Statistics II | **Problem Set 4**
Professor: Marcelo J. Moreira | **Solutions**
TA: Luan Borelli | **September 14, 2023**

---

1. *Let x_i be iid with pdf $g(x_i, \theta)$, where $\theta = (\theta_1, \theta_2)$.*

(a) *Find the asymptotic distribution of the MLE $\hat{\theta}_n = (\hat{\theta}_1, \hat{\theta}_2)$.*

Solution. Let $\theta^* = (\theta_1^*, \theta_2^*)$ be the true parameter vector. The maximum likelihood estimator $\hat{\theta}$, being the interior solution to the problem of maximizing the sample log-likelihood function $L_n(\theta) = n^{-1} \sum_{i=1}^n \ln g(x_i, \theta)$, satisfies the first order conditions

$$n^{-1} \sum_{i=1}^n \nabla_\theta \ln g(x_i, \hat{\theta}) = 0.$$

Assuming $\ln g(x_i, \theta)$ is C^2, the mean value theorem applies to the LHS and we have

$$n^{-1} \sum_{i=1}^n \nabla_\theta \ln g(x_i, \theta^*) + \left(n^{-1} \sum_{i=1}^n \nabla_{\theta\theta} \ln g(x_i, \bar{\theta}) \right) (\hat{\theta} - \theta^*) = 0,$$

for some mean value $\bar{\theta}$ "between" $\hat{\theta}$ and θ^*. By multiplying the above expression by $\sqrt{n}$ and solving for $\sqrt{n}(\hat{\theta} - \theta^*)$ we obtain

$$\sqrt{n}(\hat{\theta} - \theta^*) = - \left(n^{-1} \sum_{i=1}^n \nabla_{\theta\theta} \ln g(x_i, \bar{\theta}) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^m \nabla_\theta \ln g(x_i, \theta^*).$$

Since $\bar{\theta}$ is "between" θ^* and $\hat{\theta}$ and $\hat{\theta} \xrightarrow{p} \theta^*$, it follows that $\bar{\theta} \xrightarrow{p} \theta^*$. Thus, under standard regularity conditions,

$$\left(n^{-1} \sum_{i=1}^n \nabla_{\theta\theta} \ln g(x_i, \bar{\theta}) \right)^{-1} \xrightarrow[\text{LLN + CMT}]{p} \mathbb{E}[\nabla_{\theta\theta} \ln f(w_t; \theta^*)]^{-1} \equiv H(\theta^*)^{-1}.$$

Furthermore,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^m \nabla_\theta \ln g(x_i, \theta^*) \xrightarrow[\text{CLT}]{d} N(0, J),$$

where $J \equiv \mathbb{E}[\nabla_\theta \ln g(x_i, \theta^*) \, (\nabla_\theta \ln g(x_i, \theta^*))']$. Therefore, by Slutsky's theorem, we conclude

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} -H(\theta^*)^{-1} N(0, J) &= N(0, H(\theta^*)^{-1} J H(\theta^*)^{-1}) \\ &= N(0, J^{-1}), \end{aligned}$$

where the last equality follows from the information matrix equality, $H(\theta^*) = -J$. □

(b) *Find the asymptotic distribution of the MLE $\hat{\theta}_{1,n}$ for θ_1 when we know θ_2.*

Solution. When $\theta_2 = \theta_2^*$ is known, the log-likelihood function L_n becomes a univariate function of θ_1; that is, $L_n(\theta) = L_n(\theta_1)$. The maximum likelihood estimator $\hat{\theta}_1$, being the interior solution to the problem of maximizing $L_n(\theta_1)$, satisfies the first order condition

$$n^{-1}\sum_{i=1}^{n}\frac{d\ln g(x_i,\hat{\theta}_1)}{d\theta_1} = 0.$$

Again by the mean value theorem we have

$$n^{-1}\sum_{i=1}^{n}\frac{d\ln g(x_i,\theta_1^*)}{d\theta_1} + \left(n^{-1}\sum_{i=1}^{n}\frac{d^2\ln g(x_i,\bar{\theta}_1)}{d\bar{\theta}_1^2}\right)(\hat{\theta}_1 - \theta_1^*) = 0$$

for some mean value $\bar{\theta}_1$ between $\hat{\theta}_1$ and θ_1^*. By multiplying the above expression by $\sqrt{n}$ and solving for $\sqrt{n}(\hat{\theta}_1 - \theta_1^*)$ we obtain

$$\sqrt{n}(\hat{\theta}_1 - \theta_1^*) = -\left(n^{-1}\sum_{i=1}^{n}\frac{d^2\ln g(x_i,\bar{\theta}_1)}{d\bar{\theta}_1^2}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ln g(x_i,\theta_1^*).$$

By arguments similar to those presented in **(a)**, it follows that

$$\sqrt{n}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N\left(0, \mathbb{E}\left[\left(\frac{d\ln g(x_i,\theta_1^*)}{d\theta_1}\right)^2\right]^{-1}\right) = N(0, J_1^{-1}).$$

□

(c) *Compare your answers in itens **(a)** and **(b)**. Comment.*

Solution. Observe that in **(a)** we have

$$J = \begin{bmatrix} \mathbb{E}\left[\left(\frac{\partial \ln g(x_i,\theta^*)}{\partial\theta_1}\right)^2\right] & \mathbb{E}\left[\frac{\partial \ln g(x_i,\theta^*)}{\partial\theta_1}\frac{\partial \ln g(x_i,\theta^*)}{\partial\theta_2}\right] \\ \mathbb{E}\left[\frac{\partial \ln g(x_i,\theta^*)}{\partial\theta_1}\frac{\partial \ln g(x_i,\theta^*)}{\partial\theta_2}\right] & \mathbb{E}\left[\left(\frac{\partial \ln g(x_i,\theta^*)}{\partial\theta_2}\right)^2\right] \end{bmatrix} \equiv \begin{bmatrix} J_1 & J_2 \\ J_2 & J_3 \end{bmatrix},$$

whence

$$J^{-1} = \frac{1}{J_1J_3 - J_2^2}\begin{bmatrix} J_3 & -J_2 \\ -J_2 & J_1 \end{bmatrix}.$$

Therefore,

$$Avar(\hat{\theta}_1) = \frac{J_3}{J_1J_3 - J_2^2}.$$

In **(b)**,

$$Avar(\hat{\theta}_1) = J_1^{-1}.$$

Since variances are always non-negative,

$$\frac{J_3}{J_1 J_3 - J_2^2} = \frac{1}{J_1 - \underbrace{J_2^2/J_3}_{\geq 0}} \geq \frac{1}{J_1}.$$

So, in general, the asymptotic variance of $\hat{\theta}_1$ under unknown θ_2 is greater than it is under known θ_2, being equal only when $\text{Acov}[\hat{\theta}_1, \hat{\theta}_2] = J_2 = 0$. Intuitively, this is a natural result to expect, since when one of the parameters of interest that was previously unknown becomes known, the general uncertainty of the problem of estimating $\hat{\theta}_1$ reduces, which takes the form of a reduction in variance. This, of course, provided θ_2 is somehow related to $\hat{\theta}_1$; $J_2 \neq 0$. □

---

2. [10.4, LNs] *Let $W(X)$ be an unbiased estimator of θ.*

(a) *Show that if $W(X)$ is MVUE (minimum variance unbiased estimator), then it is unique.*

