1 2011 Theory II Problem #1

1. For a given $i = 1, \ldots, n$, let X_i and Y_i be independent exponential random variables with means $1/(\psi\lambda_i)$ and $1/\lambda_i$, respectively. Assume further that the bivariate random vectors (X_i, Y_i) are independent for $i = 1, \ldots, n$. The parameter of interest is ψ, with the λ_i's being unknown parameters which may vary for $i = 1, \ldots, n$.

   (a) Write the log-likelihood function $L_1(\psi, \lambda_1, \ldots, \lambda_n)$ based on $(X_i, Y_i),\ i = 1, \ldots, n$. Derive the score equation that defines the maximum likelihood estimator for ψ based on L_1. Denote that equation by $U_1(\psi) = 0$.

$$L_n(\psi, \boldsymbol{\lambda}) = \prod_{i=1}^n p(x_i)\,p(y_i) = \prod_{i=1}^n \psi\lambda_i e^{-\psi\lambda_i x_i}\lambda_i e^{-\lambda_i y_i} = \prod_{i=1}^n \psi\lambda_i^2 \exp\left\{-\psi\lambda_i x_i - \lambda_i y_i\right\}$$

$$\begin{aligned}\ell_n(\psi, \boldsymbol{\lambda}) &= \sum_{i=1}^n \left\{-\psi\lambda_i x_i - \lambda_i y_i + \log(\psi) + 2\log(\lambda_i)\right\} \\ &= \sum_{i=1}^n \left\{-\psi\lambda_i x_i - \lambda_i y_i + 2\log(\lambda_i)\right\} + n\log(\psi)\end{aligned}$$

   Then

$$\frac{d\ell_n(\psi, \boldsymbol{\lambda})}{d\psi} = -\sum_{i=1}^n \lambda_i x_i + \frac{n}{\psi}$$

$$\frac{d\ell_n(\psi, \boldsymbol{\lambda})}{d\lambda_i} = -\psi x_i - y_i + \frac{2}{\lambda_i}$$

   Then the score equation that defines the MLE for ψ is given by

$$0 = -\sum_{i=1}^n \widehat{\lambda}_i x_i + \frac{n}{\psi} \equiv U_1(\psi)$$

   where $\widehat{\lambda}_i$ denotes the MLE of $\lambda_i,\ i = 1, \ldots, n$.

   Is this really what the question is asking for here?

   (b) Are the standard regularity conditions for the consistency and asymptotic normality of the maximum likelihood estimators for $\psi, \lambda_1, \ldots, \lambda_n$, based on $L_1(\psi, \lambda_1, \ldots, \lambda_n)$ satisfied in this problem?

   No. MLE theory depends on a fixed number of parameters.

   (c) Assuming the regularity conditions for the maximum likelihood estimators from L_1 are satisfied, derive an explicit expression for the asymptotic variance of $\widehat{\psi}$ via the Fisher information matrix from L_1.

   We wish to calculate $\boldsymbol{I}_n(\psi, \boldsymbol{\lambda})$. We observe that

$$\frac{d^2}{d\psi^2}\ell_n(\psi, \boldsymbol{\lambda}) = -\frac{n}{\psi^2} \qquad \frac{d^2}{d\psi\,d\lambda_i}\ell_n(\psi, \boldsymbol{\lambda}) = -x_i$$

$$\frac{d^2}{d\lambda_i^2}\ell_n(\psi, \boldsymbol{\lambda}) = -\frac{2n}{\lambda_i^2} \qquad \frac{d^2}{d\lambda_i\,d\lambda_j}\ell_n(\psi, \boldsymbol{\lambda}) = 0, \quad i \neq j$$

Then

$$\boldsymbol{I}_n(\psi,\boldsymbol{\lambda}) = \begin{bmatrix} I_{11} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix}$$

where

$$I_{11} = \frac{n}{\psi^2}, \qquad \boldsymbol{I}_{12} = \boldsymbol{I}_{21}' = \left(\frac{1}{\psi^2\lambda_1}, \ldots, \frac{1}{\psi^2\lambda_n}\right), \qquad \boldsymbol{I}_{22} = \text{diag}\left(\frac{2}{\lambda_1^2}, \ldots, \frac{2}{\lambda_1^2}\right)$$

Next we want to calculate the $(1,1)$ entry of $\left[\lim_{n\to\infty}\frac{1}{n}\boldsymbol{I}_n(\psi,\boldsymbol{\lambda})\right]^{-1}$. We see that $\lim_{n\to\infty}\frac{1}{n}I_{11} = \frac{1}{\psi^2}$ and that the rest of $\lim_{n\to\infty}\frac{1}{n}\boldsymbol{I}_n(\psi,\boldsymbol{\lambda})$ is an infinite matrix of 0's; however this is clearly a singular matrix and consequently $\left[\lim_{n\to\infty}\frac{1}{n}\boldsymbol{I}_n(\psi,\boldsymbol{\lambda})\right]^{-1}$ is not defined. Thus we presume that we can understand the prompt as directing us to assume that the asymptotic variance of $\widehat{\psi}$ based on $U_1(\psi)$ as being given by $\left[\lim_{n\to\infty}\frac{1}{n}I_{11}\right]^{-1} = \psi^2$.

(d) Show that $U_1(\psi)$ depends on the data only through the ratios $T_i = X_i/Y_i$. Derive the pdf of T_i and show that it does not depend on λ_i.

Using results from (1a) we have

$$\frac{d\ell_n(\psi,\boldsymbol{\lambda})}{d\psi} = -\sum_{i=1}^n \lambda_i x_i + \frac{n}{\psi} \overset{set}{=} 0 \quad\Longrightarrow\quad \widehat{\psi}^{-1} = \frac{1}{n}\sum_{i=1}^n \widehat{\lambda}_i x_i$$

$$\frac{d\ell_n(\psi,\boldsymbol{\lambda})}{d\lambda_i} = -\psi x_i - y_i + \frac{2}{\lambda_i} \overset{set}{=} 0 \quad\Longrightarrow\quad \widehat{\lambda}_i = \frac{2}{\widehat{\psi} x_i + y_i}$$

Then

$$\widehat{\psi}^{-1} = \frac{1}{n}\sum_{i=1}^n \widehat{\lambda}_i x_i = \frac{1}{n}\sum_{i=1}^n x_i \frac{2}{\widehat{\psi} x_i + y_i} = \frac{2}{n}\sum_{i=1}^n \frac{x_i}{\widehat{\psi} x_i + y_i} = \frac{2}{n}\sum_{i=1}^n \left[\widehat{\psi} + \frac{y_i}{x_i}\right]^{-1}$$

so we see that $U_1(\psi)$ depends on the data only through X_i/Y_i, $i = 1, \ldots, n$.

