g(\theta)$ differentiable at θ_0 with $\Gamma = (\partial/\partial\theta')g(\theta_0)$ full rank $d \le k$; (iii) $g(\theta_0) = 0$; (iii′) $\sqrt{n}\,\bar{g}_n(\theta_0) \xrightarrow{d} N(0, V_0)$; (iv) stochastic equicontinuity $\sup_{\|\theta - \theta_0\| < \delta_n} \sqrt{n}\,\|\bar{g}_n(\theta) - g(\theta) - \bar{g}_n(\theta_0)\| \xrightarrow{p} 0$. EE3: $Q_n(\hat{\theta}_n) =$

$\inf_\theta Q_n(\theta) + o_p(n^{-1/2})$, $\hat{\theta}_n \xrightarrow{p} \theta_0$.]

(a) Provide estimators for V_0 and Γ and discuss their consistency under appropriate conditions.

Consider Quantile Regression with $(Y_i, X_i')'$ iid, $Y_i = X_i'\theta_0 + U_i$, where the τ-quantile of U_i conditional on X_i equals 0 for some $\tau \in (0,1)$.

(b) Derive what V_0 and Γ are in this case.

Solution.

关键洞察（11）。准则 Q_n 非光滑（含示性函数），但总体矩 $g(\theta) = \mathbb{E}(g(W_i,\theta))$ 是光滑的——期望把示性函数抹成 CDF/密度。所以 Γ 是光滑的 $g(\theta)$ 在 θ_0 的雅可比，而 V_0 是 $g(W_i,\theta_0)$ 的协方差。(a) 给样本类比估计；(b) 落到分位回归，$g = X(\mathbb{1}(Y \le X'\theta) - \tau)$，$V_0 = \tau(1-\tau)\mathbb{E}(XX')$，$\Gamma = \mathbb{E}\big(f_{U|X}(0)XX'\big)$。

(a) Estimators.

$$\hat{V}_0 = \tfrac{1}{n}\sum_{i=1}^{n} g(W_i,\hat{\theta}_n)\, g(W_i,\hat{\theta}_n)',$$

the sample second moment of $g(W_i,\hat{\theta}_n)$ (recentering at $\bar{g}_n(\hat{\theta}_n)$ is asymptotically negligible since $\bar{g}_n(\hat{\theta}_n) \xrightarrow{p} g(\theta_0) = 0$). *Consistency:* by a uniform LLN over a neighborhood of θ_0 (Glivenko–Cantelli / ULLN class condition) plus $\hat{\theta}_n \xrightarrow{p} \theta_0$ and continuity of $\theta \mapsto \mathbb{E}(g(W,\theta)g(W,\theta)')$ at θ_0, $\hat{V}_0 \xrightarrow{p} V_0 = \mathbb{E}(g(W,\theta_0)g(W,\theta_0)')$.

For $\Gamma = (\partial/\partial\theta')g(\theta_0)$, do *not* differentiate the nonsmooth $\bar{g}_n$; estimate the derivative of the *smooth* limit $g(\theta) = \mathbb{E}(g(W,\theta))$. Two standard routes: (i) a closed-form plug-in when $g(\theta)$ is known analytically (as in QR below, it involves a conditional density estimated by a kernel with bandwidth $h_n \to 0$); (ii) a numerical derivative

$$\hat{\Gamma}_{\cdot j} = \frac{\bar{g}_n(\hat{\theta}_n + \epsilon_n e_j) - \bar{g}_n(\hat{\theta}_n - \epsilon_n e_j)}{2\epsilon_n}, \quad \epsilon_n \to 0,\ \frac{1}{\sqrt{n}\epsilon_n} = O_p(1),$$

consistent under CF-NS(v), which makes the relevant local empirical numerator $o_p(n^{-1/2})$. The stronger convenient choice $\sqrt{n}\epsilon_n \to \infty$ also suffices. Either route needs the step to vanish at an appropriate rate and g to be differentiable at θ_0.

$$\boxed{\hat{V}_0 = \tfrac{1}{n}\textstyle\sum_i g(W_i,\hat{\theta}_n)g(W_i,\hat{\theta}_n)'; \quad \hat{\Gamma} = \text{numerical/kernel derivative of the smooth } g(\theta) \text{ at } \hat{\theta}_n.}$$

(b) QR case. Take $g(W_i,\theta) = X_i\big(\mathbb{1}(Y_i \le X_i'\theta) - \tau\big)$ (the QR moment; its expectation is the FOC of the check-function GMM). The conditional-τ-quantile-zero condition gives $\mathrm{P}(Y_i \le X_i'\theta_0 \,|\, X_i) = \mathrm{P}(U_i \le 0 \,|\, X_i) = \tau$, so $\mathbb{E}(g(W_i,\theta_0)) = \mathbb{E}(X_i(\tau - \tau)) = 0$, confirming $g(\theta_0) = 0$.

V_0: since $\mathbb{1}(Y_i \le X_i'\theta_0) - \tau$ has conditional mean 0 and conditional variance

$\tau(1-\tau)$ (centered Bernoulli(τ)),

$$V_0 = \mathrm{Var}(g(W_i,\theta_0)) = \mathbb{E}\big(X_iX_i'\,\mathbb{E}\big((\mathbb{1}(U_i \le 0) - \tau)^2 \,|\, X_i\big)\big) = \tau(1-\tau)\,\mathbb{E}\big(X_iX_i'\big)\,.$$

Γ: the smooth limit is $g(\theta) = \mathbb{E}\big(X_i\big(F_{U|X}(X_i'(\theta-\theta_0)) - \tau\big)\big)$; differentiating at $\theta = \theta_0$ (inner argument 0),

$$\Gamma = \frac{\partial}{\partial\theta'} g(\theta)\Big|_{\theta_0} = \mathbb{E}\big(X_iX_i'\,f_{U|X}(0 \mid X_i)\big)\,.$$

$$\boxed{\begin{aligned} &V_0 = \tau(1-\tau)\,\mathbb{E}\big(X_iX_i'\big)\,,\\ &\Gamma = \mathbb{E}\big(f_{U|X_i}(0)\,X_iX_i'\big)\,. \end{aligned}}$$

With $d = k$ (just-identified) the sandwich collapses to

$$\Gamma^{-1}V_0\Gamma^{-1\prime} = \tau(1-\tau)\,\Gamma^{-1}\mathbb{E}\big(XX'\big)\,\Gamma^{-1\prime},$$

the classical Koenker–Bassett QR variance.

Remark (Q2（510）要点).

核心反复强调：非光滑的是准则，不是总体矩。V_0 用样本协方差直接估（在 $\hat{\theta}_n$ 处取 g）；Γ 必须对光滑的 $g(\theta) = \mathbb{E}(g)$ 求导，不能对 $\bar{g}_n$ 求导（处处不可微）。QR 里 Γ 自然冒出条件密度 $f_{U|X}(0)$——这正是分位回归方差里那个出名的 nuisance 密度。

8.9 Q3（510）Heckit 选择模型：逆 Mills、probit 识别、排除限制

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：Heckit、逆 Mills、probit 与排除限制识别
历史复现：2023 选择模型、2024、2025
掌握方式：两步公式、逆 Mills 来源和识别角色都要会解释

原题 Q3 (verbatim)

In the Heckit model (Heckman, 1979), using the same notation/assumptions as in class ($y^* = x'\theta + \varepsilon$, $y = dy^*$, $d = \mathbb{1}(x'\pi_1 + z'\pi_2 \ge -\eta)$, joint normality of (ε, η)):

(a) Show that $E(y \mid x, z, d = 1) = x'\theta + \sigma_{\varepsilon\eta}\lambda(x'\pi_1 + z'\pi_2)$, where $\lambda(s) = \phi(s)/\Phi(s)$.

(b) Identify π_1 and π_2 from a probit regression of d onto x and z.

(c) Specify conditions under which one can identify θ and $\sigma_{\varepsilon\eta}$.

Solution.

思路（7）。(a) 在 $d=1$（即 $\eta \geq -(x'\pi_1+z'\pi_2)$）上对 ε 取条件期望，用双正态「$\mathbb{E}(\varepsilon\,|\,\eta)=\sigma_{\varepsilon\eta}\eta$」（归一化 $\mathrm{Var}(\eta)=1$）加单边截断的逆 Mills 比。(b) $\mathrm{P}(d=1\,|\,x,z)=\Phi(x'\pi_1+z'\pi_2)$，probit 直接识别 (π_1,π_2)。(c) 排除限制：z 中至少一个变量进选择方程但不进结果方程，使 $\lambda(\cdot)$ 有独立于 x 的变异，避免与 x 共线。

(a) Normalize $\mathrm{Var}(\eta)=1$ (only the scale of the index is identified from d). With (ε,η) jointly normal, $\mathbb{E}(\varepsilon\,|\,\eta)=\sigma_{\varepsilon\eta}\,\eta$. Conditioning on $(x,z,d=1)$, i.e. on $\eta \geq -(x'\pi_1+z'\pi_2) =: -s$,

$$\mathbb{E}(\varepsilon\,|\,x,z,d=1)=\sigma_{\varepsilon\eta}\,\mathbb{E}(\eta\,|\,\eta\geq -s)=\sigma_{\varepsilon\eta}\,\frac{\phi(-s)}{1-\Phi(-s)}=\sigma_{\varepsilon\eta}\,\frac{\phi(s)}{\Phi(s)}=\sigma_{\varepsilon\eta}\,\lambda(s),$$

using $\phi(-s)=\phi(s)$, $1-\Phi(-s)=\Phi(s)$, and the truncated-normal mean $\mathbb{E}(\eta\,|\,\eta\geq c)=\phi(c)/(1-\Phi(c))$. Since $y=y^*=x'\theta+\varepsilon$ on $\{d=1\}$,

$$\boxed{\mathbb{E}(y\,|\,x,z,d=1)=x'\theta+\sigma_{\varepsilon\eta}\,\lambda(x'\pi_1+z'\pi_2),\qquad \lambda(s)=\phi(s)/\Phi(s).}$$

(b) The selection indicator is $d=\mathbb{1}(\eta\geq -(x'\pi_1+z'\pi_2))$ with $\eta\sim N(0,1)$ independent of (x,z), so

$$\mathrm{P}(d=1\,|\,x,z)=\mathrm{P}\big(\eta\geq -(x'\pi_1+z'\pi_2)\big)=\Phi(x'\pi_1+z'\pi_2).$$

Hence a probit of d on (x,z) has index coefficients exactly (π_1,π_2), identified provided (x',z') has a nonsingular second-moment matrix (no perfect collinearity) and $\mathrm{Var}(\eta)$ is normalized to 1.

$$\boxed{(\pi_1,\pi_2)\text{ identified as the probit coefficients in }\mathrm{P}(d=1\,|\,x,z)=\Phi(x'\pi_1+z'\pi_2).}$$

(c) From (a), on the selected sample $y=x'\theta+\sigma_{\varepsilon\eta}\lambda(\hat{s})+\text{error}$ with $\hat{s}=x'\pi_1+z'\pi_2$ (π from the first stage). $\theta,\sigma_{\varepsilon\eta}$ are identified by a regression of y on $(x,\lambda(\hat{s}))$ provided the regressors are not collinear. Because $\lambda(\cdot)$ is a nonlinear function of the index it has *some* variation even when $z=x$, but identification is fragile (near-collinearity) unless there is an **exclusion restriction**:

$$\boxed{\begin{array}{l}\text{(i) }\exists\text{ a component of }z\text{ entering the selection index }(\neq 0\text{ in }\pi_2)\text{ but}\\ \quad\textit{excluded}\text{ from the outcome equation, giving }\lambda(\cdot)\text{ variation independent of }x;\\ \text{(ii) joint normality of }(\varepsilon,\eta)\text{ (so the correction is exactly }\sigma_{\varepsilon\eta}\lambda);\\ \text{(iii) full-rank / no-collinearity of }(x,\lambda(\hat{s}))\text{ on the selected sample.}\end{array}}$$

Under (i)–(iii), the second-stage OLS of y on $(x,\lambda(\hat{s}))$ identifies θ (coefficient on x) and $\sigma_{\varepsilon\eta}$ (coefficient on λ).

Remark (Q3（510）要点).

(c) 评分点在排除限制：没有「只进选择、不进结果」的工具变量 z，$\lambda(x'\pi_1)$ 会和 x 高度共线，靠的纯是 λ 的非线性弯曲来识别——脆弱且不可信。正态性是「修正项恰为 $\sigma_{\varepsilon\eta}\lambda$」的来源，也要列。

8.10 Q4（510）矩不等式检验：漂移序列、CR 临界值、size 修正

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：矩不等式、drifting sequence、AsyCS 与 Bonferroni
历史复现：2023、2024、2025 size 主线
掌握方式：极限实验与 size 修正步骤需理解；关键定义需背

原题 Q4 (verbatim)

Consider the model with moment inequalities/equalities $E_F m_j(W_i,\theta_0) \geq 0$ for $j = 1,\dots,p$ and $E_F m_j(W_i,\theta_0) = 0$ for $j = p+1,\dots,p+v$. Define $\gamma_1 = (\gamma_{1,1},\dots,\gamma_{1,p})' \in R_+^p$ with $\gamma_{1,j} = \sigma_{F,j}(\theta)^{-1}E_F m_j(W_i,\theta)$, $\sigma_{F,j}(\theta) = \mathrm{Var}_F^{1/2}(m_j(W_i,\theta))$; $\gamma_2 = \Omega = \Omega(\theta,F) = \mathrm{Corr}_F(m(W_i,\theta))$ the $k\times k$ correlation matrix; $\gamma_3 = (F,\theta)$. Define $T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta),\hat{\Sigma}_n(\theta))$, where $\hat{\Sigma}_n(\theta) = n^{-1}\sum_i(m(W_i,\theta)-\bar{m}_n(\theta))(m(W_i,\theta)-\bar{m}_n(\theta))'$ is an estimator of the asymptotic variance matrix $\Sigma(\theta)$ of $n^{1/2}\bar{m}_n(\theta)$. Assume (i) S non-increasing in m_1, (ii) $S(m,\Sigma) = S(\Delta m,\Delta\Sigma\Delta)$ for pd diagonal Δ. Under drifting sequences $\gamma_{n,h}$ with $n^{1/2}\gamma_{n,h,1}\to h_1\in[0,\infty]^p$, $\gamma_{n,h,2}\to h_2$, one has $T_n(\theta_{n,h}) \xrightarrow{d} S(h_2^{1/2}Z+(h_1,0_v),h_2)\sim J_h$, $Z\sim N(0,I_k)$, with $1-\alpha$ quantile $c_h(1-\alpha)$. Test $H_0:\theta=\theta_0$ at nominal size α (two-sided) using $T_n(\theta_0)$; h_1 cannot be consistently estimated but one can construct an asymptotically valid $(1-\beta)$-confidence region $CR_{h_1,n}(1-\beta)$ for h_1.