Solution. Suppose W' is another MVUE, and consider the estimator $W^* = \frac{1}{2}(W + W')$. Note that $\mathbb{E}_\theta[W^*] = \theta$ and

$$\begin{aligned} V_\theta[W^*] &= V_\theta\left[\frac{1}{2}W + \frac{1}{2}W'\right] \\ &= \frac{1}{4}V_\theta[W] + \frac{1}{4}V_\theta[W'] + \frac{1}{2}\text{Cov}_\theta[W, W'] \\ &\leq \frac{1}{4}V_\theta[W] + \frac{1}{4}V_\theta[W'] + \frac{1}{2}(V_\theta[W]V_\theta[W'])^{1/2} = V_\theta[W]. \end{aligned}$$

But if the above inequality is strict, then the MVUE property of W is contradicted, so we must have equality for all θ. Since the inequality is an application of Cauchy-Schwartz, we can have equality only if $W' = a(\theta)W + b(\theta)$. Now using properties of covariance, we have

$$\text{Cov}_\theta[W, W'] = \text{Cov}_\theta[W, a(\theta)W + b(\theta)] = \text{Cov}_\theta[W, a(\theta)W] = a(\theta)V_\theta[W],$$

but $\text{Cov}_\theta[W, W'] = V_\theta[W]$, since equality must hold in the previous inequality. Hence $a(\theta) = 1$ and, since $\mathbb{E}_\theta[W'] = \theta$, we must have $b(\theta) = 0$ and $W = W'$, showing that W is unique. □

(b) *Show that $W(X)$ is MVUE if and only if it is uncorrelated with all unbiased estimators U of zero (i.e., $\mathbb{E}_\theta U = 0$ for any θ).*

Solution. If $\text{Cov}[W, U] = 0$ for any unbiased estimator U of θ, then in particular, for any unbiased estimator W' of W and $U = W - W'$, we have $\text{Cov}[W, W - W'] = 0$. Observe that

$$V[W'] = V[U - W] = V[W] - 2\text{Cov}[W, U] + V[U] = V[W] + V[U] \geq V[W].$$

That is, W is MVUE. Conversely, suppose W is MVUE and we have another estimator U that satisfies $\mathbb{E}_\theta[U] = 0$ for all θ. The estimator

$$W'' = W + \alpha U,$$

where α is a constant, satisfies $\mathbb{E}_\theta[W''] = \theta$ and hence is also an unbiased estimator of θ. The variance of W'' is

$$V_\theta[W''] = V_\theta[W + \alpha U] = V_\theta[W] + 2\alpha \mathrm{Cov}_\theta[W, U] + \alpha^2 V_\theta[U].$$

If for some $\theta = \theta_0$ we have $\mathrm{Cov}_{\theta_0}[W, U] < 0$, then we can make $2\alpha \mathrm{Cov}_{\theta_0}[W, U] + \alpha^2 V_{\theta_0}[U] < 0$ by choosing $\alpha \in (0, -2\mathrm{Cov}_{\theta_0}[W, U]/V_{\theta_0}[U])$. Hence W'' will be better than W at $\theta = \theta_0$, contradicting W being MVUE. A similar argument will show that if $\mathrm{Cov}_{\theta_0}[W, U] > 0$ for any θ_0, W also cannot be MVUE. Therefore W must satisfy $\mathrm{Cov}_\theta[W, U] = 0$ for all θ, for any U satisfying $\mathbb{E}_\theta[U] = 0$. □

---

3. [R1-R5, Ch.15, Goldberger] *Prove the rules R1-R6 from Chapter 15 of Golberger.* Throughout we suppose that the $n \times 1$ random vector y has expectation vector $\mathbb{E}[y] = \mu$ and variance matrix $V[y] = \Sigma$, and write $\varepsilon = y - \mu$.

(R1) *Let $z = g + h'y$, where the scalar g and the $n \times 1$ vector h are constants. Then $\mathbb{E}[z] = g + h'\mu$ and $V[z] = h'\Sigma h$.*

Solution. From linearity of the expectation operator,

$$\mathbb{E}[z] = \mathbb{E}[g + h'y] = \mathbb{E}[g] + \mathbb{E}[h'y] = g + h'\mathbb{E}[y] = g + h'\mu.$$

Further, let $z^* = z - \mathbb{E}[z]$. Then $z^* = h'y - h'\mu = h'(y - \mu) = h'\varepsilon$, and $z^{*2} = (h'\varepsilon)^2 = (h'\varepsilon)(h'\varepsilon) = h'\varepsilon\varepsilon'h$. So

$$V[z] = \mathbb{E}[z^{*2}] = \mathbb{E}[h'\varepsilon\varepsilon'h] = h'[\varepsilon\varepsilon']h = h'V[\varepsilon]h = h'\Sigma h.$$

□

(R2) *Let $z = g + Hy$, where the $k \times 1$ vector g and the $k \times n$ matrix H are constants. Then the $k \times 1$ random vector z has $\mathbb{E}[z] = g + H\mu$ and $V[z] = H\Sigma H'$.*

Solution. From linearity of the expectation operator,

$$\mathbb{E}[z] = \mathbb{E}[g + Hy] = \mathbb{E}[g] + \mathbb{E}[Hy] = g + H\mathbb{E}[y] = g + H\mu.$$

Further, let $z^* = z - \mathbb{E}[z]$. Then $z^* = H(y - \mu) = H\varepsilon$, and $z^* z^{*'} = H\varepsilon\varepsilon'H'$. So

$$V[z] = \mathbb{E}[z^* z^{*'}] = \mathbb{E}[H\varepsilon\varepsilon'H'] = H\mathbb{E}[\varepsilon\varepsilon']H' = H\Sigma H'.$$

□

(R3) *Let $W = yy'$. Then the $n \times n$ random matrix W has expectation $\mathbb{E}[W] = \Sigma + \mu\mu'$.*

Solution. Write $yy' = (\mu + \varepsilon)(\mu + \varepsilon)' = \mu\mu' + \mu\varepsilon' + \varepsilon\mu' + \varepsilon\varepsilon'$, which, since μ is constant and $\mathbb{E}[\varepsilon] = 0$, implies $\mathbb{E}[yy'] = \mu\mu' + \Sigma$. □

(R4) *Let $w = y'y$. Then the scalar random variable w has expectation $\mathbb{E}[w] = tr(\Sigma) + \mu'\mu$.*

Solution. Write $y'y = \text{tr}(y'y) = \text{tr}(yy') = \text{tr}(W)$, so

$$\begin{aligned}\mathbb{E}[y'y] = \mathbb{E}[\text{tr}(W)] = \text{tr}(\mathbb{E}[W]) &= \text{tr}(\Sigma + \mu\mu') \\ &= \text{tr}(\Sigma) + \text{tr}(\mu\mu') = \text{tr}(\Sigma) + \text{tr}(\mu'\mu) = \text{tr}(\Sigma) + \mu'\mu,\end{aligned}$$

using the facts that trace is a linear operator, and that if AB and BA are both square matrices, then $\text{tr}(AB) = \text{tr}(BA)$. □

(R5) *Let $w = y'Ty$, where the $n \times n$ matrix T is constant. Then the random variable w has expectation $\mathbb{E}[w] = tr(T\Sigma) + \mu'T\mu$.*