Next, let $S_i = X_i/Y_i$ and $T = Y_i$. Then $S > 0$, $T > 0$ and $X_i = ST$, $Y_i = T$. Also

$$J(s,t) = \begin{vmatrix} t & s \\ 0 & 1 \end{vmatrix} = t$$

so that

$$f_{S,T}(s,t) = f_{X_i,Y_i}(st,t)\,|t| = \psi\lambda_i e^{-\psi\lambda_i st}\lambda_i e^{-\lambda_i t} t = \psi\lambda_i^2 t \exp\left\{-\lambda_i[\psi s + 1]t\right\}, \qquad s,t > 0$$

Then

$$f_S(s) = \int_0^\infty f_{S,T}(s,t)\,dt = \int_0^\infty \psi\lambda_i^2 t\exp\left\{-\lambda_i[\psi s+1]t\right\}dt$$

$$= \psi\lambda_i^2 \frac{\Gamma(2)}{\left(\lambda_i[\psi s+1]\right)^2}\int_0^\infty \frac{\left(\lambda_i[\psi s+1]\right)^2}{\Gamma(2)} t^{2-1}\exp\left\{-\lambda_i[\psi s+1]t\right\}dt$$

$$= \psi\lambda_i^2 \frac{\Gamma(2)}{\left(\lambda_i[\psi s+1]\right)^2} = \frac{\psi}{(\psi s+1)^2}, \qquad s > 0$$

(is there a mistake in this part?)

(e) Use the density of $T_1, \ldots, T_n$ to obtain a likelihood function, $L_2(\psi)$. Compare the score equation derived for ψ from L_2 with the function $U_1(\psi)$ derived in part (a). Is the maximum likelihood estimator for ψ from L_2 identical to that from L_1? Derive the asymptotic variance for the maximimum likelihood estimator based on L_2 using standard asymptotic calculations and compare with that in part (c).

$$L_2(\psi) = \prod_{i=1}^{n} f(t_i; \psi) = \prod_{i=1}^{n} \frac{\psi}{(\psi t_i + 1)^2} = \frac{\psi^n}{\prod_{i=1}^{n} (\psi t_i + 1)^2}$$

$$\ell_2(\psi) = n \log \psi - 2 \sum_{i=1}^{n} \log(\psi t_i + 1)$$

Then the score equation is given by

$$\frac{d}{d\psi} \ell_2(\psi) = \frac{d}{d\psi} \left\{ n \log \psi - 2 \sum_{i=1}^{n} \log(\psi t_i + 1) \right\} = \frac{n}{\psi} - 2 \sum_{i=1}^{n} \frac{t_i}{\psi t_i + 1} \stackrel{set}{=} 0$$

$$\implies \widehat{\psi}^{-1} = \frac{2}{n} \sum_{i=1}^{n} \frac{t_i}{\psi t_i + 1} = \frac{2}{n} \sum_{i=1}^{n} \left[\widehat{\psi} + \frac{1}{t_i} \right]^{-1} = \frac{2}{n} \sum_{i=1}^{n} \left[\widehat{\psi} + \frac{y_i}{x_i} \right]^{-1}$$

We see that the MLE for ψ based on L_2 is the same as the MLE for ψ based on L_1.

Next we wish to calculate the asymptotic variance of the MLE based on L_2. We have

$$\frac{d^2}{d\psi^2} \ell_n(\psi) = \frac{d}{d\psi} \left\{ \frac{n}{\psi} - 2 \sum_{i=1}^{n} \frac{t_i}{\psi t_i + 1} \right\} = -\frac{n}{\psi^2} + 2 \sum_{i=1}^{n} \frac{t_i^2}{(\psi t_i + 1)^2}$$

Now

$$\mathbb{E}\left[\frac{T_i^2}{(\psi T_i + 1)^2} \right] = \int_0^{\infty} \frac{t^2}{(\psi t + 1)^2} \frac{\psi}{(\psi t + 1)^2} \, dt$$

$$= \psi \int_0^{\infty} \frac{t^2}{(\psi t + 1)^4} \, dt$$

$$= \psi \int_1^{\infty} \frac{[\psi^{-1}(w-1)]^2}{w^4} \frac{dw}{\psi}$$

$$= \frac{1}{\psi^2} \int_1^{\infty} \frac{(w-1)^2}{w^4} \, dw$$

$$= \frac{1}{\psi^2} \int_1^{\infty} \left[w^{-2} - 2w^{-3} + w^{-4} \right] dw$$

$$= \frac{1}{\psi^2} \left\{ -\frac{1}{w} + \frac{1}{w^2} - \frac{1}{3w^3} \right\} \Bigg|_{w=1}^{w=\infty}$$

$$= \frac{1}{\psi^2} \left\{ 1 - 1 + \frac{1}{3} \right\} = \frac{1}{3\psi^2}$$

Therefore,

$$\mathbb{E}\left[-\frac{d^2}{d\psi^2}\,\ell_n(\psi)\right] = \frac{n}{\psi^2} - \frac{2n}{3\psi^2} = \frac{n}{3\psi^2}$$

Thus we conclude that under the appropriate regularity conditions the asymptotic variance of the estimator of ψ based on L_2 has the value

$$\left\{\lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}\left[-\frac{d^2}{d\psi^2}\,\ell_n(\psi)\right]\right\}^{-1} = \left[\lim_{n\to\infty} \frac{1}{n}\,\frac{n}{3\psi^2}\right]^{-1} = 3\psi^2$$

(f) Let $g_i(\psi) = Y_i - \psi X_i,\ i = 1,\ldots,n$ and consider estimation of ψ by solving

$$\sum_{i=1}^{n} w_i g_i(\psi) = 0$$

for ψ, where $w_i,\ i = 1,\ldots,n$ are finite constants. Determine the asymptotic variance of the estimator thus obtained and find the optimal w_i's (up to a proportionality constant). Compare the efficiency of this optimal estimator to that based on $w_i = 1,\ i = 1,\ldots,n$ and to that from $U_1(\psi)$. Is the optimal estimator usable in practice?