(a) Discuss how to construct $CR_{h_1,n}(1-\beta)$.

(b) [With $cv(1-\alpha) = \sup_{\bar{h}_1\in CR_{h_1,n}(1-\beta)} c_{(\bar{h}_1,\hat{h}_{2,n})}(1-\alpha)$ for a consistent $\hat{h}_{2,n}$ of h_2.] Provide a candidate for $\hat{h}_{2,n}$.

(c) Is the asymptotic size of the test that rejects if $T_n(\theta_0) > cv(1-\alpha)$ equal to α? Bigger? Smaller?

(d) Is there a modification rejecting if $T_n(\theta_0) > cv(1-\delta)$ for an appropriate nonrandom $\delta>0$ (depending on α,β) guaranteeing asymptotic size $\leq\alpha$? If yes, specify δ and prove your claim.

(e) If (d) is yes, what can be said about power relative to the worst-case plug-in approach?

Solution.

框架（Andrews–Soares / Andrews–Guggenberger，见 13）。h_1 是各不等式的（缩放）松弛度 $n^{1/2}\gamma_{1,j}$，不可一致估计（$O(1)$ 不消失），但可造它的置信域。临界值取「在 h_1 的置信域上对最不利配置取 sup」。(a) 反演关于 γ_1 的检验造 CR；(b) h_2 是相关矩阵，用样本相关矩阵估；(c) 原程序只能保证 size $\le \alpha+\beta$，所以一般不能保证 level α；(d) 原题把尾概率参数写成 δ，应取 $\delta=\alpha-\beta$，Bonferroni 分账后才有 size $\le \alpha$；(e) 缩小 nuisance 集会降低临界值，但改用更高分位数会提高临界值，故功效没有统一排序。

(a) Constructing $CR_{h_1,n}(1-\beta)$. $h_1=\lim n^{1/2}\gamma_{n,h,1}$ collects the scaled slackness $n^{1/2}\sigma_{F,j}^{-1}E_F m_j$ of the p inequalities. The studentized sample moment satisfies $n^{1/2}\hat\gamma_{1,j}:=n^{1/2}\bar m_{n,j}(\theta_0)/\hat\sigma_{n,j}(\theta_0)=n^{1/2}\gamma_{1,j}+\hat\sigma_{n,j}^{-1}n^{1/2}(\bar m_{n,j}-Em_j)\xrightarrow{d}h_{1,j}+Z_j$, Z_j asymptotically $N(0,1)$. Inverting, a one-sided (Bonferroni) componentwise $(1-\beta)$ CR is

$$CR_{h_1,n}(1-\beta)=\Big\{h_1\in[0,\infty]^p:\ h_{1,j}\ge \max\big(0,\ n^{1/2}\hat\gamma_{1,j}-k_{n,\beta}\big),\ j=1,\dots,p\Big\},$$

where $k_{n,\beta}$ is the $(1-\beta)$ critical value of $\max_j Z_j$ (e.g. $z_{1-\beta/p}$ by Bonferroni, or a joint-normal quantile using $\hat h_{2,n}$). Because the true h_1 is componentwise ≥ 0 and the CR is the upper set consistent with the data at level β, $\liminf \mathrm{P}(h_1\in CR_{h_1,n}(1-\beta))\ge 1-\beta$.

(b) Candidate for $\hat h_{2,n}$. $h_2=\Omega=\mathrm{Corr}_F(m(W_i,\theta_0))$ is the moment correlation matrix; estimate it by normalizing the sample covariance $\hat\Sigma_n(\theta_0)$ to a correlation matrix:

$$\boxed{\hat h_{2,n}=\hat D_n^{-1/2}\,\hat\Sigma_n(\theta_0)\,\hat D_n^{-1/2},\quad \hat D_n=\mathrm{Diag}\big(\hat\Sigma_n(\theta_0)\big).}$$

By the WLLN $\hat\Sigma_n\xrightarrow{p}\Sigma$ along the drifting sequences and by CMT (the normalization is continuous where variances are bounded away from 0), $\hat h_{2,n}\xrightarrow{p}h_2$.

(c) Asymptotic size. 按 CR 是否覆盖真 h_1 分解 rejection event：覆盖时尾概率至多 α；漏覆盖的坏事件本身至多 β。因此只能得到 $\alpha+\beta$，不能宣称 level α。 The test rejects if $T_n(\theta_0)>cv(1-\alpha)=\sup_{\bar h_1\in CR}c_{(\bar h_1,\hat h_{2,n})}(1-\alpha)$. On $\{h_1\in CR_{h_1,n}(1-\beta)\}$ (probability $\ge 1-\beta$), $cv(1-\alpha)\ge c_{(h_1,h_2)}(1-\alpha)$, so the rejection contribution on this event is asymptotically $\le\alpha$. On the complement (probability $\le\beta$) rejection is bounded only by 1. Hence the general bound is $\mathrm{P}(\text{reject})\le\alpha+\beta$ asymptotically:

$$\boxed{\text{(c) asymptotic size is bounded by } \alpha+\beta;\ \text{it need not be } \le\alpha.}$$

The word “conservative” is not justified relative to the nominal target α: although taking a supremum over a covered CR raises the critical value, the CR can miss the truth with probability as large as β, and the resulting bound $\alpha+\beta$ is *above* α. Depending on the model and dependence between the two events, the actual asymptotic size can be below, equal to, or above α; the general conclusion is only the upper bound $\alpha+\beta$.

(d) Size correction $\delta=\alpha-\beta$. 原题把 cv 的尾概率直接记为 δ：$cv(1-\delta)$。要把总预算拆成「conditional rejection δ + CR miss β」，必须令 $\delta+\beta=\alpha$，即 $\delta=\alpha-\beta$。 Yes. Assume $0<\beta<\alpha$ and take

$$\boxed{\delta=\alpha-\beta.}$$

Reject if

$$T_n(\theta_0)>cv(1-\delta):=\sup_{\bar h_1\in CR_{h_1,n}(1-\beta)} c_{(\bar h_1,\hat h_{2,n})}(1-\delta)=\sup_{\bar h_1\in CR} c_{(\bar h_1,\hat h_{2,n})}(1-\alpha+\beta).$$

Proof. Split on whether the CR covers h_1:

$$\mathrm{P}(T_n(\theta_0)>cv(1-\delta))\le \underbrace{\mathrm{P}(T_n(\theta_0)>cv(1-\delta),\ h_1\in CR)}_{\limsup\le\delta}+\underbrace{\mathrm{P}(h_1\notin CR)}_{\limsup\le\beta}.$$

On $\{h_1\in CR\}$, $cv(1-\delta)\ge c_{(h_1,h_2)}(1-\delta)$, and $T_n(\theta_0)\xrightarrow{d}J_{(h_1,h_2)}$, so the first term has $\limsup\le\delta$. The CR miss contributes at most β. Therefore

$$\boxed{\delta=\alpha-\beta:\quad \limsup_n \text{ size}\ \le\ \delta+\beta\ =\ (\alpha-\beta)+\beta\ =\ \alpha.}$$

The Bonferroni budget splits α into $\alpha-\beta$ for the test and β for the CR.

(e) Power. 两个效应方向相反：CR 把取 sup 的 nuisance 集缩小，会降低临界值；但 $\delta=\alpha-\beta<\alpha$ 使分位数从 $1-\alpha$ 提高到 $1-\alpha+\beta$，会抬高临界值。没有附加结构就不能统一排序。 The worst-case (least-favorable) plug-in uses

$$cv^{LF}(1-\alpha)=\sup_{h_1\in[0,\infty]^p} c_{(h_1,\hat h_{2,n})}(1-\alpha).$$

The Bonferroni procedure instead uses

$$cv^B=\sup_{h_1\in CR_{h_1,n}(1-\beta)} c_{(h_1,\hat h_{2,n})}(1-\alpha+\beta).$$

Restricting the supremum from the whole cone to the confidence region tends to make cv^B smaller and recover power. However, replacing the $1-\alpha$ quantile by the higher $1-\alpha+\beta$ quantile tends to make cv^B larger. Neither effect dominates in general. Hence Bonferroni can be more powerful when the CR is informative, but it is not uniformly more powerful than the worst-case test:

(e) Often a power gain is possible because the nuisance set is smaller, but there is no general or uniform power ordering.

Remark (Q4（510）要点).

> 这是 Andrews–Guggenberger 的 Bonferroni size-correction 标准考法。三件事别错：(c) 原程序只能给 size $\le \alpha+\beta$，所以未必 level α，不能称为相对 α 保守；(d) 因原题写的是 $cv(1-\delta)$，应选 $\delta=\alpha-\beta$，再用事件分解得到 $(\alpha-\beta)+\beta=\alpha$；若另一种记号写成 $cv(1-\alpha+\delta)$，那时才会写 $\delta=\beta$；(e) CR 缩小 nuisance 集与更高分位数方向相反，功效可能改善但不 uniformly dominate。$\hat{h}_{2,n}$ 就是样本相关矩阵。

8.11　Q5（510）非参 bootstrap 与子抽样对均值检验的 NRP 与一致性

归属：ECON 510（Guggenberger）　**重要度**：高　**再考判断**：高
核心考点：非参 bootstrap、subsampling、NRP 与一致性
历史复现：2024；课程 PS/讲义覆盖密集
掌握方式：Express–Condition–CLT–DCT 和 subsample rate 要会

原题 Q5 (verbatim)

Let $X_i \sim$ iid with unknown mean $\beta \in R$ and variance 1, $i=1,\dots,n$. Test $H_0:\beta=0$ vs $H_1:\beta\neq 0$ based on the statistic $n^{1/2}\bar{X}_n$, $\bar{X}_n=\sum_i X_i/n$. Two resampling schemes for the critical value:

(a) Let $X^*_{ib}\sim$ iid $\hat{F}_n$ for $i=1,\dots,n$, $b=1,\dots,B$ (say $B=1000$), $\hat{F}_n$ the EDF; define $\bar{X}^*_{nb}=\sum_i X^*_{ib}/n$. Take as critical value the $1-\alpha$ quantile of $n^{1/2}\bar{X}^*_{nb}$, $b=1,\dots,B$.

(b) Choose $s=n^{1/2}$. Consider all subsets S_j, $j=1,\dots,\binom{n}{s}$, of $\{X_1,\dots,X_n\}$ with s elements; for each compute $\bar{X}_{sj}=\sum_{i\in S_j}X_i/s$. Take as critical value the $1-\alpha$ quantile of $s^{1/2}\bar{X}_{sj}$.

What can you say about the asymptotic null rejection probability and consistency of the two procedures? Provide justifications. Would you recommend either? Explain.

Solution.

先识别题面的两个陷阱（12）。第一，备择写成 two-sided，但统计量和 $1-\alpha$ 分位数都是 *signed*；字面上只能把拒绝规则读成 upper-tail comparison。第二，bootstrap statistic 没有减去 bootstrap world 的真值 $\bar{X}_n$，所以 bootstrap critical value 会把观测统计量 $T_n=\sqrt{n}\bar{X}_n$ 自己复制进去。必须先分析题面原程序，再说明怎样修成真正的双侧检验。

Let $T_n := \sqrt{n}\bar{X}_n$. Because only a signed $1-\alpha$ critical value is supplied, interpret each literal procedure as rejecting when T_n exceeds that critical value. Assume the usual testing range $0<\alpha<1/2$. As is standard in resampling asymptotics, first analyze the ideal conditional quantile (equivalently let the number of bootstrap draws $B \to \infty$ fast enough); the finite-B qualification is stated below.