Solution. Write $y'Ty = \text{tr}(y'Ty) = \text{tr}(Tyy') = \text{tr}(TW)$. Then

$$\begin{aligned}\mathbb{E}[y'Ty] &= \mathbb{E}[\text{tr}(TW)] = \text{tr}(\mathbb{E}[TW]) = \text{tr}[T\mathbb{E}[W]] \\ &= \text{tr}(T(\Sigma + \mu\mu')) = \text{tr}(T\Sigma) + \text{tr}(T\mu\mu') \\ &= \text{tr}(T\Sigma) + \mu'T\mu.\end{aligned}$$

□

(R6) *Let $z_1 = g_1 + H_1 y$, $z_2 = g_2 + H_2 y$, where the $m_1 \times 1$ vector g_1, the $m_2 \times 1$ vector g_2, the $m_1 \times n$ matrix H_1, and the $m_2 \times n$ matrix H_2 are constants. Then $Cov[z_1, z_2] = H_1 \Sigma H_2'$.*

Solution. Let $z_1^* = z_1 - \mathbb{E}[z_1] = H_1\varepsilon$, and $z_2^* = z_2 - \mathbb{E}[z_2] = H_2\varepsilon$. Then $z_1^* z_2^{*'} = H_1\varepsilon\varepsilon' H_2'$, so

$$\text{Cov}[z_1, z_2] = \mathbb{E}[z_1^* z_2^{*'}] = H_1 \mathbb{E}[\varepsilon\varepsilon'] H_2' = H_1 \Sigma H_2'.$$

□

---

4. [7.11, LNs] *Consider the univariate Lindeberg-Feller CLT. Let the array $X_{n,m}$, $m = 1, \dots, n$, be independent zero mean random variables; if $\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2] \to \sigma^2$, and for all $\varepsilon > 0$, $\lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 I(|X_{n,m}| > \varepsilon)] = 0$, then $\sum_{t=1}^{n} X_{n,t} \xrightarrow{d} N(0, \Sigma)$.*

(a) *Give the definition of convergence in distribution: $S_n \xrightarrow{d} S$.*

Solution. A sequence $S_1, S_2, \dots$ of random vectors of dimension $k \in \mathbb{N}$ converges in distribution to a random vector S of dimension k if at all continuity points $s \in \mathbb{R}^k$ of the joint distribution $F_S(\cdot)$, $\lim_{n\to\infty} F_{S_n}(s) = F_S(s)$. In this case, we denote $S_n \xrightarrow{d} S$. □

(b) *Prove the Lyapunov condition: if there exists* $\delta > 0$ *such that* $\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[|X_{n,m}|^{2+\delta}] = 0$*, then* $\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[X_{n,m}^2 I(|X_{n,m}| > \varepsilon)] = 0$*. (In practice, we often take* $\delta = 1$*).*

Solution. Suppose there exists $\delta > 0$ such that $\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[|X_{n,m}|^{2+\delta}] = 0$. It follows that

$$\begin{aligned}
\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[X_{n,m}^2 I(|X_{n,m}| > \varepsilon)] &= \lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}\left[|X_{n,m}|^{2+\delta}\frac{I(|X_{n,m}|^\delta > \varepsilon^\delta)}{|X_{n,m}|^\delta}\right] \\
&\le \lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}\left(|X_{n,m}|^{2+\delta}\frac{1}{\varepsilon^\delta}\right) \\
&= \frac{1}{\varepsilon^\delta}\underbrace{\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[|X_{n,m}|^{2+\delta}]}_{=0} = 0.
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[X_{n,m}^2 I(|X_{n,m}| > \varepsilon)] = 0.$$

□

(c) *For each* n*, let* $Y_{n,m}$*,* $1 \le m \le n$*, be independent* k*-dimensional random vectors with* $\mathbb{E}[Y_{n,m}] = 0$*. Suppose that* $\sum_{m=1}^n \mathbb{E}[Y_{n,m}Y_{n,m}'] \to \Sigma$ *and, for all* $\varepsilon > 0$*,*

$$\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[\|Y_{n,m}\|^2 I(\|Y_{n,m}\| > \varepsilon)] = 0.$$

Show that $S_n = \sum_{t=1}^n Y_{n,m} \xrightarrow{d} N(0,\Sigma)$*.*

Solution. Take any k-dimensional $\alpha \ne 0$ and denote $Z_{n,m} \equiv \alpha' Y_{n,m}$. Since $Y_{n,m}$, $1 \le m \le n$, are independent, then $Z_{n,m}$, $1 \le m \le n$, are independent. Note that $Z_{n,m}^2 = \alpha'(Y_{n,m}Y_{n,m}')\alpha$, so we have

$$\begin{aligned}
\lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[Z_{n,m}^2] &= \lim_{n\to\infty}\sum_{m=1}^n \mathbb{E}[\alpha'(Y_{n,m}Y_{n,m}')\alpha] = \lim_{n\to\infty}\sum_{m=1}^n \alpha'\mathbb{E}[Y_{n,m}Y_{n,m}']\alpha' \\
&= \lim_{n\to\infty}\alpha'\left(\sum_{m=1}^n \mathbb{E}[Y_{n,m}Y_{n,m}']\right)\alpha = \alpha'\left(\lim_{n\to\infty}\mathbb{E}[Y_{n,m}Y_{n,m}']\right)\alpha \\
&= \alpha'\Sigma\alpha.
\end{aligned}$$

Moreover, by the Cauchy-Schwarz inequality, we have $|Z_{n,m}| = |\alpha' Y_{n,m}| \le \|\alpha\|\|Y_{n,m}\|$. Thus, if $I(|\alpha' Y_{n,m}| > \varepsilon) = 1$, then $I(\|\alpha\|\|Y_{n,m}\| > \varepsilon) = 1$. This implies that

$$I(|\alpha' Y_{n,m}| > \varepsilon) \le I(\|Y_{n,m}\| > \varepsilon/\|\alpha\|).$$

Finally,

$$\begin{aligned}\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[Z_{n,m}^2 I(|Z_{n,m}|>\varepsilon)] &= \lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[(\alpha' Y_{n,m})^2 I(|\alpha' Y_{n,m}|>\varepsilon)]\\ &\le \lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[(\|\alpha\|\|Y_{n,m}\|)^2 I(\|Y_{n,m}\|>\varepsilon/\|\alpha\|)]\\ &= \|\alpha\|^2 \underbrace{\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^2 I(\|Y_{n,m}\|)\ge\varepsilon/\|\alpha\|]}_{=0} = 0.\end{aligned}$$

Therefore, by the univariate Lindeberg-Feller CLT,

$$\alpha'\left(\sum_{m=1}^{n} Y_{n,m}\right) = \sum_{m=1}^{n}\alpha' Y_{n,m} = \sum_{m=1}^{n} Z_{n,m} \xrightarrow{d} N(0,\alpha'\Sigma\alpha).$$

Let S be a random vector with distribution $N(0,\Sigma)$. By the definition of the multivariate normal distribution, $\alpha' S \sim N(0,\alpha'\Sigma\alpha)$. Thus, we have $\alpha' S_n \xrightarrow{d} \alpha' S$. Since $\alpha \neq 0$ is arbitrary, it follows by the Crámer-Wold device that $S_n \xrightarrow{d} S$; that is, $\sum_{m=1}^{n} Y_{n,m} \xrightarrow{d} N(0,\Sigma)$. □