Define $s_n(\psi) = \sum_{i=1}^{n} w_i g_i(\psi)$. Under appropriate regularity conditions we have

$$V_n^{-1/2}(\widehat{\psi}_n - \psi) \xrightarrow{L} N(0,1)$$

where

$$V_n = \frac{\mathrm{Var}\left[s_n(\psi)\right]}{\left(\mathbb{E}\left[-d_\psi s_n(\psi)\right]\right)^2}$$

We begin by calculating V_n. Notice that $\psi X_i \sim exp(\lambda_i)$. We have

$$\mathrm{Var}\left[s_n(\psi)\right] = \mathrm{Var}\left[\sum_{i=1}^{n} w_i(Y_i - \psi X_i)\right] = \sum_{i=1}^{n} w_i^2\Big(\mathrm{Var}\left[Y_i\right] + \mathrm{Var}\left[\psi X_i\right]\Big)$$

$$= \sum_{i=1}^{n} w_i^2\left(\frac{1}{\lambda_i^2} + \frac{1}{\lambda_i^2}\right) = 2\sum_{i=1}^{n} \frac{w_i^2}{\lambda_i^2}$$

Next,

$$\mathbb{E}\left[-\frac{d}{d\psi}\,s_n(\psi)\right] = \mathbb{E}\left[-\frac{d}{d\psi}\sum_{i=1}^{n} w_i(Y_i - \psi X_i)\right] = \mathbb{E}\left[\sum_{i=1}^{n} w_i X_i\right] = \frac{1}{\psi}\sum_{i=1}^{n}\frac{w_i}{\lambda_i}$$

Combining these results we see that

$$V_n = 2\psi^2\,\frac{\displaystyle\sum_{i=1}^{n}\frac{w_i^2}{\lambda_i^2}}{\left(\displaystyle\sum_{i=1}^{n}\frac{w_i}{\lambda_i}\right)^2}$$

Next we wish to minimize V_n with respect to $w_1, \ldots, w_n$. We have

$$\frac{d}{dw_i} V_n = 2\psi^2 \frac{\frac{2w_i}{\lambda_i^2}\left(\sum_{i=1}^{n} \frac{w_i}{\lambda_i}\right)^2 - \frac{2}{\lambda_i}\left(\sum_{i=1}^{n} \frac{w_i^2}{\lambda_i^2}\right)\sum_{i=1}^{n} \frac{w_i}{\lambda_i}}{\left(\sum_{i=1}^{n} \frac{w_i}{\lambda_i}\right)^4} = \frac{4\psi^2}{\lambda_i} \frac{\frac{w_i}{\lambda_i}\sum_{i=1}^{n} \frac{w_i}{\lambda_i} - \sum_{i=1}^{n} \frac{w_i^2}{\lambda_i^2}}{\left(\sum_{i=1}^{n} \frac{w_i}{\lambda_i}\right)^3}, \quad i = 1, \ldots, n$$

so that

$$\frac{d}{dw_i} V_n \overset{set}{=} 0 \quad \Longrightarrow \quad 0 = \left[\left(\sum_{i=1}^{n} \frac{w_i}{\lambda_i}\right)^3 \frac{\lambda_i}{4\psi^2}\right] \frac{d}{dw_i} V_n = \frac{w_i}{\lambda_i}\sum_{i=1}^{n} \frac{w_i}{\lambda_i} - \sum_{i=1}^{n} \frac{w_i^2}{\lambda_i^2}, \quad i = 1, \ldots, n$$

Thus by subtracting equations, we see that

$$\begin{aligned} 0 &= \left[\frac{w_i}{\lambda_i}\sum_{i=1}^{n} \frac{w_i}{\lambda_i} - \sum_{i=1}^{n} \frac{w_i^2}{\lambda_i^2}\right] - \left[\frac{w_{i+1}}{\lambda_{i+1}}\sum_{i=1}^{n} \frac{w_i}{\lambda_i} - \sum_{i=1}^{n} \frac{w_i^2}{\lambda_i^2}\right] \\ &= \left(\frac{w_i}{\lambda_i} - \frac{w_{i+1}}{\lambda_{i+1}}\right)\sum_{i=1}^{n} \frac{w_i}{\lambda_i} \quad \Longrightarrow \quad \frac{w_i}{\lambda_i} = \frac{w_{i+1}}{\lambda_{i+1}}, \quad i = 1, \ldots, n-1 \end{aligned}$$

so that

$$\frac{w_1}{\lambda_1} = \frac{w_2}{\lambda_2} = \ldots = \frac{w_k}{\lambda_k} \overset{set}{=} k \quad \Longrightarrow \quad w_i = k\lambda_i, \quad i = 1, \ldots, n$$

which is the (non-unique) optimal choice of $\boldsymbol{w}$. Then

$$\min_{w} V_n = \frac{2\psi^2 \sum_{i=1}^{n} \frac{(k\lambda_i)^2}{\lambda_i^2}}{\left(\sum_{i=1}^{n} \frac{k\lambda_i}{\lambda_i}\right)^2} = \frac{2\psi^2 n k^2}{n^2 k^2} = \frac{2\psi^2}{n}$$

It follows that under appropriate regularity conditions,

$$\sqrt{n}(\widehat{\psi}_n - \psi) \xrightarrow{L} N\left(0, \left[\lim_{n\to\infty} \frac{1}{n}\left(\min_{w} V_n\right)^{-1}\right]^{-1}\right) \overset{d}{=} N(0, 2\psi^2)$$

Recall that we concluded in (1c) that that asymptotic variance of the estimator for ψ obtained through $U_1(\psi)$ was given by ψ^2. It follows that if in fact the asymptotic variance calculation made for this estimator was legitimate, that the estimator would be twice as efficient as the optimal estimator obtained through $s_n(\psi)$. Of course the calculation is not legitimate, and furthermore we have not even verified that the estimator obtained through $U_1(\psi)$ is consistent.

The optimal estimator obtained through $s_n(\psi)$ is not usable in practice, since the value of $\boldsymbol{\lambda}$ is not known.