(a) Literal non-recentered bootstrap. Conditionally on the data, the EDF has mean $\bar{X}_n$ and variance $\hat{\sigma}_n^2 \xrightarrow{p} 1$. The bootstrap CLT gives

$$\sqrt{n}(\bar{X}_{nb}^* - \bar{X}_n) \xrightarrow{d}{}^{*} N(0,1).$$

Therefore the conditional $1-\alpha$ quantile $c_{n,1-\alpha}^*$ of the *uncentered* statistic $\sqrt{n}\bar{X}_{nb}^*$ satisfies

$$c_{n,1-\alpha}^* = \sqrt{n}\bar{X}_n + z_{1-\alpha} + o_p(1) = T_n + z_{1-\alpha} + o_p(1).$$

The rejection inequality is consequently

$$T_n > c_{n,1-\alpha}^* \quad\iff\quad 0 > z_{1-\alpha} + o_p(1).$$

Since $z_{1-\alpha} > 0$, under H_0

$$\boxed{\mathrm{P}_0\{\text{reject by (a)}\} \longrightarrow 0.}$$

The same cancellation occurs under every fixed $\beta \neq 0$: both T_n and the critical value contain the same leading term $\sqrt{n}\beta$. Hence the power also converges to 0, not 1:

$$\boxed{\text{Procedure (a) as written has asymptotic NRP 0 and is inconsistent.}}$$

Merely saying "the NRP is not α" misses the sharper conclusion: it degenerates to zero because the uncentered bootstrap critical value tracks the observed statistic itself. If B is held literally fixed at 1000, the limiting empirical critical value is $T_n + Z_{(r)}$, where $Z_{(r)}$ is the $r = \lceil (1-\alpha)B \rceil$ order statistic of B independent $N(0,1)$ draws. Then the limiting rejection probability is

$$\mathrm{P}\big(Z_{(r)} < 0\big) = \mathrm{P}(\mathrm{Binomial}(B, 1/2) \geq r)\,,$$

an astronomically small but nonzero Monte Carlo error for conventional $(B, \alpha) = (1000, 0.05)$. It does not approach α and does not restore consistency; the clean limit 0 above is the ideal-quantile/$B \to \infty$ result exam solutions normally report.

(b) Literal signed subsampling, $s = \sqrt{n}$. For a uniformly chosen subset of size s, the finite-population/subsampling CLT gives, conditionally on the full sample,

$$\sqrt{s}(\bar{X}_{sj} - \bar{X}_n) \xrightarrow{d}{}^{*} N(0,1),$$

because $s \to \infty$ and $s/n \to 0$. Thus the signed subsampling quantile satisfies

$$c_{n,1-\alpha}^{\text{sub}} = \sqrt{s}\,\bar{X}_n + z_{1-\alpha} + o_p(1).$$

Under H_0, $T_n = O_p(1)$ and

$$\sqrt{s}\,\bar{X}_n = \sqrt{\frac{s}{n}}\,T_n = n^{-1/4}T_n = o_p(1),$$

so $c^{\text{sub}}_{n,1-\alpha} \xrightarrow{p} z_{1-\alpha}$. Since $T_n \xrightarrow{d} N(0,1)$,

$$\boxed{\mathrm{P}_0\{T_n > c^{\text{sub}}_{n,1-\alpha}\} \longrightarrow \alpha.}$$

For a fixed alternative,

$$T_n - c^{\text{sub}}_{n,1-\alpha} = (\sqrt{n} - \sqrt{s})\beta + O_p(1) = (n^{1/2} - n^{1/4})\beta + O_p(1).$$

It tends to $+\infty$ if $\beta > 0$ and to $-\infty$ if $\beta < 0$. Hence

$$\boxed{\mathrm{P}_\beta\{\text{reject by (b)}\} \to \begin{cases} 1, & \beta > 0, \\ 0, & \beta < 0. \end{cases}}$$

Procedure (b) is valid and consistent for the *right-sided* alternative, but it is *not* consistent against the stated two-sided alternative $\beta \neq 0$ because it has zero limiting power for every fixed negative β.

Repairing the problem into a genuine two-sided test. Use the absolute target statistic

$$|T_n| = \left|\sqrt{n}\bar{X}_n\right|.$$

For the bootstrap, take the $1-\alpha$ conditional quantile of the *recentered absolute root*

$$\left|\sqrt{n}(\bar{X}^*_{nb} - \bar{X}_n)\right|.$$

That critical value converges to $z_{1-\alpha/2}$, so the NRP tends to α and, for every fixed $\beta \neq 0$, $|T_n| \to \infty$ while the critical value is $O_p(1)$; power tends to 1. A two-sided subsampling repair analogously uses the $1-\alpha$ quantile of $\left|\sqrt{s}(\bar{X}_{sj} - \bar{X}_n)\right|$ (or imposes the null directly); it is first-order valid when $s \to \infty$ and $s/n \to 0$ and is consistent for both signs.

Recommendation. Recommend *neither literal procedure* for the stated two-sided problem. After repair, prefer the recentered two-sided bootstrap for this smooth root-n mean problem: it is first-order valid and can exploit higher-order refinements, whereas subsampling is more robust but usually less accurate because

> it uses only $s = o(n)$ observations per resample.
>
> Literal (a): NRP $\to 0$, power $\to 0$;
>
> literal (b): NRP $\to \alpha$, but consistent only for $\beta > 0$;
>
> genuine two-sided version: use absolute values and a recentered resampling root.

Remark (Q5（510）要点).

> 评分核心有两层。(1) **中心化**：bootstrap critical value 是 $T_n + z_{1-\alpha} + o_p(1)$，与 T_n 相消，故题面程序的 NRP 和 fixed-alternative power 都趋零；不是笼统的「不等于 α」。(2) **方向**：题面写 $H_1 : \beta \neq 0$ 却只给 signed upper-tail statistic。subsampling 在 null 下有效，但只对 $\beta > 0$ consistent，对 $\beta < 0$ power $\to 0$。真正双侧检验要比较 $|T_n|$ 与 recentered absolute bootstrap/subsampling root 的 $1-\alpha$ 分位数。见 12。

Chapter 9　2025 资格考精解（Worked Solutions）

导读 · Roadmap（先读这里，不含公式）

本章把 2025 年资格考（Pinkse 501 六题 + Guggenberger 510 五题）逐题做透——这是与用户这一届同款的最近一年，分量最重。501 全是基础硬功夫：样本方差的极限分布、p 阶矩下样本均值的收敛率、χ^2_n 尾概率的 limsup、二元正态的截断条件期望、依概率收敛不蕴含期望收敛、Markov+Jensen 型不等式、指数分布的累积量加条件 MLE、以及一个 $[0,1]$ 上分段函数的可测性。510 是进阶套路：识别三概念与 size 三概念、非光滑 GMM 的 V_0, Γ 估计与分位回归实例、五个正态均值检验的临界值/功效/无 UMP、LASSO 的 oracle 性质与 Ridge= 增广 OLS、以及 Heckit 选择模型。每题先给原题（逐字），再给中文思路与 *English formal derivation*；用到的定理一律指回前两部分对应章节。

9.1　501 · Q1（渐近理论：样本方差、收敛率、尾概率 limsup）

归属：ECON 501（Pinkse）　**重要度**：高（必会）　**再考判断**：高
核心考点：样本方差、仅有 p 阶矩的率与尾概率 limsup
历史复现：2019–2025 渐近理论连续高频
掌握方式：样本方差 IF、von Bahr–Esseen/MZ 率和 Markov 界均需掌握

原题 501 · Q1（Asymptotic theory）

(a) Let $\hat\sigma^2, \sigma_0^2$ denote the sample and population variance, respectively, of an i.i.d. sample of size n with finite fourth moments. Derive the limit distribution of $\sqrt{n}(\hat\sigma^2 - \sigma_0^2)$.

(b) Establish the convergence rate of $\bar x$ if $\bar x$ is the sample mean of an i.i.d. sequence of random variables with $0 < \mathbb{E}(\|x\|^p) < \infty$ for some $1 < p < \infty$. Your answer should be a function of p and work for all such values of p.

(c) Let $\hat\sigma^2$ be the sample variance of an i.i.d. sample (of size n) of unit variance random variables with finite fourth moments. Let z_n be a χ^2_n-distributed random variable independent of $\hat{\sigma}^2$. Derive $\limsup_{n\to\infty} \mathrm{P}\big(n\hat{\sigma}^2 > z_n\big)$.

Solution.

思路。三小问层层加码：(a) 是标准 CLT+delta，方法全在 4；(b) 考「矩越少收敛越慢」，答案是 p 的函数；(c) 是把两个独立的「和」各自 CLT 化、看差的符号概率，结论是 1/2 而且极限存在故 limsup 即极限。记 $\mu_4 := \mathbb{E}\big((x-\mu)^4\big)$。

(a) Write $\mu = \mathbb{E}(x_1)$, $\sigma_0^2 = \mathbb{E}\big((x_1-\mu)^2\big)$, $\mu_4 = \mathbb{E}\big((x_1-\mu)^4\big) < \infty$. The (bias-uncorrected) sample variance is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 - (\bar{x}-\mu)^2.$$

The second term is $O_p(n^{-1})$ (since $\bar{x}-\mu = O_p(n^{-1/2})$), so $\sqrt{n}(\bar{x}-\mu)^2 = o_p(1)$ and it drops out after scaling. Let $w_i := (x_i-\mu)^2$, i.i.d. with mean σ_0^2 and variance $\mathrm{Var}(w_1) = \mathbb{E}\big((x_1-\mu)^4\big) - \sigma_0^4 = \mu_4 - \sigma_0^4 < \infty$. By the CLT (4),

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} w_i - \sigma_0^2\right) \xrightarrow{d} N\big(0,\ \mu_4-\sigma_0^4\big).$$

Adding the $o_p(1)$ term and using Slutsky,

$$\boxed{\sqrt{n}\left(\hat{\sigma}^2-\sigma_0^2\right) \xrightarrow{d} N\big(0,\ \mu_4-\sigma_0^4\big), \qquad \mu_4 = \mathbb{E}\big((x_1-\mu)^4\big)}$$

有限四阶矩的作用就在保证 $\mathrm{Var}(w_1) = \mu_4-\sigma_0^4$ 有限、CLT 才成立。用 $n-1$ 当分母只改 $O_p(n^{-1})$ 因子，不影响 $\sqrt{n}$ 级极限。正态时 $\mu_4 = 3\sigma_0^4$，方差退化为 $2\sigma_0^4$。

(b) 这是 Marcinkiewicz–Zygmund 型结论：手头只有 p 阶矩（$1<p<\infty$）时，样本均值围绕其极限的统一保证阶随 p 变。$p \ge 2$ 回到通常的 $\sqrt{n}$；$1<p<2$ 的 worst-case polynomial exponent 是 $1-1/p$。这里的 “rate” 不能被误读为每个满足该矩条件的分布都有同一个非退化极限。 Center at $\mu = \mathbb{E}(x_1)$ (which exists since $p>1$). We seek a normalization r_n whose order can be guaranteed using only the stated p-th moment.

Case $p \ge 2$: second moments are finite, the CLT gives $\sqrt{n}(\bar{x}-\mu) = O_p(1)$, i.e. rate $n^{1/2}$.

Case $1<p<2$: By the Marcinkiewicz–Zygmund / von Bahr–Esseen inequality (Thm 2.15), for i.i.d. mean-zero $y_i = x_i - \mu$ with $\mathbb{E}(|y_1|^p) < \infty$ and $1 \le p \le 2$,

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} y_i\right|^p\right) \le C_p\, n\, \mathbb{E}(|y_1|^p).$$

Hence $\mathbb{E}\big(\big|n^{-1}\sum_{i=1}^n y_i\big|^p\big) \le C_p\, n^{1-p}\mathbb{E}(|y_1|^p)$, so by Markov, for any $M>0$,

$$\mathrm{P}\Big(n^{1-1/p}\,|\bar{x}-\mu|>M\Big) \le \frac{n^{(1-1/p)p}\,\mathbb{E}(|\bar{x}-\mu|^p)}{M^p} = \frac{n^{p-1}\cdot C_p n^{1-p}\mathbb{E}(|y_1|^p)}{M^p} = \frac{C_p\mathbb{E}(|y_1|^p)}{M^p},$$

uniformly in n. Thus $n^{1-1/p}(\bar{x}-\mu)=O_p(1)$. The Marcinkiewicz–Zygmund strong law gives the sharper statement

$$n^{-1/p}\sum_{i=1}^n (x_i-\mu) \longrightarrow 0 \quad \text{a.s.,} \qquad 1<p<2,$$

so in fact $n^{1-1/p}(\bar{x}-\mu)=o_p(1)$. This does not supply a universal nondegenerate limit: a particular distribution may have finite variance and hence the faster root-n rate, while heavy-tailed distributions can make the exponent $1-1/p$ essentially sharp.