(d) *Propose a multivariate version of the Liapunov's theorem.*

Solution. If there exists $\delta > 0$ such that

$$\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^{2+\delta}] = 0,$$

then $\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^2 I(\|Y_{n,m}\|>\varepsilon)] = 0$. For a proof, suppose there exists $\delta > 0$ such that $\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^{2+\delta}] = 0$. It follows that

$$\begin{aligned}\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^2 I(\|Y_{n,m}\|>\varepsilon)] &= \lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}\left[\|Y_{n,m}\|^{2+\delta}\frac{I(\|Y_{n,m}\|^\delta>\varepsilon^\delta)}{\|Y_{n,m}\|^\delta}\right]\\ &\le \lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}\left[\|Y_{n,m}\|^{2+\delta}\frac{1}{\varepsilon^\delta}\right]\\ &= \frac{1}{\varepsilon^\delta}\underbrace{\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^{2+\delta}]}_{=0} = 0.\end{aligned}$$

Therefore $\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[\|Y_{n,m}\|^2 I(\|Y_{n,m}\|>\varepsilon)] = 0$, as desired. □

5. [16.29, LNs] *Consider the regression $y = X\beta + u$, where $\mathbb{E}[u|X] = 0$ and $V[u|X] = \sigma^2 I_N$ with unknown σ^2.*

(a) *Find the Wald statistic for $H_0 : R\beta - r = 0$. Derive its distribution under H_0.*

Solution. Let R be a full rank $q \times k$ matrix. The Wald statistic for $H_0 : R\beta - r = 0$ is

$$W = (R\hat{\beta} - r)'\hat{V}_{R\hat{\beta}}^{-1}(R\hat{\beta} - r) = \sqrt{n}(R\hat{\beta} - r)'\hat{V}_{R\beta}^{-1}\sqrt{n}(R\hat{\beta} - r),$$

where $\hat{V}_{R\hat{\beta}}$ is some (consistent) covariance matrix estimator for $R\hat{\beta}$ and $\hat{V}_{R\beta} \equiv n\hat{V}_{R\hat{\beta}}$.

We know that $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2)$. Under the null, $R\beta = r$. Therefore, by the multivariate delta method,

$$\sqrt{n}(R\hat{\beta} - r) = \sqrt{n}(R\hat{\beta} - R\beta) \underset{H_0}{\xrightarrow{d}} RN(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2) = N(0, \underbrace{R\mathbb{E}[x_i x_i']^{-1}R'\sigma^2}_{\equiv V_{R\hat{\beta}}}) \equiv Z.$$

Since $\hat{V}_{R\hat{\beta}} \xrightarrow{p} V_{R\beta}$, it follows by the continuous mapping theorem and Slutsky's theorem that

$$W \underset{H_0}{\xrightarrow{d}} Z'[R\mathbb{E}[x_i x_i']^{-1}R'\sigma^2]^{-1}Z \sim \chi^2_q.$$

Thus the Wald statistic for $H_0 : R\beta - r = 0$ is asymptotically chi-squared distributed with q degrees of freedom. □

(b) *Show that the Wald statistic for $R\beta - r = 0$ equals the largest Wald statistic among all one-dimensional tests for restrictions of the form $c'(R\beta - r) = 0$. Comment.*

Solution. The Wald statistic for restrictions of the form $c'(R\beta - r) = 0$ is

$$W_c = [c'(R\hat{\beta} - r)]'\hat{V}_{c'R\hat{\beta}}^{-1}c'(R\hat{\beta} - r).$$

Notice that

$$\hat{V}_{c'R\hat{\beta}} = c'\hat{V}_{R\hat{\beta}}c,$$

Write

$$\begin{aligned}
W_c &= [c'(R\hat{\beta} - r)]'[c'\hat{V}_{R\hat{\beta}}c]^{-1}c'(R\hat{\beta} - r) \\
&= \frac{[c'(R\hat{\beta} - r)]'c'(R\hat{\beta} - r)}{c'\hat{V}_{R\hat{\beta}}c} \\
&= \frac{[c'(R\hat{\beta} - r)]^2}{c'\hat{V}_{R\hat{\beta}}c} \\
&= \frac{[v'\hat{V}_{R\hat{\beta}}^{-1/2}(R\hat{\beta} - r)]^2}{v'v} \\
\text{(Cauchy-Schwartz)} &\leq \frac{v'v[\hat{V}_{R\hat{\beta}}^{-1/2}(R\hat{\beta} - r)]'\hat{V}_{R\hat{\beta}}^{-1/2}(R\hat{\beta} - r)}{v'v} \\
&= (R\hat{\beta} - r)'\hat{V}_{R\hat{\beta}}^{-1}(R\hat{\beta} - r) = W,
\end{aligned}$$

where $v = \hat{V}_{R\hat{\beta}}^{1/2} c$ and $\hat{V}_{R\hat{\beta}}^{1/2}$ is such that $\hat{V}_{R\hat{\beta}}^{1/2}\hat{V}_{R\hat{\beta}}^{1/2} = \hat{V}_{R\hat{\beta}}$.

Observe that, in particular, one could set $c = e_i$ to be a canonical selector vector that selects the i-th restriction in $Rb - r$. Suppose instead of the original procedure of jointly testing all the restrictions, $H_0 : Rb - r = 0$, one proposes a new different procedure: testing each restriction separately, by performing a sequence of q independent tests $H_0 : e_i'(Rb - r) = 0$, $i = 1, \dots, q$, with associated Wald statistic W_i, and then claiming that the joint restrictions are rejected if at least one of the separate tests rejects the null. The above result tells us that $W_i \leq W$ for all $i = 1, \dots, q$. We reject the null when, for a given $1 - \alpha$ quantile k, the Wald statistic is greater than k. Thus for some i and some α we could have quantiles k_1 and k_q such that $k_1 < W_i < W < k_q$, where k_1 and k_q are the $1 - \alpha$ quantiles of the χ_1^2 and χ_q^2 distributions, respectively. In this case, the new procedure would reject the null, while the original one would not. This shows that the new proposed procedure is problematic. Indeed, it could even be used for cheating: for example, one could purposely seek a significance level such that $k_1 < W_i < W < k_q$ for *some* i so that we reject the null based on the new procedure when actually — based on the original correct one — we should not! □

---

6. *Consider the model*

$$y_i = x_i'\beta + u_i$$

where (x_i', u_i) are iid with $u_i|x_i$ have the density $f(u) \in C^2$ (with support $-\infty < u < \infty$). Assume that

$$\mathbb{E}[U] = \int_{-\infty}^{\infty} u f(u) = 0$$
$$\text{and } V[U] = \mathbb{E}[U^2] = \int_{-\infty}^{\infty} u^2 f(u) = \sigma^2.$$

(a) *Use transformation of variables to show that the (conditional) pdf of $y_i|x_i$ is given by $g(y_i|x_i) = f(y_i - x_i'\beta)$.*