2 2011 Theory II Problem #2

2. In this problem, we consider the univariate density

$$p(y;\,\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,y^{\alpha-1}(1-y)^{\beta-1}, \qquad 0<y<1 \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function, and $\alpha > 0$ and $\beta > 0$. One may reparamterize (1) in terms of (μ,ϕ), such that

$$p(y;\,\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma\big((1-\mu)\phi\big)}\,y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \qquad 0<y<1 \tag{2}$$

where $0<\mu<1$, $\phi>0$, $\mathbb{E}[y]=\mu$, and $\mathrm{Var}[y]=\mu(1-\mu)/(1+\phi)$

(a) Find explicit expressions for (μ,ϕ) in terms of (α,β).

We have

$$\begin{cases} \phi &= \alpha+\beta \\ \mu\phi &= \quad\alpha \end{cases} \qquad\Longrightarrow\qquad \mu(\alpha+\beta)=\alpha \qquad\Longrightarrow\qquad \mu=\frac{\alpha}{\alpha+\beta}$$

so that

$$(\mu,\phi) = \left(\frac{\alpha}{\alpha+\beta},\ \alpha+\beta\right)$$

(b) Let $Y_1,\ldots,Y_n$ be a random sample from the density in (2). Show that the joint density for $Y_1,\ldots,Y_n$ belongs to the multivariate exponential family of distributions identify the canonical statistics and parameters, determine its rank, and find the joint complete sufficient statistics for (μ,ϕ).

$$\begin{aligned}
p(\mathbf{y};\,\mu,\phi) &= \prod_{i=1}^{n} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma\big((1-\mu)\phi\big)}\,y_i^{\mu\phi-1}(1-y_i)^{(1-\mu)\phi-1} \\
&= \left[\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma\big((1-\mu)\phi\big)}\right]^n \left(\prod_{i=1}^{n} y_i\right)^{\mu\phi-1} \left(\prod_{i=1}^{n}(1-y_i)\right)^{(1-\mu)\phi-1} \\
&= \exp\left\{(\mu\phi-1)\sum_{i=1}^{n}\log y_i + \big[(1-\mu)\phi-1\big]\sum_{i=1}^{n}\log(1-y_i) + n\log\left[\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma\big((1-\mu)\phi\big)}\right]\right\} \\
&= \exp\left\{\mu\phi\left[\sum_{i=1}^{n}\log y_i - \sum_{i=1}^{n}\log(1-y_i)\right] + \phi\sum_{i=1}^{n}\log(1-y_i)\right. \\
&\qquad\qquad \left. + \, n\log\left[\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma\big((1-\mu)\phi\big)}\right] - \sum_{i=1}^{n}\log y_i - \sum_{i=1}^{n}\log(1-y_i)\right\} \\
&= \exp\left\{\theta_1\left[\sum_{i=1}^{n}\log y_i - \sum_{i=1}^{n}\log(1-y_i)\right] + \theta_2\sum_{i=1}^{n}\log(1-y_i)\right. \\
&\qquad\qquad \left. + \, n\log\left[\frac{\Gamma(\theta_2)}{\Gamma(\theta_1)\,\Gamma(\theta_2-\theta_1)}\right] - \sum_{i=1}^{n}\log y_i - \sum_{i=1}^{n}\log(1-y_i)\right\}
\end{aligned}$$

$$= \exp\left\{\theta_1 \sum_{i=1}^n \log\left(\frac{y_i}{1-y_i}\right) + \theta_2 \sum_{i=1}^n \log(1-y_i) \right.$$
$$\left. + \, n \log\left[\frac{\Gamma(\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2-\theta_1)}\right] - \sum_{i=1}^n \log y_i - \sum_{i=1}^n \log(1-y_i)\right\}$$

which is the form of an exponential family of dimension 2 where $(\theta_1, \theta_2) \equiv (\mu\phi, \phi)$. Define

$$\Theta \equiv \left\{(\theta_1, \theta_2)\colon \theta_1 = \mu\phi,\, \theta_2 = \phi,\, \mu > 0,\, \phi > 0\right\} = \left\{(\theta_1, \theta_2)\colon \theta_1, \theta_2 > 0\right\}$$

It is an elementary result from analysis that the set $B(\boldsymbol{x}, \epsilon) \equiv \{\boldsymbol{x} \in \mathbb{R}^2\colon ||\boldsymbol{x}|| < \epsilon\}$ is open. Consider $\boldsymbol{\theta} = (1,1) \in \Theta$ and let $0 < \epsilon < 1$; then a simple proof-by-contradiction shows that $B(\theta, \epsilon) \subset \Theta$ and hence the family of densities in question is a full-rank exponential family (has rank 2). Define

$$\boldsymbol{y}^* \equiv \left(\sum_{i=1}^n \log\left(\frac{y_i}{1-y_i}\right), \quad \sum_{i=1}^n \log(1-y_i)\right)$$

It follows from exponential family theory that $\boldsymbol{y}^*$ is complete sufficient for $\boldsymbol{\theta}$. Furthermore, since (μ, ϕ) has a 1-1 correspondence with (θ_1, θ_2), we find that $\boldsymbol{y}^*$ is complete sufficient for (μ, ϕ).

(c) Now, suppose that $Y_1, \ldots, Y_n$ are independent random variables, where each Y_i, $i = 1, \ldots, n$, follows the density in (2) with unknown mean μ_i and unknown precision ϕ. Suppose that $\boldsymbol{X}_i$ is a $p \times 1$ vector of covariates, with $g(\mu_i) = \boldsymbol{X}_i'\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients and $g(\cdot)$ is an arbitrary known link function. Define $\xi = (\boldsymbol{\beta}, \phi)$.

(i) Derive the score function for ξ and show that the expectation of the score function equals 0 at the true value of ξ.