Combining the two cases, the convergence rate is

$$\boxed{\bar{x}-\mu = O_p\Big(n^{-\min\{1-1/p,\ 1/2\}}\Big), \quad r_n = n^{\min\{1-1/p,\ 1/2\}}}$$

min 在 $p=2$ 处接合：$1-1/p$ 从 0^+（$p\to 1$）升到 $1/2$（$p=2$），之后被 $1/2$ 封顶。对 $1<p<2$，M–Z 强大数律其实给 $r_n(\bar{x}-\mu)\to 0$；写 O_p 是为了用同一个公式覆盖全部 p。“收敛率” 在这里只是由给定矩条件可保证的 polynomial order，不是声称存在唯一极限分布。

(c) 两个对象 $n\hat{\sigma}^2$ 与 $z_n=\chi^2_n$ 都是 n 个 i.i.d. 项之和，各自中心化除以 $\sqrt{n}$ 后 CLT 成正态，且二者**独立**。要的是 $\mathrm{P}\big(n\hat{\sigma}^2>z_n\big)=\mathrm{P}\big((n\hat{\sigma}^2-n)-(z_n-n)>0\big)$ 的 limsup。 Unit variance means $\sigma_0^2=1$, so $\mathbb{E}\big(n\hat{\sigma}^2\big)=n-1$ (uncorrected sample variance) and $\mathbb{E}(z_n)=n$. Centering each by n and scaling by $\sqrt{n}$ leaves only an $O(n^{-1/2})$ mean drift in the first term (since $(\mathbb{E}\big(n\hat{\sigma}^2\big)-n)/\sqrt{n}=-1/\sqrt{n}\to 0$), asymptotically negligible:

$$\frac{n\hat{\sigma}^2-n}{\sqrt{n}} = \sqrt{n}\left(\hat{\sigma}^2-1\right) \xrightarrow{d} N(0,\ \mu_4-1) \quad \text{by (a) with } \sigma_0^2=1,$$

$$\frac{z_n-n}{\sqrt{n}} = \frac{\sum_{j=1}^n(\zeta_j^2-1)}{\sqrt{n}} \xrightarrow{d} N(0,\ 2), \qquad \zeta_j \overset{i.i.d.}{\sim} N(0,1),\ \mathrm{Var}\big(\zeta_1^2\big)=2.$$

The two are independent (given), so jointly

$$\frac{(n\hat{\sigma}^2-n)-(z_n-n)}{\sqrt{n}} \xrightarrow{d} N(0,\ (\mu_4-1)+2) = N(0,\ \mu_4+1).$$

Therefore

$$\mathrm{P}\big(n\hat{\sigma}^2>z_n\big) = \mathrm{P}\left(\frac{(n\hat{\sigma}^2-n)-(z_n-n)}{\sqrt{n}}>0\right) \longrightarrow \mathrm{P}(N(0,\mu_4+1)>0)=\tfrac{1}{2}.$$

The limit exists, hence the limsup equals it:

$$\boxed{\limsup_{n\to\infty} \mathrm{P}\big(n\hat{\sigma}^2 > z_n\big) = \tfrac{1}{2}}$$

两点细节：其一，中心化都用 n（虽然未校正样本方差有 $\mathbb{E}\big(n\hat{\sigma}^2\big) = n-1$，但这点 $O(1)$ 偏差除以 $\sqrt{n}$ 后是 $-1/\sqrt{n} \to 0$，均值漂移可忽略；$\mathbb{E}(z_n) = n$ 精确）；其二，正态过 0 的概率 $= 1/2$ 与方差 $\mu_4 + 1$ 无关——只要方差有限正态对称即可。μ_4 进答案的唯一作用是保证 CLT 成立、极限是真正态。limsup $=$ lim $= 1/2$。

Remark (Q1 三问要点).

(a) $\hat{\sigma}^2$ 的渐近方差 $\mu_4 - \sigma_0^4$ (不是 $2\sigma_0^4$, 那只是正态特例); (b) 收敛率 $n^{\min\{1-1/p,1/2\}}$, $p \geq 2$ 封顶在 $\sqrt{n}$; (c) 答案 1/2 且 limsup 即极限。三问分别考 delta/CLT、矩与率的关系、独立和之差的符号概率，全是 4 的招牌。

9.2 501 · Q2（二元正态：截断条件期望）

归属：ECON 501（Pinkse） **重要度**：高（必会） **再考判断**：高
核心考点：二元正态的截断条件期望
历史复现：2017、2018、2020、2021、2024、2025
掌握方式：先回归分解，再做单边截断正态

原题 501 · Q2

Suppose that x, y are jointly normal with mean zero, unit variance, and correlation ρ for which $\rho^2 = 3/4$. Compute $\mathbb{E}(y \mid y \leq x,\ x = 0)$.

Solution.

思路。先 *condition on* $x = 0$：二元正态的条件分布 $y \mid x = 0 \sim N\left(0,\ 1-\rho^2\right)$（见 3 的条件正态公式 $\mathbb{E}(y \mid x) = \rho x$、$\mathrm{Var}(y \mid x) = 1-\rho^2$）。再把事件 $\{y \leq x\}$ 在 $x = 0$ 处化成 $\{y \leq 0\}$，剩下就是一个**中心正态的左截断均值**。

Condition first on $x = 0$. By the bivariate-normal conditional law (3),

$$y \mid x = 0 \ \sim\ N\big(\rho \cdot 0,\ 1-\rho^2\big) = N\big(0,\ 1-\rho^2\big).$$

With $\rho^2 = 3/4$, the conditional standard deviation is $\sigma := \sqrt{1-\rho^2} = \sqrt{1/4} = 1/2$. On the event $x = 0$, the condition $y \leq x$ becomes $y \leq 0$. So we need the truncated mean of a centered normal:

$$\mathbb{E}(y \mid y \leq 0,\ x = 0) = \mathbb{E}(y \mid y \leq 0), \qquad y \sim N(0, \sigma^2).$$

For $y \sim N(0, \sigma^2)$, the left-truncated mean (inverse Mills ratio at 0) is

$$\mathbb{E}(y \mid y \le 0) = -\sigma \frac{\phi(0)}{\Phi(0)} = -\sigma \frac{1/\sqrt{2\pi}}{1/2} = -\sigma \sqrt{\frac{2}{\pi}}.$$

Plugging $\sigma = 1/2$:

$$\boxed{\mathbb{E}(y \mid y \le x,\ x = 0) = -\frac{1}{2}\sqrt{\frac{2}{\pi}} = -\frac{1}{\sqrt{2\pi}} \approx -0.399}$$

推导左截断均值：对 $y \sim N(0, \sigma^2)$，$\mathbb{E}(y\,\mathbb{1}\{y \le 0\}) = \int_{-\infty}^{0} t\,\frac{1}{\sigma}\phi(t/\sigma)\,dt = -\sigma\phi(0)$（用 $\frac{d}{dt}\phi(t/\sigma) = -\frac{t}{\sigma^2}\phi(t/\sigma)$），再除以 $\mathrm{P}(y \le 0) = 1/2$ 得 $-2\sigma\phi(0) = -\sigma\sqrt{2/\pi}$。这里 ρ 的符号不影响——只有 ρ^2 进入条件方差。

Remark (Q2 易错点).

别被 ρ 带偏。一旦 condition on $x = 0$，条件均值 $\rho \cdot 0 = 0$，ρ 只通过 $1 - \rho^2$ 进方差；事件 $\{y \le x\}$ 在 $x = 0$ 退化成 $\{y \le 0\}$。最终答案 $-1/\sqrt{2\pi}$ 恰好与 $\phi(0)$ 同形，是巧合（$\sigma\sqrt{2/\pi}\cdot$ 系数约掉）。

9.3 501 · Q3（依概率收敛 $\not\Rightarrow$ 期望收敛）

归属：ECON 501（Pinkse） **重要度**：中高 **再考判断**：中高
核心考点：密度上界与依概率收敛前提冲突
历史复现：收敛陷阱见 2020–2025；该空前提结构仅 2025
掌握方式：先检查前提可满足性；记 $P(|X_n| \le \epsilon) \le 2\epsilon$

原题 501 · Q3

Let x_n for any n be continuously distributed random variables with bounded support and density $f_n \le 1$. Suppose further that $x_n \xrightarrow{p} 0$ (converges in probability to 0). Prove or disprove the statement: “$\mathbb{E}(x_n)$ necessarily converges to zero.”

Solution.

思路与关键观察。先答常见层面：依概率收敛*一般*不蕴含期望收敛——缺一致可积（UI）时，「远端低概率尖峰」就是标准反例，故*若没有* $f_n \le 1$ 这条，命题为假。但本题恰恰加了 $f_n \le 1$，它是*决定性*约束：密度被 1 封顶时，任一长度 2ε 的区间至多承载概率 2ε，于是 $\mathrm{P}(|x_n| \le \varepsilon) \le 2\varepsilon$、即 $\mathrm{P}(|x_n| > \varepsilon) \ge 1 - 2\varepsilon$ *对所有* n 成立。固定小 ε（如 $\varepsilon = \frac{1}{4}$）该下界 $\ge \frac{1}{2}$、与 n 无关、*永不*趋于 0，故 $x_n \xrightarrow{p} 0$ 在 $f_n \le 1$ 下*根本不可能*发生。两个前提（$f_n \le 1$ 与 $x_n \xrightarrow{p} 0$）因此*互斥*：满足前提的随机变量列*不存在*，蕴含式*空虚为真*。这才是逻辑自洽的最终判定——不能既判「假」又拿不出合规反例。

(I) Without the constraint $f_n \le 1$ the statement is false. Convergence in probability does not control tails; absent uniform integrability, $\mathbb{E}(x_n) \not\to 0$ in general. Textbook counterexample (a point-mass spike, so $f_n \le 1$ is *violated*):

$$x_n = \begin{cases} 0, & \text{w.p. } 1 - \frac{1}{n}, \\ n, & \text{w.p. } \frac{1}{n}. \end{cases}$$

Then $\mathrm{P}(|x_n| > \varepsilon) = \frac{1}{n} \to 0$ so $x_n \xrightarrow{p} 0$, yet $\mathbb{E}(x_n) = n \cdot \frac{1}{n} = 1 \not\to 0$. The implication holds only under uniform integrability (Vitali's theorem). This is the content the question is testing —but note the spike puts mass $1 - \frac{1}{n}$ at the single point 0, which is a density violation, not a density ≤ 1.

(II) The constraint $f_n \le 1$ rules out $x_n \xrightarrow{p} 0$ entirely. For any density with $f_n \le 1$, every n and every $\varepsilon > 0$ give

$$\mathrm{P}(|x_n| \le \varepsilon) = \int_{-\varepsilon}^{\varepsilon} f_n(t)\,dt \le 2\varepsilon \cdot 1 = 2\varepsilon \quad \Longrightarrow \quad \mathrm{P}(|x_n| > \varepsilon) \ge 1 - 2\varepsilon.$$

Take $\varepsilon = \frac{1}{4}$: then $\mathrm{P}\big(|x_n| > \frac{1}{4}\big) \ge \frac{1}{2}$ for *every* n, so $\mathrm{P}\big(|x_n| > \frac{1}{4}\big) \not\to 0$ and $x_n \xrightarrow{p} 0$ *fails by definition.* There is no escape: a bounded density cannot concentrate mass near 0. Trying to "smear the spike" confirms this —to put probability $1 - \frac{1}{n} \approx 1$ in a shrinking window $[0, \delta_n]$ requires density $\approx 1/\delta_n \to \infty > 1$, illegal.

(III) Resolution: the hypotheses are jointly unsatisfiable, so the implication is vacuously TRUE. Because no random-variable sequence can satisfy both $f_n \le 1$ and $x_n \xrightarrow{p} 0$, the set of objects to which the claim applies is empty. A conditional with an unsatisfiable antecedent is vacuously true, so:

$$\boxed{\text{Under } f_n \le 1,\ x_n \xrightarrow{p} 0 \text{ is impossible } (\mathrm{P}(|x_n| > \varepsilon) \ge 1 - 2\varepsilon);\ \text{hence the implication holds } \textit{vacuously}.}$$

务必注意逻辑闭合：如果坚持判「命题为假」，就必须给出一个同时满足 $f_n \le 1$、$x_n \xrightarrow{p} 0$ 且 $\mathbb{E}(x_n) \not\to 0$ 的反例；而 (II) 证明这种序列*不存在*，所以「假 + 拿不出反例」是不自洽的。正确判定是*空虚为真*。命题真正想考的 UI 反例（见 (I)）只在去掉 $f_n \le 1$ 时才成立——这正是出题点：识别出 $f_n \le 1$ 把反例堵死了。

Remark (Q3 正确判定与考点).

两个得分层次：(a) 一般原理——依概率收敛不蕴含期望收敛，桥梁是一致可积（Vitali），无约束时「远端尖峰」即反例（见 (I)）；(b) 本题约束 $f_n \le 1$ 逼出 $\mathrm{P}(|x_n| > \varepsilon) \ge 1 - 2\varepsilon$，使 $x_n \xrightarrow{p} 0$ 不可能、前提自相矛盾，故蕴含式*空虚为真*。考场最稳的写法：先点 (a) 说明「没有 $f_n \le 1$ 时命题假、UI 是关键」，再用 (b) 指出题给的 $f_n \le 1$ 恰好让前提无法实现、命题空虚成立。不要在守 $f_n \le 1$ 的前提下硬判「假」却给不出合规反例。见 4。

9.4 501 · Q4（Markov + Jensen 型不等式，无独立性）

归属：ECON 501（Pinkse） **重要度**：高 **再考判断**：高
核心考点：Markov 与 Jensen 型不等式
历史复现：不等式见 2023、2024、2025
掌握方式：从非负变量逐步代入；每个等号条件写清

原题 501 · Q4

Suppose that for any i, x_i is such that $\mathbb{E}(x_i) = 3$ and $\mathrm{P}(x_i \geq 1) = 1$. Show that $\mathrm{P}(\sum_{i=1}^n \log x_i > 3n) \leq 1/3$. You may **not** assume anything beyond the conditions stated (e.g. no independence, no identical distribution).

Solution.

思路。不要先对随机变量取期望；要先在目标事件上对样本点 $x_1, \dots, x_n$ 用算术–几何平均不等式。事件中的对数和大于 $3n$，等价于几何平均大于 e^3；而算术平均不小于几何平均，所以该事件必然推出 $\bar{x}_n > e^3$。最后只需对非负的 $\bar{x}_n$ 用 Markov。整个论证只用 $x_i \geq 1$、$\mathbb{E}(x_i) = 3$ 与期望线性，不需要独立或同分布。方法见 2。

Let

$$A_n := \left\{ \sum_{i=1}^n \log x_i > 3n \right\}.$$

Step 1 (rewrite the event as a statement about the geometric mean). On A_n,

$$\frac{1}{n} \sum_{i=1}^n \log x_i > 3.$$

Exponentiation is strictly increasing, so

$$\exp\left(\frac{1}{n} \sum_{i=1}^n \log x_i \right) > e^3.$$

Because $x_i \geq 1 > 0$ almost surely, every logarithm is well-defined, and

$$\exp\left(\frac{1}{n} \sum_{i=1}^n \log x_i \right) = \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

Thus, on A_n, the geometric mean of $x_1, \dots, x_n$ exceeds e^3.