Solution. Recall that if a continuous random variable X has pdf f_X, then an increasing 1-to-1 transformation $Y = h(X)$ of this random variable has pdf $f_X(h^{-1}(y)) \cdot |\frac{\partial h^{-1}(y)}{\partial y}|$. Here y_i is an increasing one-to-one transformation of u_i, which has density f. The inverse transformation is $h^{-1}(u) = u - x_i'\beta$. Therefore the pdf of y_i is $f(h^{-1}(y_i)) \cdot |\frac{\partial h^{-1}(y_i)}{\partial y_i}| = f(y_i - x_i'\beta) = f(u_i)$. □

(b) *Find the likelihood of $y = (y_1, \dots, y_n)$ conditional on $X = (x_1, \dots, x_n)'$.*

Solution. The likelihood is $\mathcal{L}(\beta) = \prod_{i=1}^n f(y_i - x_i'\beta)$. □

(c) *State the Gauss-Markov theorem.*

Solution. In the *homoskedastic* linear regression model, if $\tilde{\beta}$ is a linear unbiased estimator of β, then $V[\tilde{\beta}|X] \geq \sigma^2(X'X)^{-1}$. □

(d) *We will show in item **(f)** that the asymptotic variance of $\sqrt{n}(\tilde{\beta}-\beta^*)$ can be smaller than the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta^*)$, where $\tilde{\beta}$ is the MLE and $\hat{\beta}$ is the OLS estimator. Explain why this result does not contradict the Gauss-Markov theorem.*

Solution. When the MLE estimator lacks linearity and/or unbiasedness, which is perfectly possible, it falls outside the scope of the Gauss-Markov theorem. Consequently, the asymptotic variance of $\sqrt{n}(\tilde{\beta}-\beta^*)$ can be smaller than that of $\sqrt{n}(\hat{\beta}-\beta^*)$ without posing any contradictions. □

(e) *Find the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta^*)$.*

Solution. Write $\sqrt{n}(\hat{\beta}-\beta^*) = \left(n^{-1}\sum_{i=1}^n x_i x_i'\right)^{-1}\sqrt{n}\left(n^{-1}\sum_{i=1}^n x_i u_i\right)$. By standard LLN, CMT, CLT, and Slutsky arguments it follows that $\sqrt{n}(\hat{\beta}-\beta^*) \xrightarrow{d} N(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2)$. □

(f) *Show algebraically that (i) the asymptotic variance of $\sqrt{n}(\tilde{\beta}-\beta^*)$ is no larger than the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta^*)$; and (ii) give a necessary and sufficient condition on the density $f(u)$ for the asymptotic variance of $\hat{\beta}$ and $\tilde{\beta}$ to be the same.*

Solution. Under standard regularity conditions, taking logs of the likelihood function obtained in **(a)**, using first-order conditions and appealing to the mean value theorem, one can show just as in Exercise 1 that $\sqrt{n}(\tilde{\beta}-\beta_0) \xrightarrow{d} N(0, J^{-1})$, where $J = \mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 x_i x_i'\right]$. Recall that $A-B$ is PSD if and only if $B^{-1}-A^{-1}$ is PSD. Therefore

$$
\begin{aligned}
& \left(\frac{1}{\sigma^2}\mathbb{E}\left[x_i x_i'\right]\right)^{-1} - \left(\mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 x_i x_i'\right]\right)^{-1} \gtrsim 0 \\
\iff & \mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 x_i x_i'\right] - \frac{1}{\sigma^2}\mathbb{E}\left[x_i x_i'\right] \gtrsim 0 \\
\text{(LIE)} \iff & \mathbb{E}\left[\mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 \middle| x_i\right] x_i x_i'\right] - \mathbb{E}\left[\frac{1}{\sigma^2} x_i x_i'\right] \gtrsim 0. \qquad (1)
\end{aligned}
$$

From Cauchy-Schwartz inequality,

$$
\underbrace{\mathbb{E}[u^2|x_i]}_{=\sigma^2}\,\mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 \middle| x_i\right] \geq \left(\underbrace{\mathbb{E}\left[u\frac{f'(u_i)}{f(u_i)} \middle| x_i\right]}_{=-1}\right)^2 = 1,
$$
[1]

whence

$$
\mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 \middle| x_i\right] \geq \frac{1}{\sigma^2}.
$$

[1] Observe that $\mathbb{E}\left[u_i\frac{f'(u_i)}{f(u_i)} \middle| x_i\right] = \int_{-\infty}^{\infty} u_i \frac{f'(u_i)}{f(u_i)} f(u_i)\, du_i = \int_{-\infty}^{\infty} u_i f'(u_i) du_i = u_i f(u_i)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(u_i)\, du_i$, by integration by parts. Since f is a pdf, the second term equals 1. You can show that the first term is zero.

Therefore (1) holds and hence

$$\text{Avar}(\hat{\beta}) - \text{Avar}(\tilde{\beta}) \succsim 0.$$

A necessary and sufficient condition on the density $f(u_i)$ for the asymptotic variance of $\hat{\beta}$ and $\tilde{\beta}$ to be the same is $\mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 \middle| x_i\right] = 1/\sigma^2$. A simple sufficient condition is u_i being normally distributed. Observe that in this case we would have $f'(u_i)/f(u_i) = -u_i/\sigma^2$, whence $\mathbb{E}\left[\left(\frac{f'(u_i)}{f(u_i)}\right)^2 \middle| x_i\right] = 1/\sigma^2$. □

---

7. [16.34, LNs] *Consider the following regression model with the explanatory variable being a time trend:*

$$y_i = i\beta + u_i, \quad i = 1, \ldots, N,$$

where the u_i are i.i.d. with $\mathbb{E}[u_i] = 0$, $V[u_i] = \sigma^2$, and $\mathbb{E}|u_i|^3 = C$.

(a) *Show that*

$$\hat{\beta}_N = \frac{\sum_{i=1}^N iy_i}{\sum_{i=1}^N i^2}$$

is an unbiased estimator of β.

Solution. Observe that

$$\hat{\beta}_N = \frac{\sum_{i=1}^N i(i\beta + u_i)}{\sum_{i=1}^N i^2} = \beta + \frac{\sum_{i=1}^N iu_i}{\sum_{i=1}^N i^2},$$

whence

$$\mathbb{E}[\hat{\beta}_N] = \mathbb{E}[\beta] + \mathbb{E}\left[\frac{\sum_{i=1}^N iu_i}{\sum_{i=1}^N i^2}\right] = \beta + \frac{\sum_{i=1}^N i\mathbb{E}[u_i]}{\sum_{i=1}^N i^2} = \beta.$$

□

(b) *Show that*

$$V[\hat{\beta}_N] = \frac{6\sigma^2}{N(N+1)(2N+1)}.$$

Hint: it may be helpful to know that $\sum_{i=1}^N i^2 = \frac{1}{6}N(N+1)(2N+1)$.