We wish to derive the score function for $\boldsymbol{\mu}, \phi$. We have

$$L_n(\boldsymbol{\mu}, \phi) = \prod_{i=1}^n \frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma\big((1-\mu_i)\phi\big)} y_i^{\mu_i\phi-1}(1-y_i)^{(1-\mu_i)\phi-1}$$
$$= \prod_{i=1}^n \left(\frac{y_i}{1-y_i}\right)^{\mu_i\phi-1} (1-y_i)^{\phi} \frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma\big((1-\mu_i)\phi\big)}$$

and

$$\ell_n(\boldsymbol{\mu}, \phi) = \sum_{i=1}^n \left\{(\phi\mu_i - 1)\log\left(\frac{y_i}{1-y_i}\right) + \phi\log(1-y_i) \right.$$
$$\left. + \log\left[\Gamma(\phi)\right] - \log\left[\Gamma(\mu_i\phi)\right] - \log\left[\Gamma\big((1-\mu_i)\phi\big)\right]\right\}$$

Define $\psi(z) = \dfrac{d\log\Gamma(z)}{dz}$. Notice that for $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ and $g(\mu_i) = \eta_i$

$$\frac{d\mu_i}{d\eta_i} = \left(\frac{d\eta_i}{d\mu_i}\right)^{-1} = \left(\frac{dg(\mu_i)}{d\mu_i}\right)^{-1} = \frac{1}{\dot{g}(\mu_i)}$$

Then

$$
\begin{aligned}
\frac{d}{d\boldsymbol{\beta}}\,\ell_n(\boldsymbol{\mu},\phi) &= \sum_{i=1}^{n}\frac{d}{d\boldsymbol{\beta}}\,\ell_i(\mu_i,\phi)\\
&= \sum_{i=1}^{n}\frac{d\ell_i(\mu_i,\phi)}{d\mu_i}\,\frac{d\mu_i}{\eta_i}\,\frac{d\eta_i}{d\boldsymbol{\beta}}\\
&= \sum_{i=1}^{n}\left\{\phi\,\log\left(\frac{y_i}{1-y_i}\right)-\frac{d}{d\mu_i}\big[\Gamma(\mu_i\phi)\big]-\frac{d}{d\mu_i}\,\log\big[\Gamma\big((1-\mu_i)\phi\big)\big]\right\}\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i\\
&= \sum_{i=1}^{n}\left\{\phi\,\log\left(\frac{y_i}{1-y_i}\right)-\frac{d\log\big[\Gamma(\mu_i\phi\big]}{d(\mu_i\phi)}\,\frac{d\mu_i\phi}{d\mu_i}-\frac{d\log\big[\Gamma\big((1-\mu_i)\phi\big)\big]}{d\big((1-\mu_i)\phi\big)}\,\frac{d(1-\mu_i)\phi}{d\mu_i}\right\}\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i\\
&= \phi\sum_{i=1}^{n}\left\{\log\left(\frac{y_i}{1-y_i}\right)-\psi(\mu_i\phi)+\psi\big[(1-\mu_i)\phi\big]\right\}\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{d}{d\phi}\,\ell_n(\boldsymbol{\mu},\phi) &= \sum_{i=1}^{n}\Bigg\{\mu_i\log\left(\frac{y_i}{1-y_i}\right)+\log(1-y_i)+\frac{d}{d\phi}\,\log\big[\Gamma(\phi)\big]\\
&\qquad -\frac{d}{d\phi}\log\big[\Gamma(\mu_i\phi)\big]-\frac{d}{d\phi}\log\big[\Gamma\big((1-\mu_i)\phi\big)\big]\Bigg\}\\
&= \sum_{i=1}^{n}\Bigg\{\mu_i\log\left(\frac{y_i}{1-y_i}\right)+\log(1-y_i)+\frac{d}{d\phi}\,\log\big[\Gamma(\phi)\big]\\
&\qquad -\frac{d\log\big[\Gamma(\mu_i\phi)\big]}{d(\mu_i\phi)}\,\frac{d\mu_i\phi}{d\phi}-\frac{d\log\big[\Gamma\big((1-\mu_i)\phi\big)\big]}{d\phi}\,\frac{d(1-\mu_i)\phi}{d\phi}\Bigg\}\\
&= \sum_{i=1}^{n}\left\{\mu_i\log\left(\frac{y_i}{1-y_i}\right)+\log(1-y_i)+\psi(\phi)-\mu_i\,\psi(\mu_i\phi)-(1-\mu_i)\,\psi\big[(1-\mu_i)\phi\big]\right\}
\end{aligned}
$$

So we have found the score function. Now we want to show that the expectation of this function is $\mathbf{0}$. We have

$$
\mathbb{E}\left[\frac{d}{d(\boldsymbol{\beta},\boldsymbol{\phi})}\,\ell_n(\boldsymbol{\mu},\phi)\right]=\mathbb{E}\left\{\sum_{i=1}^{n}\left[\frac{d\boldsymbol{\theta}_i}{d(\boldsymbol{\beta},\phi)}\right]'\frac{d\ell_i(\mu_i,\phi)}{d\theta_i}\right\}=\sum_{i=1}^{n}\left[\frac{d\boldsymbol{\theta}_i}{d(\boldsymbol{\beta},\phi)}\right]'\mathbb{E}\left[\frac{d\ell_i(\mu_i,\phi)}{d\theta_i}\right]
$$

But

$$
\mathbb{E}\left[\frac{d\ell_i(\mu_i,\phi)}{d\boldsymbol{\theta}_i}\right]=\mathbb{E}\left\{\frac{d}{d\boldsymbol{\theta}_i}\Big[\log\big[h(y_i)\big]+\big[T(Y_i)\big]'\boldsymbol{\theta}_i-b(\boldsymbol{\theta}_i)\Big]\right\}=\mathbb{E}\big[T(Y_i)\big]-\dot b(\boldsymbol{\theta}_i)=\mathbf{0}
$$

for $i=1,\ldots,n$ so that the desired result is reached.