Step 2 (AM–GM gives an event inclusion). For every realization with positive x_i, the arithmetic–geometric mean inequality gives

$$\bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i \;\geq\; \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

Consequently,

$$A_n \subseteq \{\bar{x}_n > e^3\}.$$

This is a deterministic inequality applied pointwise; it makes no probabilistic independence assumption.

Step 3 (Markov on the arithmetic mean). Since $x_i \geq 1$, $\bar{x}_n \geq 0$ almost surely. Moreover, by linearity of expectation,

$$\mathbb{E}(\bar{x}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(x_i) = \frac{1}{n}(n \cdot 3) = 3,$$

again without independence. Therefore Markov's inequality and the event inclusion yield

$$\mathrm{P}(A_n) \leq \mathrm{P}\big(\bar{x}_n > e^3\big) \leq \frac{\mathbb{E}(\bar{x}_n)}{e^3} = \frac{3}{e^3}.$$

Finally, $e = \sum_{k=0}^{\infty} 1/k! > 1 + 1 + \frac{1}{2} + \frac{1}{6} = \frac{8}{3}$, hence

$$e^3 > \left(\frac{8}{3}\right)^3 = \frac{512}{27} > 9 \quad\Longrightarrow\quad \frac{3}{e^3} < \frac{3}{9} = \frac{1}{3}.$$

We have actually proved the stronger bound

$$\boxed{\mathrm{P}\big(\textstyle\sum_{i=1}^{n} \log x_i > 3n\big) \;\leq\; \dfrac{3}{e^3} \;<\; \dfrac{1}{3}.}$$

判分人想看的链条： $\sum_i \log x_i > 3n \Rightarrow (\prod_i x_i)^{1/n} > e^3 \Rightarrow \bar{x}_n > e^3$（AM–GM），再由 $\mathbb{E}(\bar{x}_n) = 3$ 与 Markov 得 $\mathrm{P}(()\,A_n) \leq 3/e^3 < 1/3$。每个箭头都不需要独立性。

Remark (Q4 的关键辨析：先搬动事件，而不是先取期望).

若先用 $\mathbb{E}(\log x_i) \leq \log 3$，再对 $\sum_i \log x_i$ 用 Markov，只能得到 $\log 3/3 \approx 0.366 > 1/3$，这个界不够强。正确突破口是把 Jensen/AM–GM *逐样本地*用于事件本身：$n^{-1}\sum_i \log x_i \leq \log \bar{x}_n$，于是目标事件包含在 $\{\bar{x}_n > e^3\}$ 中，再对 $\bar{x}_n$ 用 Markov。也不要直接对 $\prod_i x_i$ 用 Markov，因为在没有独立性时不能把 $\mathbb{E}(\prod_i x_i)$ 拆成 $\prod_i \mathbb{E}(x_i)$。见 2。

9.5 501 · Q5（指数分布累积量 + 条件 MLE）

归属： ECON 501（Pinkse） **重要度：** 高 **再考判断：** 高
核心考点： 指数分布累积量与条件 MLE
历史复现： 指数分布见 2020/2021/2025；MLE 多年高频
掌握方式： 累积量公式可背；条件似然需现场因子分解

原题 501 · Q5

Let $\{u_i\}$ be i.i.d. standard exponential (i.e. u_i has density $\exp(-u)\,\mathbb{1}(u \geq 0)$).

(a) Determine the cumulants of u_i.

(b) Suppose further that $y_i = \theta_0 x_i u_i$ where x_i has bounded support on the positive halfline and $\theta_0 > 0$. Determine the (conditional) maximum likelihood estimator $\hat{\theta}$ of θ_0 if the u_i's are independent of the x_i's.

Solution.

思路。(a) 累积量从累积量母函数 $K(t) = \log M(t)$ 读系数，标准指数的 $M(t) = (1-t)^{-1}$，$K(t) = -\log(1-t)$ 的泰勒系数直接给出 $\kappa_k = (k-1)!$。(b) 给定 x_i，y_i 是均值 $\theta_0 x_i$ 的指数分布；写条件对数似然、对 θ 求 FOC。累积量见 2，条件 MLE 见 5。

(a) Cumulants. The MGF of a standard exponential is

$$M(t) = \mathbb{E}\big(e^{tu}\big) = \int_0^\infty e^{tu} e^{-u}\, du = \frac{1}{1-t}, \qquad t < 1.$$

The cumulant generating function is

$$K(t) = \log M(t) = -\log(1-t) = \sum_{k=1}^{\infty} \frac{t^k}{k} = \sum_{k=1}^{\infty} \frac{(k-1)!}{k!}\, t^k.$$

The k-th cumulant is $k!$ times the coefficient of t^k:

$$\boxed{\kappa_k = k! \cdot \frac{1}{k} = (k-1)!, \qquad k = 1, 2, 3, \ldots}$$

so $\kappa_1 = 1$ (mean), $\kappa_2 = 1$ (variance), $\kappa_3 = 2$, $\kappa_4 = 6$, etc. 校验：$\kappa_1 = \mathbb{E}(u) = 1$、$\kappa_2 = \mathrm{Var}(u) = 1$（指数 Exp(1) 均值方差皆 1），对。三阶累积量 $= \mathbb{E}\big((u-1)^3\big) = 2$，对。

(b) Conditional MLE. Given x_i, since $u_i = y_i/(\theta_0 x_i) \sim \mathrm{Exp}(1)$ and $x_i, \theta_0 > 0$, the variable $y_i \mid x_i$ is exponential with mean $\theta_0 x_i$:

$$f(y_i \mid x_i; \theta) = \frac{1}{\theta x_i} \exp\!\left(-\frac{y_i}{\theta x_i}\right) \mathbb{1}\{y_i \geq 0\}.$$

用变量替换 $u = y/(\theta x)$，$du/dy = 1/(\theta x)$，密度 $e^{-u} \cdot |du/dy| = \frac{1}{\theta x} e^{-y/(\theta x)}$。 The conditional log-likelihood (treating x_i as given; $u_i \perp x_i$ makes the x-marginal free of θ, so it drops out) is

$$\ell_n(\theta) = \sum_{i=1}^{n} \left[-\log(\theta x_i) - \frac{y_i}{\theta x_i}\right] = -n \log\theta - \sum_{i=1}^{n} \log x_i - \frac{1}{\theta} \sum_{i=1}^{n} \frac{y_i}{x_i}.$$

First-order condition:

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}\frac{y_i}{x_i} = 0 \implies \theta = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}.$$

The second derivative at this point is $\frac{n}{\theta^2} - \frac{2}{\theta^3}\sum\frac{y_i}{x_i} = \frac{n}{\theta^2} - \frac{2n}{\theta^2} = -\frac{n}{\theta^2} < 0$, a maximum. Hence

$$\boxed{\hat\theta = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}}$$

条件 MLE $=$ y_i/x_i 的样本均值。直觉：$y_i/x_i = \theta_0 u_i$，是*无偏*的尺度信号（$\mathbb{E}(y_i/x_i \mid x_i) = \theta_0\mathbb{E}(u_i) = \theta_0$），其样本均值自然估 θ_0。$x_i \perp u_i$ 保证 x 的边际不含 θ，所以条件似然 $=$ 完整似然在 θ 上的有效部分。

Remark (Q5 累积量与 MLE 的连贯).

(a) 给出 $\kappa_1 = \mathbb{E}(u) = 1$ 正是 (b) 里 $\mathbb{E}(y_i/x_i \mid x_i) = \theta_0 \cdot 1 = \theta_0$ 的根据——条件 MLE 无偏。$\hat\theta$ 也是*矩估计*（用 $\mathbb{E}(u) = 1$），两法在此重合，因为指数尺度族里样本均值就是充分统计量的 MLE。见 5。

9.6 501 · Q6（测度论：$[0,1]$ 上分段函数的可测性）

归属：ECON 501（Pinkse） **重要度**：高 **再考判断**：高
核心考点：分段函数的 Borel 可测性
历史复现：测度论见 2017、2018、2022、2023、2024、2025
掌握方式：用阈值逆像生成 Borel；逐段处理边界

原题 501 · Q6

Consider the probability space $(\Omega, \mathcal{B}, \lambda)$ with $\Omega = [0,1]$, $\mathcal{B}$ the Borel field, and λ the Lebesgue measure. Consider the function $x : \Omega \to \mathbb{R}$ with

$$x(\omega) = \begin{cases} \sqrt{\omega}, & \text{if } \omega \text{ is rational,} \\ 1-\omega^2, & \text{if } \omega \text{ is irrational.} \end{cases}$$

Is $x(\omega)$ a random variable? Show that it is (isn't).

Solution.

思路。是随机变量（Borel 可测）。关键：$\mathbb{Q} \cap [0,1]$ 与其补集（无理数）都是 Borel 集（$\mathbb{Q}$ 可数 $\Rightarrow$ Borel）；在每一片上 x 都是*连续函数*（$\sqrt{\omega}$、$1-\omega^2$）故 Borel；分段定义在两个 Borel 片上、各用 Borel 函数，整体 Borel 可测。用 1 的「好集法 / 生成元上验证」即可，但这里更直接：把 $\{x \le t\}$ 拆成两片各自的 Borel 原像之并。

Yes, x is a random variable (Borel measurable).

The two pieces are Borel. Let $A := \mathbb{Q} \cap [0,1]$. Since $\mathbb{Q}$ is countable, $A = \bigcup_{q\in A}\{q\}$ is a countable union of singletons (each closed, hence Borel), so $A \in \mathcal{B}$. Its complement $A^c = [0,1] \setminus A$ (the irrationals) is also Borel.

Reduce to two restrictions. For any $t \in \mathbb{R}$,

$$\{\omega : x(\omega) \le t\} = \underbrace{\left(A \cap \{\omega : \sqrt{\omega} \le t\}\right)}_{\text{rational piece}} \cup \underbrace{\left(A^c \cap \{\omega : 1-\omega^2 \le t\}\right)}_{\text{irrational piece}}.$$

两片的「内层」集合各是一个连续函数的下水平集，故是 Borel；与 Borel 集 A 或 A^c 取交仍 Borel；两 Borel 集之并 Borel。The maps $g(\omega) = \sqrt{\omega}$ and $h(\omega) = 1-\omega^2$ are continuous on $[0,1]$, hence Borel measurable, so $\{g \le t\}, \{h \le t\} \in \mathcal{B}$. Therefore each of the two pieces is an intersection of two Borel sets (Borel), and their union is Borel:

$$\{x \le t\} \in \mathcal{B} \qquad \text{for every } t \in \mathbb{R}.$$

Since the sub-level sets $\{x \le t\}$ generate measurability and all lie in $\mathcal{B}$, x is $\mathcal{B}/\mathcal{B}(\mathbb{R})$-measurable:

$$\boxed{x \text{ is a (Borel) random variable.}}$$

一般原理：若 $\Omega = \bigcup_k B_k$ 是可数个 Borel 集的（不交）并，且 x 在每个 B_k 上等于一个 Borel 函数 x_k，则 x Borel 可测，因为 $\{x \le t\} = \bigcup_k (B_k \cap \{x_k \le t\})$ 是可数个 Borel 集之并。本题 $k = 2$（有理 / 无理两片），是该原理的最简情形。注意连续性几乎处处与可测性无关——即使 x 在每个有理点都跳变、处处不连续，它依然 Borel 可测。

Remark (Q6 可测 $\neq$ 连续).

这函数处处不连续（有理点 $\sqrt{\omega}$ 与邻近无理点 $1-\omega^2$ 一般不等），但仍是随机变量。可测性只要求原像落在 σ 域里，与连续/光滑无关。判分关键句：$\mathbb{Q}$ 可数 $\Rightarrow$ Borel；分段在 Borel 集上拼接 Borel 函数 $\Rightarrow$ Borel。见 1。

9.7 510 · Q1（识别三概念 + size 三概念）

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：点/部分/弱识别与 size 三概念
历史复现：识别见 2022–2025；size 见 2023–2025
掌握方式：定义逐字背准，尤其 sup 与 limsup 顺序

原题 510 · Q1

(a) Explain precisely what is meant by i) point identification, ii) partial identification, and iii) weak identification. Which concept implies which (if any)?

Provide meaningful examples.

(b) Define precisely what is meant for a test to i) control the limiting null rejection probability, ii) have correct size, and iii) have correct asymptotic size. Provide an example of empirical interest where a test controls the limiting null rejection probability but whose asymptotic size exceeds the nominal size.

Solution.

思路。两组概念都是 510 的核心定义，分别见 7 与 13。(a) 识别三连：点识别 = 识别集是单点；部分识别 = 识别集是真子集（非单点）；弱识别 = 每个固定 DGP 下点识别，但识别信息漂向零、标准渐近失真。(b) size 三连的关键是先 *sup* 还是先 *lim*：limiting NRP 是逐点极限；correct size 是有限样本 sup；correct asymptotic size 是 $\limsup_n \sup_{\text{null}}$（sup 在前、lim 在后，要 uniform）。

(a) Identification concepts. Let P denote the DGP and $\Theta_I(P) \subseteq \Theta$ the identified set (all parameter values observationally equivalent to the truth under P).

i) *Point identification*: $\Theta_I(P) = \{\theta_0\}$ is a singleton —the data distribution pins θ_0 down uniquely.

ii) *Partial identification*: $\Theta_I(P)$ is a proper subset of Θ that is *not* a singleton — the data restrict θ_0 to a set but cannot pin it to a point.

iii) *Weak identification*: θ_0 *is* point identified for each fixed P, but identification is "near failure" —formally, along drifting-parameter (local-to-non-identification) sequences the information about θ_0 shrinks to zero, so standard normal/χ^2 asymptotics are poor approximations.