Solution.

$$V[\hat{\beta}_N] = V\left[\beta + \frac{\sum_{i=1}^N iu_i}{\sum_{i=1}^N i^2}\right] = V\left[\frac{\sum_{i=1}^N iu_i}{\sum_{i=1}^N i^2}\right] = \frac{\sum_{i=1}^N i^2 V[u_i]}{\left(\sum_{i=1}^N i^2\right)^2}$$

$$= \frac{\sum_{i=1}^N i^2}{\left(\sum_{i=1}^N i^2\right)^2} V[u_i] = \frac{1}{\sum_{i=1}^N i^2}\sigma^2 = \frac{6\sigma^2}{N(N+1)(2N+1)}.$$

□

(c) *Prove that $N^{1/2}(\hat{\beta}_N - \beta) \xrightarrow{p} 0$ using Chebyshev's inequality, and comment how this result differs from what we have seen in class.*

Solution. By Chebyshev's inequality

$$0 \geq P(|\sqrt{N}(\hat{\beta}_N - \beta)| \geq \varepsilon) \geq \frac{NV[\hat{\beta}_N]}{\varepsilon^2} = \frac{1}{\varepsilon^2}\frac{6\sigma^2}{(N+1)(2N+1)} \to 0.$$

Therefore $\sqrt{N}(\hat{\beta}_N - \beta) \xrightarrow{p} 0$. □

(d) *Find r^* such that $N^{r^*}(\hat{\beta}_N - \beta)$ converges in distribution to a nondegenerate distribution.*

Solution. Write

$$N^{r^*}(\hat{\beta}_N - \beta) = N^{r^*}\frac{\sum_{i=1}^N iu_i}{\sum_{j=1}^N j^2} = \sqrt{N}\left(N^{-1}\sum_{i=1}^N \underbrace{\frac{iu_i}{\sum_{j=1}^N j^2}N^{r^*+1/2}}_{\equiv X_{ni}}\right).$$

Observe that $\mathbb{E}[X_{ni}] = 0$. In order for Lindeberg-Feller CLT to be applied, the first condition we need to verify is the Lyapunov condition. For this, it suffices to show that for some $\delta > 0$ we have

$$\lim_{N\to\infty}\frac{1}{\bar{\sigma}_n^{2+\delta}}\sum_{i=1}^N \mathbb{E}|X_{ni}|^{2+\delta} = 0,$$

where $\bar{\sigma}_n^2 = \sum_{i=1}^N V[X_{ni}]$. Take $\delta = 1$ and observe that

$$\begin{aligned}\sum_{i=1}^N \mathbb{E}|X_{ni}|^3 &= \sum_{i=1}^N \mathbb{E}\left[\frac{i^3|u_i^3|}{\left(\sum_{j=1}^N j^2\right)^3}N^{3r^*+3/2}\right] \\ &= \sum_{i=1}^N \frac{i^3}{\left(\sum_{j=1}^N j^2\right)^3}N^{3r^*+3/2}\mathbb{E}|u_i^3| \\ &= \frac{\left(\sum_{i=1}^N i^3\right)N^{3r^*+3/2}}{\left(\sum_{i=1}^N i^2\right)^3}C.\end{aligned}$$

Recall that $\sum_i^N i^2$ has degree 3 in N, so the denominator has degree 27 in N. One can show that $\sum_{i=1}^N i^3 = \frac{N^2(N+1)}{4}$, so the numerator has degree $4 + (3r^* + 3/2)$ in N. Thus, provided $r^* \leq 7.1666\ldots$, the above limit is zero. Now, it rests to verify that $\bar{\sigma}_n^{2+\delta} = \bar{\sigma}_n^3$ does not converge to zero, which is equivalent to $\bar{\sigma}_n^2$ not converging to zero. We have

$$\bar{\sigma}_n^2 = \sum_{i=1}^N \frac{6\sigma^2 N^{2r^*}}{(N+1)(2N+1)} = \frac{6\sigma^2 N^{2r^*+1}}{(N+1)(2N+1)}.$$

For this to not converge to zero we must have $2r^* + 1 \geq 2$, and so $r^* \geq 1/2$. Therefore, for any $1/2 \leq r^* \leq 7.1666\ldots$ the Lyapunov condition holds, and hence the Lindeberg-Feller CLT holds. Now it rests to verify that the average variance $n^{-1}\bar{\sigma}_n^2$ of X_{ni} converges to a constant, say $\bar{\sigma}^2$. Observe that the variance of X_{ni} is

$$V[X_{ni}] = \frac{6\sigma^2}{N(N+1)(2N+1)} N^{2r^*+1} = \frac{6\sigma^2 N^{2r^*}}{(N+1)(2N+1)}.$$

Therefore we need to ensure that

$$N^{-1}\sum_{i=1}^{N} V[X_{ni}] = N^{-1}\sum_{i=1}^{N} \frac{6\sigma^2}{(N+1)(2N+1)} N^{2r^*} = \frac{6\sigma^2 N^{2r^*}}{(N+1)(2N+1)} \xrightarrow{p} \bar{\sigma}^2 < \infty.$$

For this to happen we need to ensure that the denominator grows at least as fast as the numerator. That is, the maximum degree of N in the denominator must be at least as large as the degree of N in the numerator. Therefore we need $2r^* \leq 2$, which implies $r^* \leq 1$. However, the question asks specifically for a *nondegenerate* distribution; that is, we additionally must have $\bar{\sigma}^2 \neq 0$. Observe that for any $r^* < 1$, $\bar{\sigma}^2 = 0$. Therefore, we must have $r^* = 1$. □

(e) *Find the limiting distribution of* $N^{r^*}(\hat{\beta}_N - \beta)$, *for* r^* *found in part* ***(d)***.

Solution. For $r^* = 1$ we have

$$\bar{\sigma}^2 = \lim_{N\to\infty} \frac{6\sigma^2 N^2}{(N+1)(2N+1)} = 3\sigma^2.$$

It follows by the Lindeberg-Feller CLT that $N^{r^*}(\hat{\beta}_N - \beta) = N(\hat{\beta}_N - \beta) \xrightarrow{d} N(0, 3\sigma^2)$. □

---

8. [9.19, Hansen] *An economist estimates* $Y = X_1'\beta_1 + X_2\beta_2 + e$ *by least squares and tests the hypothesis* $H_0 : \beta_2 = 0$ *against* $H_1 : \beta_2 \neq 0$. *Assume* $\beta_1 \in \mathbb{R}^k$ *and* $\beta_2 \in \mathbb{R}$. *She obtains a Wald statistic* $W = 0.34$. *The sample size is* $n = 500$.

(a) *What is the correct degrees of freedom for the* χ^2 *distribution to evaluate the significance of the Wald statistic?*

Solution. $q = 1$. The dimension of β_2 is 1, and the hypothesis is that β_2 is zero, which is one restriction. □

(b) *The Wald statistic* W *is very small. Indeed, is it less than the 1% quantile of the appropriate* χ^2 *distribution? If so, should you reject* H_0? *Explain your reasoning.*

Solution. Yes. The 1% quantile of a χ_1^2 is ≈ 0.00016, which is less than $W = 0.34$. No. The Wald test rejects for large values of W_n, when $W_n \geq c$ for some c. A test which rejects for small W_n can have correct Type I error, but will have low power. An asymptotic α% test rejects H_0 if $W_n \geq c_\alpha$ where c_α is the $1-\alpha$ quantile (the upper α quantile), that is $P(W_n \geq c_\alpha) = 1 - P(W_n < c_\alpha) = \alpha$. An asymptotic 5% test rejects if $W_n \geq 3.84$, and an asymptotic 1% test rejects for $W_n \geq 6.63$. Since W_n is far smaller, H_0 is not rejected. The question about the 1% quantile is misleading. It is not the 1% critical value. It is also not sensible to talk about test with 99% Type I error. So there is no sense in which 0.00016 is a reasonable critical value for a test. □

---

9. [9.20, Hansen] *You are reading a paper, and it reports the results from two nested OLS regressions:*

$$Y_i = X_{1i}'\tilde{\beta}_1 + \tilde{e}_i$$
$$Y_i = X_{1i}'\hat{\beta}_1 + X_{2i}'\hat{\beta}_2 + \hat{e}_i.$$

Some summary statistics are reported. For the short regression, $R^2 = .20$, $\sum_{i=1}^n \tilde{e}_i^2 = 106$, the number of coefficients is 5 and $n = 50$. For the long regression, $R^2 = .26$, $\sum_{i=1}^n \hat{e}_i^2 = 100$, the number of coefficients is 8 and $n = 50$. You are curious if the estimate $\hat{\beta}_2$ is statistically different from the zero vector. Is there a way to determine an answer from this information? Do you have to make any assumptions (beyond the standard regularity conditions) to justify your answer?