(ii) Show that the Fisher information matrix of ξ is given by

$$
\boldsymbol{I}(\xi)=\begin{pmatrix}\boldsymbol{I}_{\beta\beta} & \boldsymbol{I}_{\beta\phi}\\ \boldsymbol{I}_{\phi\beta} & \boldsymbol{I}_{\phi\phi}\end{pmatrix}
$$

where $\boldsymbol{I}_{\boldsymbol{\beta\beta}} = \phi \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}$, $\boldsymbol{I}_{\boldsymbol{\beta}\phi} = \boldsymbol{I}'_{\phi\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{T}\boldsymbol{c}$, $\boldsymbol{I}_{\phi\phi} = \mathrm{tr}(D)$, $\boldsymbol{c} = (c_1, \ldots, c_n)'$ with $c_j = \phi\left\{\dot{\psi}(\mu_j\phi)\mu_j - \dot{\psi}\left[(1-\mu_j)\phi\right]\right\}$, $\boldsymbol{T} = \mathrm{diag}\left(\frac{1}{\dot{g}(\mu_1)}, \ldots, \frac{1}{\dot{g}(\mu_n)}\right)$, $\dot{g}(z) = \frac{d}{dz}g(z)$, $\psi(z) = \frac{d}{dz}\log\Gamma(z)$, $\dot{\psi}(z) = \frac{d}{dz}\psi(z)$, $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_n)$ with $d_j = \dot{\psi}(\mu_j\phi)\mu_j^2 + \dot{\psi}\left[(1-\mu_j)\phi\right](1-\mu_j)^2 - \dot{\psi}(\phi)$, and $\boldsymbol{W} = \mathrm{diag}(w_1, \ldots, w_n)$ with

$$w_j = \phi\left\{\dot{\psi}(\mu_j\phi) + \dot{\psi}\left[(1-\mu_j)\phi\right]\right\}$$

Let

$$\begin{cases} y_i^* = \log\left(\dfrac{y_i}{1-y_i}\right) \\ \mu_i^* = \psi(\mu_i\phi) - \psi\left[(1-\mu_i)\phi\right] \end{cases} \qquad \text{for each } i = 1, \ldots, n$$

Then from (2c.i) we have

$$\begin{aligned} \frac{d}{d\boldsymbol{\beta}}\ell_n(\boldsymbol{\mu}, \phi) &= \phi\sum_{i=1}^n \left\{\log\left(\frac{y_i}{1-y_i}\right) - \psi(\mu_i\phi) + \psi\left[(1-\mu_i)\phi\right]\right\}\frac{1}{\dot{g}(\mu_i)}\boldsymbol{x}_i \\ &= \phi\sum_{i=1}^n (y_i^* - \mu_i^*)\frac{1}{\dot{g}(\mu_i)}\boldsymbol{x}_i \\ &= \phi\boldsymbol{X}'\boldsymbol{T}\left(\boldsymbol{y}^* - \boldsymbol{\mu}^*\right) \end{aligned}$$

where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix}, \qquad \boldsymbol{T} = \mathrm{diag}\left(\frac{1}{\dot{g}(\mu_1)}, \ldots, \frac{1}{\dot{g}(\mu_n)}\right), \qquad \boldsymbol{y}^* = \begin{bmatrix} y_1^* \\ \vdots \\ y_n^* \end{bmatrix}, \qquad \boldsymbol{\mu}^* = \begin{bmatrix} \mu_1^* \\ \vdots \\ \mu_n^* \end{bmatrix}$$

Next,

$$\begin{aligned} \frac{d^2}{d\beta_j\,d\beta_k}\ell_n(\boldsymbol{\mu}, \phi) &= \sum_{i=1}^n \frac{d}{d\beta_j}\left\{\frac{d\ell_i(\mu_i, \phi)}{d\beta_k}\right\} \\ &= \sum_{i=1}^n \frac{d}{d\beta_j}\left\{\frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{d\eta_i}{d\beta_k}\right\} \\ &= \sum_{i=1}^n \frac{d}{d\beta_j}\left\{\frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\frac{d\mu_i}{d\eta_i}\right\}x_{ik} \\ &= \sum_{i=1}^n \left\{\left(\frac{d}{d\beta_j}\frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\right)\frac{d\mu_i}{d\eta_i} + \frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\left(\frac{d}{d\beta_j}\frac{d\mu_i}{d\eta_i}\right)\right\}x_{ik} && \text{(product rule)} \\ &= \sum_{i=1}^n \left\{\left(\frac{d}{d\mu_i}\frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\right)\frac{d\mu_i}{d\eta_i}\frac{d\eta_i}{d\beta_j}\frac{d\mu_i}{d\eta_i} + \frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\left(\frac{d}{d\beta_j}\frac{d\mu_i}{d\eta_i}\right)\right\}x_{ik} && \text{(chain rule)} \end{aligned}$$

$$= \sum_{i=1}^{n} \frac{d^2 \ell_i(\mu_i, \phi)}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{d\eta_i}{d\beta_j} + \frac{d\ell_i(\mu_i, \phi)}{d\mu_i} \left(\frac{d}{d\beta_j} \frac{d\mu_i}{d\eta_i}\right) x_{ik}$$

$$= \sum_{i=1}^{n} \frac{d^2 \ell_i(\mu_i, \phi)}{d\mu_i^2} \left(\frac{1}{\dot{g}(\mu_i)}\right)^2 x_{ij} x_{ik} + \sum_{i=1}^{n} \frac{d\ell_i(\mu_i, \phi)}{d\mu_i} \left(\frac{d}{d\beta_j} \frac{d\mu_i}{d\eta_i}\right) x_{ik}$$

Notice that from (3c.i) we obtain

$$\mathbb{E}\left[\frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\right] = \left(\frac{d\boldsymbol{\theta}_i}{d\mu_i}\right)' \mathbb{E}\left[\frac{d\ell_i(\mu_i, \phi)}{d\boldsymbol{\theta}_i}\right] = 0$$

Additionally (using a previous calculation)

$$\frac{d^2}{d\mu_i^2} \ell_i(\mu_i, \phi) = \phi \frac{d}{d\mu_i} \left\{\log\left(\frac{y_i}{1-y_i}\right) - \psi(\mu_i\phi) + \psi\big[(1-\mu_i)\phi\big]\right\}$$

$$= \phi \left(\frac{d\psi(\mu_i\phi)}{d(\mu_i\phi)} \frac{d\mu_i\phi}{d\mu_i} + \frac{d\psi\big[(1-\mu_i)\phi\big]}{d\big((1-\mu_i)\phi\big)} \frac{d(1-\mu_i)\phi}{d\mu_i}\right)$$

$$= \phi\Big(-\phi\, \dot{\psi}(\mu_i\phi) - \phi\, \dot{\psi}\big[(1-\mu_i)\phi\big]\Big)$$

$$= -\phi^2 \left(\dot{\psi}(\mu_i\phi) + \dot{\psi}\big[(1-\mu_i)\phi\big]\right)$$