Implications. Under the definitions just given, point identification and partial identification are *mutually exclusive*: a singleton is not a non-singleton, so neither concept implies the other. (Some authors use "set identification" broadly enough to include singleton sets; under that broader vocabulary point identification is a special case of set identification, but it is *not* a special case of *partial* identification as defined here.) Weak identification is not a third identified-set shape; it describes the *strength* of identification along sequences of DGPs. In the standard weak-IV usage, every fixed DGP with a nonzero first stage is point identified, while a sequence of first stages approaches the unidentified boundary. Thus weak identification entails point identification at each fixed nonboundary DGP, but its limit experiment may be unidentified. It neither implies nor is implied by partial identification. 蕴含关系一句话：按本题的非单点定义，点识别与部分识别互斥、彼此不蕴含；弱识别是局部序列意义的「识别强度趋零」，通常逐个非边界 DGP 仍

点识别，但极限可不识别；弱识别与部分识别无蕴含。见 7。 **Examples.**

Point: $y = x'\theta_0 + \varepsilon$, $\mathbb{E}(x\varepsilon) = 0$, $\mathbb{E}\big(xx'\big)$ full rank —OLS pins θ_0.

Partial: interval-censored outcome $y \in [y_L, y_U] \Rightarrow \mathbb{E}(y \mid x) \in [\mathbb{E}(y_L \mid x), \mathbb{E}(y_U \mid x)]$.

Weak: IV $y = \beta w + \varepsilon$, $w = \pi z + v$ with $\pi = c/\sqrt{n}$ (near-zero first stage).

(b) Size concepts. Let $\mathrm{RP}_n(P) = \mathrm{P}_{H_0}(\text{reject})$ be the null rejection probability at sample size n under null DGP $P \in \mathcal{P}_0$.

i) *Controls the limiting null rejection probability*: for *each fixed* $P \in \mathcal{P}_0$, $\limsup_{n\to\infty} \mathrm{RP}_n(P) \le \alpha$ (a *pointwise* statement).

ii) *Correct (finite-sample) size*: $\sup_{P\in\mathcal{P}_0} \mathrm{RP}_n(P) \le \alpha$ for *all* n (sup over the null, exact, every sample size).

iii) *Correct asymptotic size*: $\mathrm{AsySz} := \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} \mathrm{RP}_n(P) \le \alpha$ —the *sup-before-lim* (uniform) version. This dominates (i): always $\sup_P \limsup_n \le \limsup_n \sup_P$.

(i) 与 (iii) 的差别全在 sup 与 lim 的顺序：(i) 先固定 P 取 n-极限（逐点）；(iii) 每个 n 先在 null 上 sup、再取极限（uniform）。$\sup_P \limsup_n \le \limsup_n \sup_P$，故 (iii) 比 (i) 强。两者可严格不等——这正是下面例子的点。 **Example (limiting NRP controlled, asymptotic size $> \alpha$).** *Weak-IV t-test* (or near-unit-root / parameter-on-boundary). Consider the IV t-statistic for β with first stage $w = \pi z + v$. For each *fixed* $\pi \ne 0$, as $n \to \infty$ the t-statistic is asymptotically $N(0,1)$, so the limiting NRP is exactly α —(i) holds. But the convergence is *not uniform* in π near 0: along the drifting sequence $\pi_n = c/\sqrt{n}$, the t-statistic has a non-normal (weak-IV) limit whose rejection probability can far exceed α. Taking the sup over π (including the worst drifting sequence) before the limit,

$$\mathrm{AsySz} = \lim_n \sup_\pi \sup \mathrm{RP}_n > \alpha.$$

$$\boxed{\text{Weak-IV } t\text{-test: limiting NRP} = \alpha \text{ (each fixed } \pi \ne 0\text{), but AsySz} > \alpha \text{ (uniformity fails near } \pi = 0\text{).}}$$

这就是 510 反复强调的「逐点对、uniform 错」：固定 π 的 textbook 渐近骗你检验是 level-α，但 π 可以随 n 漂向弱识别区，sup 把最坏漂移序列抓出来，asymptotic size 被顶到 α 以上。同理出现在 unit-root、边界参数、moment inequality。见 13、14。

Remark (Q1 顺序就是一切).

两组概念的灵魂都在「顺序」：识别上，弱识别是「点识别但信息 $\to 0$」；size 上，asymptotic size 是「sup 在前、lim 在后」。Guggenberger 的整套 uniformity 方法（13）就是为了让 $\limsup_n \sup_P \mathrm{RP}_n \le \alpha$，而不仅是逐点。

9.8 510 · Q2（非光滑 GMM：V_0, Γ 估计与分位回归）

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：非光滑 GMM 的方差估计与 QR 特化
历史复现：2023、2024、2025 连续同构
掌握方式：三年 Q2 一起练到可闭卷完整写出

原题 510 · Q2

Under Assumptions EE3 and CF-NS the limiting distribution of the GMM estimator $\hat\theta_n$ with nonsmooth stochastic criterion function $Q_n(\theta) = \|\bar g_n(\theta)\|$ (with $\bar g_n(\theta) = n^{-1}\sum_{i=1}^n g(W_i,\theta)$, $g(\theta) = \mathbb{E}(g(W_i,\theta))$) is

$$\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{d} N\big(0,\ (\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1}\big)\,.$$

CF-NS: (i) θ_0 interior; (ii) $g(\theta)$ differentiable at θ_0 with $\Gamma = (\partial/\partial\theta')g(\theta_0)$ full rank $d \le k$; (iii) $g(\theta_0) = 0$; (iv) $\sqrt{n}\bar g_n(\theta_0) \xrightarrow{d} N(0, V_0)$; plus a stochastic-equicontinuity sup condition over shrinking neighborhoods. EE3: $Q_n(\hat\theta_n) = \inf_\theta Q_n(\theta) + o_p(n^{-1/2})$ and $\hat\theta_n \xrightarrow{p} \theta_0$. 注：原卷把两条假设都标成 (iii)（$g(\theta_0) = 0$ 与 $\sqrt{n}\bar g_n \xrightarrow{d} N(0, V_0)$），并把 sup/随机等度连续条件单列；此处按 (i)–(iv) 重新顺号、把 sup 条件并入，便于与 PDF 第 3 页逐条对照，内容不变。

(a) Provide estimators for V_0 and Γ and discuss their consistency under appropriate conditions.

(b) Consider Quantile Regression, $(Y_i, X_i')'$ i.i.d., $Y_i = X_i'\theta_0 + U_i$, where the τ-quantile of U_i conditional on X_i equals 0 for some $\tau \in (0,1)$. Derive what V_0 and Γ are in this case.

(c) Find a simple expression for $(\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1}$ in the quantile-regression case.

Solution.

思路。整题就是 11 的 worked example，直接对齐。(a) V_0 是 $\sqrt{n}\bar g_n(\theta_0)$ 的协方差，用样本二阶矩在 $\hat\theta_n$ 处估即可；Γ 是总体雅可比，因准则非光滑不能用 $g(W_i,\cdot)$ 的样本导数，得用数值微分（Powell 型带宽 ε_n）或已知闭式。(b)(c) 分位回归：矩 $g = X(\tau - \mathbb{1}\{Y \le X'\theta\})$，$V_0 = \tau(1-\tau)\mathbb{E}(XX')$、$\Gamma = -\mathbb{E}\big(f_{U|X}(0)XX'\big)$，且 $d = k$ 时夹心塌缩。

(a) Estimators of V_0 and Γ.

V_0: Since $V_0 = \mathrm{Var}(g(W_1,\theta_0)) = \mathbb{E}(g(W_1,\theta_0)g(W_1,\theta_0)')$ (because $g(\theta_0) = \mathbb{E}(g(W_1,\theta_0)) = 0$), estimate by the sample second moment at $\hat\theta_n$:

$$\boxed{\hat V_0 = \frac{1}{n}\sum_{i=1}^n g(W_i,\hat\theta_n)\, g(W_i,\hat\theta_n)'}$$

Consistency: by a uniform LLN over a neighborhood of θ_0 plus $\hat{\theta}_n \xrightarrow{p} \theta_0$ and continuity of $\theta \mapsto \mathbb{E}(g(W_1,\theta)g(W_1,\theta)')$ at θ_0, $\hat{V}_0 \xrightarrow{p} V_0$. (Each $g(W_i,\cdot)$ may be discontinuous in θ, but its *second moment* is continuous in θ at θ_0 under CF-NS, which is what the ULLN needs.)

Γ*:* The map $\theta \mapsto g(W_i,\theta)$ is *nonsmooth* (e.g. indicator), so the sample average of pathwise derivatives is unavailable/zero a.e. Instead $\Gamma = \partial g(\theta_0)/\partial\theta'$ is the Jacobian of the *smooth population* moment $g(\theta) = \mathbb{E}(g(W_1,\theta))$. Estimate by a *numerical (finite) difference* of $\bar{g}_n$ with bandwidth ε_n:

$$\boxed{\hat{\Gamma}_{\cdot j} = \frac{\bar{g}_n(\hat{\theta}_n + \varepsilon_n e_j) - \bar{g}_n(\hat{\theta}_n - \varepsilon_n e_j)}{2\varepsilon_n}, \qquad j = 1,\dots,d}$$

where e_j is the j-th unit vector. *Consistency:* Lecture 16 takes $\varepsilon_n \to 0$ and $1/(\sqrt{n}\varepsilon_n) = O_p(1)$. Under CF-NS(v), the local empirical numerator is $o_p(n^{-1/2})$, so the amplified remainder is still $o_p(1)$; differentiability and $\hat{\theta}_n \xrightarrow{p} \theta_0$ then give $\hat{\Gamma} \xrightarrow{p} \Gamma$. The stronger convenient choice $\sqrt{n}\varepsilon_n \to \infty$ also suffices. (If Γ has a known closed form —as in QR below —use it directly with a plug-in density estimate.) Γ 不能用 $\frac{1}{n}\sum \partial_\theta g(W_i,\hat{\theta}_n)$，因为非光滑（指标函数导数几乎处处为 0）。数值微分的两难是：$\varepsilon_n \to 0$ 才瞄准 local derivative，但 step 太小会放大 empirical remainder。可背 $\varepsilon_n = n^{-a}$、$0 < a < 1/2$ 作为安全选择。见 11。

(b) Quantile regression: V_0 and Γ. The check-function FOC gives the moment

$$g(W_i,\theta) = X_i\left(\tau - \mathbb{1}\{Y_i \le X_i'\theta\}\right), \qquad W_i = (Y_i, X_i').$$

V_0*:* At θ_0, the conditional τ-quantile restriction gives $\mathrm{P}(Y_i \le X_i'\theta_0 \mid X_i) = \tau$, so $\mathbb{1}\{Y_i \le X_i'\theta_0\} \mid X_i \sim \text{Bernoulli}(\tau)$, whence $\mathbb{E}((\tau - \mathbb{1}\{\cdot\}) \mid X_i) = 0$ and $\mathrm{Var}(\tau - \mathbb{1}\{\cdot\} \mid X_i) = \tau(1-\tau)$. Therefore

$$V_0 = \mathbb{E}\big(gg'\big) = \mathbb{E}\big(X_iX_i'\,(\tau - \mathbb{1}\{Y_i \le X_i'\theta_0\})^2\big) = \mathbb{E}\big(X_iX_i'\,\mathbb{E}\big((\tau-\mathbb{1})^2 \mid X_i\big)\big)$$

$$\boxed{V_0 = \tau(1-\tau)\,\mathbb{E}\big(X_iX_i'\big)}$$

Γ*:* Differentiate the population moment $g(\theta) = \mathbb{E}(X_i(\tau - \mathbb{1}\{Y_i \le X_i'\theta\})) = \mathbb{E}\big(X_i(\tau - F_{U|X}(X_i'\theta - X_i'\theta_0 \mid X_i))\big)$. Using $\partial_\theta F_{U|X}(X_i'(\theta-\theta_0) \mid X_i) = f_{U|X}(X_i'(\theta-\theta_0) \mid X_i)X_i'$, at $\theta = \theta_0$,

$$\boxed{\Gamma = \frac{\partial g(\theta_0)}{\partial\theta'} = -\mathbb{E}\big(X_iX_i'\,f_{U|X}(0 \mid X_i)\big)}$$

(the sign is immaterial in the sandwich; many texts drop it or define g with opposite sign). Here $d = k = \dim(X_i)$. 非光滑的来源是指标函数；但期望把指标抹成条件 CDF $F_{U|X}$，对 θ 可导，导数落在 θ_0 处即条件密度 $f_{U|X}(0 \mid X)$。这正是 Γ 里冒出条件密度的原因。

(c) Simplification. In QR, Γ is $k\times k$ and full rank, hence *square and invertible.* The sandwich collapses: with Γ square, $(\Gamma'\Gamma)^{-1}\Gamma' = \Gamma^{-1}$, so

$$(\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1} = \Gamma^{-1} V_0\, \Gamma^{-1\prime}.$$

Plugging in $V_0 = \tau(1-\tau)\mathbb{E}(XX')$ and defining $J := \mathbb{E}\big(f_{U|X}(0\mid X)XX'\big) = -\Gamma$ (the sign cancels),

$$\boxed{(\Gamma'\Gamma)^{-1}\Gamma' V_0 \Gamma(\Gamma'\Gamma)^{-1} = \tau(1-\tau)\, J^{-1}\mathbb{E}\big(XX'\big)\, J^{-1}}$$

Under $f_{U|X}(0\mid X) \equiv f$, $J = f\,\mathbb{E}(XX')$ and it reduces further to

$$\frac{\tau(1-\tau)}{f^2}\,\big(\mathbb{E}\big(XX'\big)\big)^{-1}.$$

这是分位回归的经典渐近方差 Koenker–Bassett 形式。$d = k$ 使「广义」夹心退化成 $\Gamma^{-1}V_0\Gamma^{-1\prime}$；同方差时 $J = f\mathbb{E}(XX') = -\Gamma$，两个 $\mathbb{E}(XX')$ 约掉一个，留 $\frac{\tau(1-\tau)}{f^2}(\mathbb{E}(XX'))^{-1}$。见 11。

Remark (Q2 非光滑的两处关键).