Solution. The question asks if the estimate is statistically different than zero. This is asking for a statistical test.[2] The specified null hypothesis is that $\beta_2 = 0$. The general Wald statistic would be appropriate, but cannot be calculated from the information. However, the Wald statistic assuming homoskedasticity can be calculated. Thus the key assumption required is that the error is conditionally homoskedastic: $\mathbb{E}[e_i^2|x_i] = \sigma^2$, a constant. The Wald statistic for (1) versus (2) assuming homoskedasticity is

$$\begin{aligned}
W &= n\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \\
&= n\left(\frac{\sum_{i=1}^n \tilde{e}_i - \sum_{i=1}^n \hat{e}_i^2}{\sum_{i=1}^n \hat{e}_i^2}\right)\frac{n-k}{k_2} \\
&= 50\left(\frac{106 - 100}{100}\right) \\
&= 3,
\end{aligned}$$

where k_2 is the dimension of x_{2i}, which is 3 since (2) has 3 more coefficients than (1). A 5% asymptotic Wald test compares this with the 5% critical value of the χ_3^2 distribution, which is about 7.8. Since 3 is less than 7.8, you don't reject. While you may have memorized this, the mean of χ_3^2 is 3, so the observed value of 3 is certainly less than the 5% quantile. Alternatively, the 5% critical value of the χ_1^2 is $1.96^2 = 3.86$, which must be smaller than the critical value of the χ_3^2 distribution, so it is easy to conclude that the observed value of 3 is smaller than the critical value.

[2] In contrast, model selection asks which model fits better.

As an alternative to the Wald statistic, you could compute the F statistic, and reject the hypothesis if the F statistic exceeds the 5% critical value from the F distribution with degrees of freedom — 3,42. This is appropriate if you add the additional assumption that the error is independent of the regressors and Gaussian. □

---

11. [23.4, Hansen] *Take the model* $Y = \beta_1 \exp(\beta_2 X) + e$ *with* $\mathbb{E}[e \mid X] = 0$.

(a) *Are the parameters* (β_1, β_2) *identified?*

Solution. No. Assume the model is correctly specified, so there exists a parameter value $\beta^* = (\beta_1^*, \beta_2^*)$ satisfying $\mathbb{E}[Y|X = x] = \beta_1^* \exp(\beta_2^* X)$. The parameters (β_1, β_2) are point identified if there is a unique (β_1, β_2) such that $\beta_1 \exp(\beta_2 X) = \beta_1^* \exp(\beta_2^* X)$. In other words, if $\beta_1 \exp(\beta_2 X) \neq \beta_1^* \exp(\beta_2^* X)$ whenever $(\beta_1, \beta_2) \neq (\beta_1^*, \beta_2^*)$. Fix $\beta_1 = \beta_1^* = 0$. For any $\beta_2, \beta_2^* \in \mathbb{R}$ we have $\beta_1 \exp(\beta_2 X) = \beta_1^* \exp(\beta_2^* X) = 0$. Therefore the parameters (β_1, β_2) are not identified in general. □

(b) *Find an expression to calculate the covariance matrix of the NLLS estimators* $(\hat{\beta}_1, \hat{\beta}_2)$.

Solution. We know that the asymptotic covariance matrix of NLLS estimators is given by

$$\begin{aligned} \text{Avar}[\hat{\beta}] &= Q^{-1}\Omega Q^{-1} \\ &= \mathbb{E}[m_{\beta i} m_{\beta i}']^{-1} \mathbb{E}[m_{\beta i} m_{\beta i}' e_i^2] \mathbb{E}[m_{\beta i} m_{\beta i}']^{-1}, \end{aligned}$$

where $m_{\beta i} \equiv \partial m(x_i, \beta^*)/\partial \beta$. For $m(x, \beta) = \beta_1 \exp(\beta_2 x)$ we have

$$m_{\beta i} = \begin{bmatrix} \exp(\beta_2^* x_i) \\ \beta_1^* \exp(\beta_2^* x_i) x_i \end{bmatrix}.$$

The estimate $\hat{m}_{\beta i}$ is obtained by replacing (β_1, β_2) with the NLLS estimator $(\hat{\beta}_1, \hat{\beta}_2)$. The covariance matrix components are then estimated as

$$\begin{aligned} \hat{Q} &= \frac{1}{n} \sum_{i=1}^{n} \hat{m}_{\theta i} \hat{m}_{\theta i}', \\ \text{and } \hat{\Omega} &= \frac{1}{n} \sum_{i=1}^{n} \hat{m}_{\theta i} \hat{m}_{\theta i}' \hat{e}_i^2, \end{aligned}$$

where $\hat{e}_i = Y_i - m(X_i, \hat{\theta})$ are the NLLS residuals. Thus,

$$\begin{aligned} \hat{\text{Avar}}[\hat{\beta}] &= \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1} \\ &= \left[\frac{1}{n} \sum_{i=1}^{n} \hat{m}_{\theta i} \hat{m}_{\theta i}'\right]^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \hat{m}_{\theta i} \hat{m}_{\theta i}' \hat{e}_i^2\right) \left[\frac{1}{n} \sum_{i=1}^{n} \hat{m}_{\theta i} \hat{m}_{\theta i}'\right]^{-1}. \end{aligned}$$

□

---

12. [25.10, Hansen] *Find the first-order condition for the probit MLE $\hat{\beta}^{probit}$.*

Solution. Let (Y, X) be random with $Y \in \{0, 1\}$ and $X \in \mathbb{R}^k$. The Probit model is

$$P[Y = 1|X = x] = \Phi(x'\beta),$$

where $\Phi(u)$ is the standard normal distribution function. To construct the likelihood we need the distribution of an individual observation. Recall that if Y is Bernoulli, such that $P[Y = 1] = p$ and $P[Y = 0] = 1 - p$, then Y has the probability mass function

$$\pi(y) = p^y(1-p)^{1-y}, \quad y \in \{0, 1\}.$$

In the Probit model, Y is conditionally Bernoulli, so its conditional probability mass function is

$$\pi(Y|X) = \Phi(X'\beta)^Y[1 - \Phi(X'\beta)]^{1-Y} = \Phi(X'\beta)^Y\Phi(-X'\beta)^{1-Y} = \Phi(Z'\beta),$$

where $Z = X$ if $Y = 1$ and $Z = -X$ if $Y = 0$. Taking logs and summing across observations we obtain the log-likelihood function:

$$L_n(\beta) = \sum_{i=1}^{n} \log \Phi(Z_i'\beta).$$

The likelihood score is

$$S_n(\beta) = \frac{\partial}{\partial\beta}L_n(\beta) = \sum_{i=1}^{n} Z_i \frac{\phi(Z_i'\beta)}{\Phi(Z_i'\beta)} = \sum_{i=1}^{n} Z_i\lambda(Z_i'\beta),$$

where $\lambda(Z_i'\beta) \equiv \phi(Z_i'\beta)/\Phi(Z_i'\beta)$ is known as the inverse Mills ratio. Therefore the first-order conditions are

$$\sum_{i=1}^{n} Z_i\lambda(Z_i'\beta) = 0.$$

□

13. [17.2, LNs] *Consider the model*

$$y_i = I(x_i'\beta^* + u_i > 0),$$

where β^* *is the true parameter and* (x_i', u_i) *are iid with* $u_i|x_i$ *having a* $N(0,1)$ *distribution with cdf* Φ.