Thus putting these last few pieces together we see that

$$\mathbb{E}\left[-\frac{d^2 \ell_n(\boldsymbol{\mu}, \phi)}{d\beta_j\, d\beta_k}\right] = \sum_{i=1}^{n} \mathbb{E}\left[-\frac{d^2 \ell_i(\mu_i, \phi)}{d\mu_i^2}\right]\left(\frac{1}{\dot{g}(\mu_i)}\right)^2 x_{ij} x_{ik} - \sum_{i=1}^{n} \mathbb{E}\left[\frac{d\ell_i(\mu_i, \phi)}{d\mu_i}\right]\left(\frac{d}{d\beta_j} \frac{d\mu_i}{d\eta_i}\right) x_{ik}$$

$$= \sum_{i=1}^{n} \phi^2 \left(\dot{\psi}(\mu_i\phi) + \dot{\psi}\big[(1-\mu_i)\phi\big]\right) x_{ij} x_{ik}$$

$$= \phi \sum_{i=1}^{n} w_i x_{ij} x_{ik}$$

where $w_i = \phi \left(\dot{\psi}(\mu_i\phi) + \dot{\psi}\big[(1-\mu_i)\phi\big]\right) \frac{1}{\dot{g}(\mu_i)}$.

Furthermore,

$$\mathbb{E}\left[-\frac{d^2 \ell_n(\boldsymbol{\mu}, \phi)}{d\boldsymbol{\beta}\, d\boldsymbol{\beta}'}\right] = \phi \boldsymbol{X}' \boldsymbol{W} \boldsymbol{X} \qquad \text{where} \quad \boldsymbol{W} = \text{diag}(w_1, \ldots, w_n)$$

Next (from a previous calculation)

$$\frac{d^2}{d\phi^2}\,\ell_n(\boldsymbol{\mu},\phi) = \frac{d}{d\phi}\sum_{i=1}^n \Big\{\mu_i \log\Big(\frac{y_i}{1-y_i}\Big) + \log(1-y_i) + \psi(\phi) - \mu_i\,\psi(\mu_i\phi) - (1-\mu_i)\,\psi\big[(1-\mu_i)\phi\big]\Big\}$$

$$= \sum_{i=1}^n \Big\{\dot\psi(\phi) - \mu_i\,\frac{d\psi(\mu_i\phi)}{d\mu_i\phi}\,\frac{d\mu_i\phi}{d\phi} - (1-\mu_i)\,\frac{d\psi\big[(1-\mu_i)\phi\big]}{d\big((1-\mu_i)\phi\big)}\,\frac{d(1-\mu_i)\phi}{d\phi}\Big\}$$

$$= \sum_{i=1}^n \Big\{\dot\psi(\phi) - \mu_i^2\,\dot\psi(\mu_i\phi) - (1-\mu_i)^2\,\dot\psi\big[(1-\mu_i)\phi\big]\Big\} = -\sum_{i=1}^n d_i$$

where $d_i = \mu_i^2\dot\psi(\mu_i\phi) + (1-\mu_i)^2\dot\psi\big[(1-\mu_i)\phi\big] - \dot\psi(\phi),\ i = 1,\ldots,n$. It follows that

$$\mathbb{E}\Big[-\frac{d}{d\phi^2}\,\ell_n(\boldsymbol{\mu},\phi)\Big] = \sum_{i=1}^n d_i$$

(this is different from the prompt, but seems right)

Next, we observe that (using some previous calculations)

$$\frac{d}{d\phi\, d\boldsymbol{\beta}}\,\ell_n(\boldsymbol{\mu},\phi) = \frac{d}{d\phi}\Big[\phi\sum_{i=1}^n (y_i^* - \mu_i^*)\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i\Big]$$

$$= \sum_{i=1}^n (y_i^* - \mu_i^*)\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i - \phi\sum_{i=1}^n \frac{d\mu_i^*}{d\phi}\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i$$

$$= \sum_{i=1}^n (y_i^* - \mu_i^*)\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i - \phi\sum_{i=1}^n \Big(\mu_i\,\psi(\mu_i\phi) + (1-\mu_i)\,\psi\big[(1-\mu_i)\phi\big]\Big)\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i$$

$$= \sum_{i=1}^n (y_i^* - \mu_i^*)\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i - \sum_{i=1}^n c_i\,\frac{1}{\dot g(\mu_i)}\,\boldsymbol{x}_i$$

Let us denote $\boldsymbol{\theta}_i = (\theta_{i1},\theta_{i2}),\ i = 1,\ldots,n$. We saw in (2b) that

$$p(y_i;\,\mu_i,\phi) = \exp\Big\{\theta_{i1}y_i^* + \theta_{i2}\log(1-y_i) - b(\boldsymbol{\theta}_i) + c(y_i)\Big\}$$

where

$$b(\boldsymbol{\theta}_i) = -\log\Big[\frac{\Gamma(\theta_{i2})}{\Gamma(\theta_{i1})\,\Gamma(\theta_{i2}-\theta_{i1})}\Big] = \log\big[\Gamma(\theta_{i1})\big] + \log\big[\Gamma(\theta_{i2}-\theta_{i1})\big] - \log\big[\Gamma(\theta_{i2})\big]$$

and $\boldsymbol{\theta}_i = (\mu_i\phi,\phi)$. We further observe that

$$\frac{d}{d\theta_{i1}}\,\ell_i(\mu_i,\phi) = \psi(\theta_{i1}) + \psi(\theta_{i2}-\theta_{i1}) = \psi(\mu_i\phi) - \psi\big[(1-\mu_i)\phi\big] = \mu_i^*, \qquad i = 1,\ldots,n$$

Since from exponential family theory

$$\mathbb{E}\,Y_i^* = \frac{d}{d\theta_{i1}}\,b(\boldsymbol{\theta}_i) = \mu_i^*$$

we see that

$$\mathbb{E}\left[-\frac{d}{d\phi\, d\boldsymbol{\beta}}\,\ell_n(\boldsymbol{\mu},\phi)\right] = -\sum_{i=1}^{n}(\mathbb{E}\,Y_i^* - \mu_i^*)\,\frac{1}{\dot{g}(\mu_i)}\,\boldsymbol{x}_i + \sum_{i=1}^{n} c_i\,\frac{1}{\dot{g}(\mu_i)}\,\boldsymbol{x}_i$$

$$= \sum_{i=1}^{n} c_i\,\frac{1}{\dot{g}(\mu_i)}\,\boldsymbol{x}_i$$

$$= \boldsymbol{X}'\boldsymbol{T}\boldsymbol{c} \qquad\qquad \text{where} \quad \boldsymbol{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

In conclusion we have determined that

$$\boldsymbol{I}(\boldsymbol{\mu},\phi) = \begin{bmatrix} \phi\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{T}\boldsymbol{c} \\ \boldsymbol{c}'\boldsymbol{T}\boldsymbol{X} & \sum_{i=1}^{n} d_i \end{bmatrix}$$

where $\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{T}, \boldsymbol{c}$ and $d_1,\ldots,d_n$ are as defined above.