两处「非光滑被期望救活」：(1) Γ 不能样本求导，但总体 $g(\theta)$ 是光滑的（指标的期望 = CDF），导数处冒出条件密度；(2) V_0 反而不受非光滑影响（指标平方 = 指标，Bernoulli 方差 $\tau(1-\tau)$）。估计时 V_0 用样本二阶矩、Γ 用数值微分或核密度插值。见 11。

9.9 510 · Q3（五个正态均值检验：临界值、功效、无 UMP）

归属：ECON 510（Guggenberger） **重要度**：高（必会） **再考判断**：高
核心考点：五检验临界值、功效与无 UMP
历史复现：2018、2023、2025；NP/UMP 多年出现
掌握方式：临界值、功效函数和互相击败的逻辑都要会

原题 510 · Q3

Suppose $X_i \sim N(\mu_i, 1)$, $i = 1,\ldots,n$ independent. Consider five tests of H_0 : $\mu_1 = \cdots = \mu_n = 0$ vs H_1 : “not H_0” with acceptance regions

$$A_1 = \{\textstyle\sum_{i=1}^n x_i \le k_1\},\quad A_2 = \{|\textstyle\sum_{i=1}^n x_i| \le k_2\},\quad A_3 = \{x_i^2 \le k_3\ \forall i\},$$
$$A_4 = \{x_1^2 \le k_4\},\quad A_5 = \{\textstyle\sum_{i=1}^n x_i^2 \le k_5\},$$

where k_j (possibly depending on n) are chosen so each test has size α for some $\alpha \in (0,1)$.

(a) Determine k_j, $j = 1, \ldots, 5$.

(b) Let $n = 2$. Determine the power functions of the first and the fourth test.

(c) Show that none of the five tests is (weakly) more powerful than all the other tests considered here uniformly over the alternative space $\{(\mu_1, \mu_2) : -\infty < \mu_1, \mu_2 < \infty\}$ $(n = 2)$.

Solution.

思路。(a) 在 H_0 下各统计量的零分布是已知的（正态/卡方/独立卡方），临界值由 size $= \alpha$ 反解。拒绝是落在 A_j 之外。(b) $n = 2$ 时把功效写成非中心化的正态尾。(c) 每个检验在某方向最优、在另一方向被打败 $\Rightarrow$ 无一致最优。理论框架见 6。

(a) Critical constants (reject when outside A_j). Under H_0:

- *Test 1* (reject if $\sum x_i > k_1$): $\sum x_i \sim N(0, n)$, one-sided, so $k_1 = \sqrt{n}\, z_{1-\alpha}$.
- *Test 2* (reject if $|\sum x_i| > k_2$): two-sided, $k_2 = \sqrt{n}\, z_{1-\alpha/2}$.
- *Test 3* (reject if $\max_i x_i^2 > k_3$): the $x_i^2 \sim \chi_1^2$ are independent, so $\mathrm{P}(\text{accept}) = \left(\mathrm{P}\left(\chi_1^2 \le k_3\right)\right)^n = 1 - \alpha$, giving $k_3 = q_{\chi_1^2}\left((1-\alpha)^{1/n}\right)$ (Šidák).
- *Test 4* (reject if $x_1^2 > k_4$): $x_1^2 \sim \chi_1^2$, so $k_4 = q_{\chi_1^2}(1-\alpha) = z_{1-\alpha/2}^2$.
- *Test 5* (reject if $\sum x_i^2 > k_5$): $\sum x_i^2 \sim \chi_n^2$, so $k_5 = q_{\chi_n^2}(1-\alpha)$.

$$\boxed{k_1 = \sqrt{n} z_{1-\alpha}, \quad k_2 = \sqrt{n} z_{1-\alpha/2}, \quad k_3 = q_{\chi_1^2}\big((1-\alpha)^{1/n}\big), \quad k_4 = q_{\chi_1^2}(1-\alpha), \quad k_5 = q_{\chi_n^2}(1-\alpha)}$$

where $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ and $q_{\chi_m^2}(p)$ is the p-quantile of χ_m^2.

(b) Power functions, $n = 2$.

Test 1 ($\sum x_i > k_1$, $k_1 = \sqrt{2}\, z_{1-\alpha}$): under (μ_1, μ_2), $\sum x_i = x_1 + x_2 \sim N(\mu_1 + \mu_2, 2)$, so

$$\boxed{\beta_1(\mu_1, \mu_2) = \mathrm{P}(\textstyle\sum x_i > k_1) = 1 - \Phi\left(\dfrac{k_1 - (\mu_1 + \mu_2)}{\sqrt{2}}\right) = \Phi\left(\dfrac{\mu_1 + \mu_2}{\sqrt{2}} - z_{1-\alpha}\right)}$$

Test 4 ($x_1^2 > k_4$, $k_4 = z_{1-\alpha/2}^2$): $x_1 \sim N(\mu_1, 1)$, reject iff $|x_1| > \sqrt{k_4} = z_{1-\alpha/2}$, so

$$\boxed{\beta_4(\mu_1, \mu_2) = \mathrm{P}\big(|x_1| > z_{1-\alpha/2}\big) = 1 - \Phi\big(z_{1-\alpha/2} - \mu_1\big) + \Phi\big(-z_{1-\alpha/2} - \mu_1\big)}$$

which depends on μ_1 only (and equals α at $\mu_1 = 0$). β_1 只通过 $\mu_1 + \mu_2$ 进入（沿 $(1,1)$ 方向），β_4 只通过 μ_1（完全忽略 μ_2）。这两条单调/对称结构正是 (c) 的弹药。

(c) No one of the five tests uniformly dominates the other four. The clean proof uses the Neyman–Pearson lemma, so no unproved comparison of complicated power functions is needed.

First rule out Tests 2–5 as uniformly best. Fix any $c > 0$ and consider the *simple* alternative

$$H_{1,c} : (\mu_1, \mu_2) = (c, c).$$

Relative to $H_0 : (\mu_1, \mu_2) = (0, 0)$, the likelihood ratio is

$$\frac{f_{c,c}(x_1, x_2)}{f_{0,0}(x_1, x_2)} = \exp\{c(x_1 + x_2) - c^2\},$$

which is strictly increasing in $x_1 + x_2$. By the Neyman–Pearson lemma, the unique (up to null sets) most powerful size-α test against $H_{1,c}$ rejects for large $x_1 + x_2$. That is exactly Test 1, because its rejection region is

$$\{x_1 + x_2 > \sqrt{2}\, z_{1-\alpha}\}.$$

Each of Tests 2, 3, 4, and 5 has a rejection region that differs from this half-plane on a set of positive Lebesgue probability. Since the likelihood ratio has a continuous distribution, the strict part of the Neyman–Pearson lemma gives

$$\beta_1(c, c) > \beta_j(c, c), \qquad j = 2, 3, 4, 5.$$

Therefore none of Tests 2–5 can be weakly more powerful than Test 1 at *every* alternative.

Now rule out Test 1 as uniformly best. Take $(\mu_1, \mu_2) = (M, -M)$. Then $x_1 + x_2 \sim N(0, 2)$, exactly as under H_0, so Test 1 has power

$$\beta_1(M, -M) = \alpha \quad \text{for every } M.$$

But Test 4 rejects when $|x_1| > z_{1-\alpha/2}$ and $x_1 \sim N(M, 1)$, so

$$\beta_4(M, -M) \longrightarrow 1 \qquad (M \to \infty).$$

Thus for all sufficiently large M, $\beta_4(M, -M) > \beta_1(M, -M)$. Test 1 therefore cannot uniformly dominate Test 4.

We have exhibited a failure of uniform dominance for every candidate: Tests 2–5 lose to Test 1 at (c, c), while Test 1 loses to Test 4 at $(M, -M)$. Hence

$$\boxed{\text{No UMP test exists over } \{(\mu_1, \mu_2) \in \mathbb{R}^2\}\text{: each test is best in one direction, beaten in another.}}$$

深层原因：每个简单备择方向都有自己的 Neyman–Pearson 半空间。(c, c) 的最优方向是 $x_1 + x_2$，而 $(M, -M)$ 的信号与这个方向正交。二维备择不可能让同一个拒绝域同时对所有方向都是 NP 最优。上面的论证不仅给直觉，还逐一排除了五个候选。见 6。

Remark (Q3 无 UMP 的根).

一句话：UMP 要求所有备择点的 NP 最优检验**方向一致**；二维均值里 $(1,1)$ 方向与 $(1,0)$ 方向的最优拒绝域不同，故无 UMP。Test 1 **还有偏**（$\mu_1+\mu_2<0$ 时功效 $<\alpha$），连「无偏」都不满足。考场展示 2–3 个反向支配点即可。见 6。

9.10 510 · Q4（LASSO oracle 性质 + Ridge= 增广 OLS）

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：LASSO oracle 与 Ridge 增广 OLS
历史复现：2023、2025
掌握方式：oracle 条件需背；增广设计矩阵必须能现场构造

原题 510 · Q4

Consider $y = X\theta^* + w \in \mathbb{R}^n$ with $X \in \mathbb{R}^{n\times p}$; estimate θ^* from (y, X), possibly $p \gg n$.

(a) Define the LASSO estimator. What do we mean by the "oracle property" of LASSO?

(b) Present a set of high-level assumptions on θ^*, the penalty λ_n, X, and w such that the oracle property holds. Critically discuss these assumptions.

(c) Describe the finite-sample performance of LASSO relative to Ridge under various DGPs.

(d) Show that the Ridge estimator can be computed via OLS after augmenting $y \to (y', 0_p')'$ and $X \to (X', \sqrt{\lambda_n} I_p)'$.

Solution.

思路。全题对齐 15。(a) 写 LASSO 目标 + oracle 定义（选对支撑 + 非零系数渐近等于 oracle OLS）。(b) 列稀疏性、beta-min、restricted eigenvalue/irrepresentable、$\lambda_n \asymp \sqrt{(\log p)/n}$、sub-Gaussian 噪声，并批判（irrepresentable 强且不可检验；plain LASSO 有偏、真 oracle 常需 adaptive LASSO）。(c) 稀疏 LASSO 胜、稠密/共线 Ridge 稳。(d) 增广最小二乘代数恒等。

(a) LASSO and the oracle property.

$$\boxed{\hat\theta_{\text{LASSO}} = \arg\min_{\theta\in\mathbb{R}^p} \frac{1}{2n}\|y - X\theta\|_2^2 + \lambda_n \|\theta\|_1}$$

The ℓ_1 penalty induces *sparsity* (corner solutions set coordinates exactly to 0). Let $S = \{j : \theta_j^* \neq 0\}$ be the true support, $s = |S|$. The *oracle property* holds if, as $n \to \infty$ (with prob. $\to 1$),

(i) *Support/sign consistency*: $\mathrm{P}\left(\widehat{S} = S\right) \to 1$ (selects exactly the true nonzeros, correct signs);

(ii) *Oracle efficiency*: the estimates of the nonzero coefficients $\hat{\theta}_S$ have the same asymptotic distribution as the OLS estimator $\hat{\theta}_S^{\text{oracle}}$ that *knew* S in advance, i.e. $\sqrt{n}(\hat{\theta}_S - \theta_S^*) \xrightarrow{d} N(0, \Sigma_S)$ with Σ_S the oracle covariance.

oracle =「事后诸葛亮」: 仿佛有人提前告诉你哪些系数是 0,你只对真支撑做 OLS。oracle 性质 = 选得准(支撑)+ 估得好(非零部分像 oracle OLS 一样有效)。

(b) High-level assumptions + critique. 先区分两个不同结论:下面的高维条件可给 plain LASSO 预测率、估计率,外加更强条件时的支撑恢复;但它们不能自动给出"支撑恢复 + oracle OLS 极限分布"这一完整 oracle property。

(A1) *Sparsity*: $s = |S|$ small relative to sample, $s \log p/n \to 0$.

(A2) *Beta-min (minimum signal)*: $\min_{j\in S} \left|\theta_j^*\right| \geq c\sqrt{(\log p)/n}$ (nonzeros bounded away from 0 fast enough to be detected).

(A3) *Design*: restricted eigenvalue / compatibility condition on X (for ℓ_2 rates), and the *irrepresentable (incoherence) condition* $\left\|X_{S^c}' X_S (X_S' X_S)^{-1}\right\|_\infty < 1$ for *sign/support* recovery.

(A4) *Penalty*: for high-dimensional rates and support recovery, plain LASSO typically uses $\lambda_n \asymp \sqrt{(\log p)/n}$ (large enough to dominate the largest noise score). For the *full* oracle property, replace it by an adaptive penalty $\lambda_n \sum_j \hat{w}_j |\theta_j|$ or refit OLS after consistent selection.