(a) *Find* $\mathbb{E}[y_i|x_i;\beta]$*, the conditional expectation of* y_i *given* x_i.

Solution. The event $y_i = 1$ is the same as $x_i'\beta^* + u_i > 0$. Thus

$$\begin{aligned} P[y_i = 1 \mid x_i] &= P[x_i'\beta^* + u_i > 0 \mid x_i] = P[u_i > -x_i'\beta^* \mid x_i] \\ &= 1 - P[u_i < -x_i'\beta^* \mid x_i] = 1 - \Phi(-x_i'\beta^*) = \Phi(x_i'\beta^*). \end{aligned}$$

Therefore

$$\mathbb{E}[y_i|x_i;\beta] = \Phi(x_i'\beta^*) \cdot 1 + [1 - \Phi(x_i'\beta^*)] \cdot 0 = \Phi(x_i'\beta^*).$$

□

(b) *Find* $V[y_i \mid x_i;\beta]$*, the conditional variance of* y_i *given* x_i*. Hint:* y_i *is a Bernoulli random variable.*

Solution. Observe that the event $y_i^2 = 1$ is also the same as $x_i'\beta^* + u_i > 0$. Therefore just as in **(a)** we have $\mathbb{E}[y_i^2|x_i;\beta] = \Phi(x_i'\beta^*)$. It follows that

$$\begin{aligned} V[y_i|x_i;\beta] &= \mathbb{E}[y_i^2|x_i;\beta] - \mathbb{E}[y_i|x_i;\beta]^2 \\ &= \Phi(x_i'\beta^*) - \Phi(x_i'\beta^*)^2 \\ &= \Phi(x_i'\beta^*)[1 - \Phi(x_i'\beta^*)]. \end{aligned}$$

□

(c) *Consider the estimator* $\hat\beta$ *which minimizes*

$$\sum_{i=1}^{n} (y_i - \mathbb{E}[y_i|x_i,\beta])^2.$$

Find the limiting distribution of $\sqrt{n}(\hat\beta - \beta^*)$.

Solution. Notice that since $\mathbb{E}[y_i|x_i,\beta] = \Phi(x_i'\beta)$, $\hat\beta$ can be understood as a nonlinear least squares estimator with $m(x_i,\beta) = \Phi(x_i'\beta)$ for a Probit model $y_i = \Phi(x_i'\beta) + e_i$. Therefore

$$\begin{aligned} \text{Avar}[\hat\beta] &= Q^{-1}\Omega Q^{-1} \\ &= \mathbb{E}[m_{\beta i}m_{\beta i}']^{-1}\mathbb{E}[m_{\beta i}m_{\beta i}'e_i^2]\mathbb{E}[m_{\beta i}m_{\beta i}']^{-1}, \end{aligned}$$

where $m_{\beta i} \equiv \partial m(x_i,\beta^*)/\partial\beta$. For $m(x_i,\beta) = \Phi(x_i'\beta)$ we have

$$m_{\beta i} = x_i\phi(x_i'\beta^*).$$

Thus, by observing that $V[y_i|x_i;\beta] = V[e_i|x_i;\beta] = \Phi(x_i'\beta^*)[1-\Phi(x_i'\beta^*)]$,

$$\begin{aligned}
\sqrt{n}(\hat\beta-\beta^*) &\xrightarrow{d} N(0,\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2e_i^2x_ix_i']\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1})\\
\text{(LIE)} &= N(0,\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2V[y_i|x_i;\beta^*]x_ix_i']\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1})\\
&= N(0,\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2\Phi(x_i'\beta^*)[1-\Phi(x_i'\beta^*)]x_ix_i']\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}).
\end{aligned}$$

□

(d) *Consider the estimator $\tilde\beta$ which minimizes*

$$\sum_{i=1}^n \frac{(y_i-\mathbb{E}[y_i|x_i,\beta])^2}{V[y_i|x_i;\beta^*]}.$$

Find the limiting distribution of $\sqrt{n}(\tilde\beta-\beta^*)$.

Solution. Observe that

$$\sum_{i=1}^n \frac{(y_i-\mathbb{E}[y_i|x_i,\beta])^2}{V[y_i|x_i;\beta^*]} = \sum_{i=1}^n \left(\frac{y_i}{V[y_i|x_i;\beta^*]^{1/2}} - \frac{\mathbb{E}[y_i|x_i,\beta]}{V[y_i|x_i;\beta^*]^{1/2}}\right)^2.$$

In a similar fashion to what was done in item **(c)**, $\tilde\beta$ can be understood as a nonlinear least squares estimator for a rescaled (by $V[y_i|x_i;\beta^*]^{1/2}$) Probit model

$$\frac{y_i}{V[y_i|x_i;\beta^*]^{1/2}} = \frac{\Phi(x_i'\beta)}{V[y_i|x_i;\beta^*]^{1/2}} + \varepsilon_i.$$

.

This is equivalent to

$$y_i = \Phi(x_i'\beta) + V[y_i|x_i;\beta^*]^{1/2}\varepsilon_i.$$

Thus from **(c)** we have the relation $e_i = V[y_i|x_i;\beta^*]^{1/2}\varepsilon_i$, whence it follows that

$$\sqrt{n}(\hat\beta-\beta^*) \xrightarrow{d} N(0,\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2V[y_i|x_i;\beta^*]\varepsilon_i^2x_ix_i']\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}).$$

But notice that $V[\varepsilon_i] = V[e_i]/V[y_i|x_i;\beta^*] = V[y_i|x_i;\beta^*]/V[y_i|x_i;\beta^*] = 1$. Therefore by LIE

$$\begin{aligned}
\sqrt{n}(\hat\beta-\beta^*) &\xrightarrow{d} N(0,\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2V[y_i|x_i;\beta^*]x_ix_i']\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1})\\
&= N(0,\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2\Phi(x_i'\beta^*)[1-\Phi(x_i'\beta^*)]\mathbb{E}[\phi(x_i'\beta^*)^2x_ix_i']^{-1}).
\end{aligned}$$

This is exactly the same asymptotic distribution obtained in item **(c)**. □

(e) *Compare the asymptotic variance of $\hat\beta$ and $\tilde\beta$. Explain your answer.*

Solution. The asymptotic variances of $\hat\beta$ and $\tilde\beta$ are exactly the same. Division by $V[y_i|x_i;\beta^*]$ just provides a useful normalization that makes the error variance unitary. □