(A5) *Noise*: w_i i.i.d. (or independent) sub-Gaussian, mean 0, variance σ^2.

Why plain LASSO cannot have the full oracle property under the usual regular fixed-p asymptotics. Let $Q = \operatorname{plim} X'X/n$ be positive definite and suppose the noise score has a nondegenerate $n^{-1/2}$ limit. On the event $\widehat{S} = S$, the KKT equation on the active coordinates gives

$$\sqrt{n}(\hat{\theta}_S - \theta_S^*) = Q_{SS}^{-1}\left(\frac{X_S' w}{\sqrt{n}} - \sqrt{n}\lambda_n \operatorname{sgn}(\theta_S^*)\right) + o_p(1).$$

For this to have the *same centered limit* as oracle OLS, the penalty bias must vanish, hence

$$\sqrt{n}\lambda_n \longrightarrow 0. \qquad \text{(oracle efficiency)}$$

For an inactive coordinate to be set exactly to zero with probability tending to one, however, its KKT inequality must dominate a random score of order $n^{-1/2}$:

$$\left|\frac{1}{n} X_j'(y - X\hat{\theta})\right| \leq \lambda_n, \qquad j \notin S.$$

Under a nondegenerate score (and the usual irrepresentable condition), selection consistency therefore requires

$$\sqrt{n}\lambda_n \longrightarrow \infty. \qquad \text{(kill inactive noise)}$$

The two displayed requirements contradict each other. Thus there is *no* regular tuning sequence for plain LASSO that simultaneously delivers exact selection and the centered oracle-OLS limit.

How adaptive LASSO removes the conflict. Take $\hat{w}_j = 1/\left|\hat{\theta}_j^{\text{init}}\right|^{\gamma}$ with a root-n consistent initial estimator and $\gamma > 0$. Then $\hat{w}_j = O_p(1)$ on S, while $\hat{w}_j = O_p(n^{\gamma/2})$ on S^c. With our normalized objective, the compatible conditions

$$\sqrt{n}\lambda_n \to 0, \qquad n^{(\gamma+1)/2}\lambda_n \to \infty$$

make the penalty negligible on true signals but overwhelming on zero coordinates. Together with beta-min, design, and noise regularity, adaptive LASSO obtains both support consistency and the oracle limit. Post-LASSO OLS gives the same first-order conclusion if the first-stage support estimate is consistent.

Critical discussion. The irrepresentable condition is strong, essentially uncheckable, and easily violated when relevant and irrelevant regressors are highly correlated. Beta-min rules out local-to-zero signals—the very cases where model selection is intrinsically unstable. Consequently the oracle statement is pointwise and nonuniform near the model-selection boundary; it should not be read as a guarantee of uniformly valid inference after selection. 判分重点：先定义 oracle 的两部分，再写出 plain LASSO 的调参矛盾：杀掉零坐标要 $\sqrt{n}\lambda_n \to \infty$，非零坐标像 oracle OLS 一样无偏却要 $\sqrt{n}\lambda_n \to 0$。所以严格 oracle 通常指 adaptive LASSO（或一致选模后的 post-LASSO OLS），不能把普通 LASSO 的 prediction/selection 结果偷换成 oracle efficiency。irrepresentable 与 beta-min 也都是很强、且导致非一致性的假设。见 15。

(c) LASSO vs Ridge in finite samples.

- *Sparse DGP* (few large nonzeros, many exact zeros): LASSO dominates —it sets irrelevant coefficients to 0, giving lower MSE and interpretable selection; Ridge only shrinks, never zeros, keeping noise from all p regressors.
- *Dense DGP* (many small coefficients of similar magnitude): Ridge dominates —shrinking everything proportionally beats LASSO's all-or-nothing selection, which kills true small signals.
- *Highly correlated regressors*: Ridge is more stable (handles collinearity gracefully, shares weight across correlated predictors); LASSO arbitrarily picks one of a correlated group and zeros the rest, making selection unstable across samples. (Elastic net interpolates.)

口诀：稀疏选 LASSO、稠密选 Ridge、共线 Ridge 稳。MSE 权衡取决于 DGP 的稀疏度与相关结构。

(d) Ridge as augmented OLS. The Ridge estimator solves

$$\hat{\theta}_{\text{R}} = \arg\min_{\theta} \ \|y - X\theta\|_2^2 + \lambda_n \|\theta\|_2^2 = (X'X + \lambda_n I_p)^{-1} X'y.$$

Augment $\tilde{y} = \binom{y}{0_p} \in \mathbb{R}^{n+p}$ and $\tilde{X} = \binom{X}{\sqrt{\lambda_n}\, I_p} \in \mathbb{R}^{(n+p)\times p}$. The augmented OLS objective is

$$\left\|\tilde{y} - \tilde{X}\theta\right\|_2^2 = \|y - X\theta\|_2^2 + \left\|0_p - \sqrt{\lambda_n}\, I_p\theta\right\|_2^2 = \|y - X\theta\|_2^2 + \lambda_n \|\theta\|_2^2,$$

identical to the Ridge objective. Hence the OLS estimator on $(\tilde{y}, \tilde{X})$ is

$$(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} = \left(X'X + \lambda_n I_p\right)^{-1}\left(X'y + \sqrt{\lambda_n} I_p \cdot 0_p\right) = (X'X + \lambda_n I_p)^{-1}X'y.$$

$$\boxed{\hat{\theta}_{\mathrm{R}} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \;\text{ with }\; \tilde{X} = \begin{pmatrix} X \\ \sqrt{\lambda_n} I_p \end{pmatrix},\ \tilde{y} = \begin{pmatrix} y \\ 0_p \end{pmatrix}}$$

增广行 $\sqrt{\lambda_n} I_p$ 配上响应 0，就是「把每个 θ_j 往 0 拉、惩罚力度 λ_n」的伪观测。代数上 $\tilde{X}'\tilde{X} = X'X + \lambda_n I_p$、$\tilde{X}'\tilde{y} = X'y$，正好是 Ridge 的正规方程。题面 $\sqrt{\lambda_n} I_p$ 的根号不可省，否则惩罚变成 λ_n^2。见 15。

Remark (Q4 plain LASSO 不 oracle).

最易丢分处：把「LASSO 有 oracle 性质」当成无条件真理。实际上 plain LASSO 因 ℓ_1 对大系数的常数偏置一般不同时满足选择一致 + oracle 效率；Zou(2006) 的 adaptive LASSO（加权惩罚）才补上。irrepresentable 条件是支撑恢复的强假设。见 15。

9.11 510 · Q5（Heckit 选择模型：逆 Mills、识别）

归属：ECON 510（Guggenberger） **重要度**：极高（必会） **再考判断**：极高
核心考点：Heckit 逆 Mills 与两阶段识别
历史复现：2023 选择识别、2024、2025
掌握方式：probit–IMR–outcome equation 三步形成固定模板

原题 510 · Q5

In the Heckit model (Heckman, 1979), with $y^* = x'\theta + \varepsilon$, $y = d\,y^*$, $d = \mathbb{1}(x'\pi_1 + z'\pi_2 \geq -\eta)$, and joint normality of (ε, η):

(a) Show that $\mathbb{E}(y \mid x, z, d = 1) = x'\theta + \sigma_{\varepsilon\eta}\,\lambda(x'\pi_1 + z'\pi_2)$, where $\lambda(s) = \phi(s)/\Phi(s)$.

(b) Identify π_1 and π_2 from a probit regression of d onto x and z.

(c) Specify conditions under which one can identify θ and $\sigma_{\varepsilon\eta}$.

Solution.

思路。Heckit 三连，全在 7 的 worked example。(a) 选择 $d = 1 \Leftrightarrow \eta \geq -(x'\pi_1 + z'\pi_2)$，对选中样本 $\mathbb{E}(y \mid d = 1) = x'\theta + \mathbb{E}(\varepsilon \mid \eta \geq -s)$；用二元正态 $\mathbb{E}(\varepsilon \mid \eta) = \sigma_{\varepsilon\eta}\eta$

（$\mathrm{Var}(\eta)=1$）与截断正态均值 $\mathbb{E}(\eta \mid \eta \ge -s)=\lambda(s)$ 得逆 Mills 修正项。(b) probit 直接给 (π_1, π_2)。(c) 排他性约束（z 中有变量不进 x）保证 λ 不与 x 共线。

(a) Conditional mean with inverse Mills ratio. Normalize $\mathrm{Var}(\eta)=1$ (the probit scale normalization). Let $s := x'\pi_1 + z'\pi_2$. The selection rule is

$$d = 1 \iff \eta \ge -s.$$

For the selected subsample,

$$\mathbb{E}(y \mid x, z, d=1) = x'\theta + \mathbb{E}(\varepsilon \mid x, z,\ \eta \ge -s) = x'\theta + \mathbb{E}(\varepsilon \mid \eta \ge -s)\,,$$

using $(\varepsilon, \eta) \perp (x, z)$. With (ε, η) jointly normal and $\mathrm{Var}(\eta)=1$, the regression of ε on η is $\mathbb{E}(\varepsilon \mid \eta) = \sigma_{\varepsilon\eta}\,\eta$ (since $\mathbb{E}(\varepsilon \mid \eta) = \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2}\eta = \sigma_{\varepsilon\eta}\eta$). By iterated expectations,

$$\mathbb{E}(\varepsilon \mid \eta \ge -s) = \sigma_{\varepsilon\eta}\,\mathbb{E}(\eta \mid \eta \ge -s)\,.$$

The truncated standard-normal mean is $\mathbb{E}(\eta \mid \eta \ge -s) = \dfrac{\phi(-s)}{1-\Phi(-s)} = \dfrac{\phi(s)}{\Phi(s)} = \lambda(s)$ (using $\phi(-s)=\phi(s)$, $1-\Phi(-s)=\Phi(s)$). Therefore

$$\boxed{\mathbb{E}(y \mid x, z, d=1) = x'\theta + \sigma_{\varepsilon\eta}\,\lambda(x'\pi_1 + z'\pi_2), \qquad \lambda(s) = \frac{\phi(s)}{\Phi(s)}}$$

逆 Mills $\lambda(s)=\phi(s)/\Phi(s)$ 来自截断正态：$\mathbb{E}(\eta \mid \eta > -s) = \phi(s)/\Phi(s)$。修正项系数 $\sigma_{\varepsilon\eta}$（ε 与选择误差 η 的协方差）正是选择偏误的源头——若 $\sigma_{\varepsilon\eta}=0$（无选择偏误），对选中样本直接 OLS 即一致。

(b) Identify (π_1, π_2) from probit. The choice probability is

$$\mathrm{P}(d=1 \mid x, z) = \mathrm{P}\big(\eta \ge -(x'\pi_1 + z'\pi_2)\big) = \Phi(x'\pi_1 + z'\pi_2)$$

(by $\eta \sim N(0,1)$ and symmetry $\mathrm{P}(\eta \ge -s) = \Phi(s)$). This is exactly a *probit* of d on (x, z):

$$\boxed{(\pi_1, \pi_2) \text{ are the probit coefficients of } d \text{ on } (x, z) \text{ (under the normalization } \sigma_\eta = 1).}$$

第一步 probit 把指标 $s = x'\pi_1 + z'\pi_2$ 估出来，于是 $\hat{\lambda}(\hat{s})$ 可算，作为第二步回归的「构造回归元」。$\sigma_\eta = 1$ 是必需的尺度归一（probit 只识别 π/σ_η）。

(c) Conditions to identify θ and $\sigma_{\varepsilon\eta}$. Among the selected $(d=1)$ observations, run

$$\mathbb{E}(y \mid x, z, d=1) = x'\theta + \sigma_{\varepsilon\eta}\,\lambda(\hat{s}), \qquad \hat{s} = x'\hat{\pi}_1 + z'\hat{\pi}_2.$$

This is a linear regression of y on $(x,\ \lambda(\hat{s}))$; θ and $\sigma_{\varepsilon\eta}$ are identified iff the regressors $[x,\ \lambda(\hat{s})]$ are not (asymptotically) collinear among the selected sample. Conditions:

(i) (π_1, π_2) identified from the first-stage probit (part b), so $\lambda(\cdot)$ is known.

(ii) *Exclusion restriction*: at least one variable in z enters the selection equation but is *excluded* from the outcome equation x. Then $\lambda(\hat{s})$ varies independently of x and is not collinear with x.

(iii) Full-rank condition: $\mathbb{E}([x, \lambda(\hat{s})]\,[x, \lambda(\hat{s})]' \mid d = 1)$ is nonsingular.

排他性约束是灵魂：没有 z 中的排他变量，$\lambda(\hat{s})$ 是 x 的非线性函数，识别只靠 λ 的非线性（曲率），在 λ 近似线性的区间会弱识别、极脆弱。有了排他变量，$\lambda(\hat{s})$ 有独立变异，$\theta, \sigma_{\varepsilon\eta}$ 稳健可识别。

$$\boxed{\theta, \sigma_{\varepsilon\eta} \text{ identified} \Longleftarrow (\pi_1, \pi_2) \text{ known } + \text{ exclusion restriction (a } z \text{ excluded from } x) + \text{ rank.}}$$

Remark (Q5 排他约束是 Heckit 命门).

Heckit 的全部识别力可归结一句：用 probit 估出 $\lambda(\hat{s})$ 当回归元，再在选中样本里 OLS。但若 $z \subseteq x$（无排他变量），$\lambda(x'\pi_1)$ 只是 x 的非线性变换，识别只靠 λ 的弯曲——这是著名的「Heckit 仅靠泛函形式识别」脆弱性。排他性约束（exclusion restriction）让识别建立在外生变异而非分布假设上。见 7。