(d) Let $\widehat{\xi} = (\widehat{\boldsymbol{\beta}}, \widehat{\phi})$ denote the maximimum likelihood estimator of ξ. Derive from first principles the asymptotic distribution of $\widehat{\xi}$, properly normalized.

Let

$$\boldsymbol{s}_n(\xi) = \begin{pmatrix} \dfrac{d}{d\,\boldsymbol{\beta}}\,\ell_n(\boldsymbol{\beta},\phi) \\ \dfrac{d}{d\,\phi}\,\ell_n(\boldsymbol{\beta},\phi) \end{pmatrix}$$

Then by the Taylor expansion, we have

$$\boldsymbol{0} = \boldsymbol{s}_n(\widehat{\xi}) = \boldsymbol{s}_n(\xi) + \left[\frac{d}{d\,\xi'}\,\boldsymbol{s}_n(\xi)\right](\widehat{\xi}-\xi) \;+\; o_p(1)$$

$$\implies \quad \left[\frac{d}{d\,\xi'}\,\boldsymbol{s}_n(\xi)\right]^{1/2}(\widehat{\xi}-\xi) = \left[-\frac{d}{d\,\xi'}\,\boldsymbol{s}_n(\xi)\right]^{-1/2}\boldsymbol{s}_n(\xi) \;+\; o_p(1)$$

If we are willing to assume that

$$\left\|\left(-\frac{d}{d\,\xi'}\,\boldsymbol{s}_n(\xi)\right) - I_n(\xi)\right\| \xrightarrow{a.s.} 0$$

then by Lindeberg's CLT and under some additional assumptions we obtain that

$$\left[-\frac{d}{d\,\xi'}\,\boldsymbol{s}_n(\xi)\right]^{-1/2}\boldsymbol{s}_n(\xi) \xrightarrow{L} N(\boldsymbol{0},\boldsymbol{I})$$

By Slutsky's theorem we can further conclude that

$$[I(\boldsymbol{x})]^{1/2}(\widehat{\xi}-\xi) \xrightarrow{L} N(\boldsymbol{0},\boldsymbol{I})$$

Discussion: the assumption that we have placed on the asymptotic equivalence of the negative gradient of $\boldsymbol{s}_n$ and the Information matrix is far stronger than is needed to establish this result; however it seems like the only reasonable way to answer this question in a testing situation. Note that we cannot use the strong law of large numbers to obtain this result because we do not have identically distributed random variables. To see a statement / proof of this result under weaker assumptions see Shao pages 292-295, or to see the original source see Fahrmeir and Kaufmann (1985).

(e) The multivariate generalization of the distribution in (2) is called the Dirichlet distribution, which may be defined as follows. Let $r_1, \dots, r_k$ be independent random variables with $r_j \sim gamma(\alpha_j, 1),\ j = 1, \dots, k$. The gamma$(a, b)$ density is given by $f(r) = \frac{b^a}{\Gamma(a)} r^{a-1} e^{-br}$ for $r > 0,\ a > 0,\ b > 0$. Define $s = \sum_{j=1}^{k} r_j$ and $q_j = r_j/s,\ j = 1, \dots, k$. The joint densisty of $(q_1, \dots, q_{k-1})$ is called the Dirichlet density. Derive the joint density of $(q_1, \dots, q_{k-1})$.

Let $S = \sum_{i=1}^{n} R_i,\ Q_i = R_i/S,\ i = 1, \dots, k-1$. Then

$$0 < Q_1, \dots, Q_{k-1} < 1, \quad \sum_{i=1}^{k-1} Q_i < 1, \quad S > 0$$

and

$$R_i = SQ_i, \quad i = 1, \dots, k-1$$
$$R_k = S - \sum_{i=1}^{k-1} R_i = S - \sum_{i=1}^{k-1} SQ_i = S\left(1 - \sum_{i=1}^{k-1} Q_i\right)$$

Also since the determinant operator is invariant to elementary row operations we see that by adding each row to the last row we obtain

$$J(q_1, \dots, q_{k-1}, s) = \begin{vmatrix} s & 0 & \cdots & 0 & q_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & s & q_{k-1} \\ -s & \cdots & \cdots & -s & 1 - \sum_{i=1}^{k-1} q_i \end{vmatrix} = \begin{vmatrix} s & 0 & \cdots & 0 & q_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \vdots & & \ddots & s & q_{k-1} \\ 0 & \cdots & \cdots & 0 & 1 \end{vmatrix} = s^{k-1}$$

Then

$$\begin{aligned} &f_{Q_1,\dots,Q_{k-1},S}(q_1, \dots, q_{k-1}, s) \\ &= f_{R_1,\dots,R_k}\left(sq_1, \dots, sq_{k-1}, s\left(1 - \textstyle\sum_{i=1}^{k-1} q_i\right)\right) \left| J(q_1, \dots, q_{k-1}, s) \right| \\ &= \left[\prod_{i=1}^{k-1} f_{R_i}(sq_i)\right] f_{R_k}\left(s\left(1 - \textstyle\sum_{j=1}^{k-1} q_j\right)\right) s^{k-1} \\ &= \left[\prod_{i=1}^{k-1} \Gamma(\alpha_i)(sq_i)^{\alpha_i - 1} e^{-sq_i}\right] \Gamma(\alpha_k)\left[s\left(1 - \textstyle\sum_{j=1}^{k-1} q_j\right)\right]^{\alpha_k - 1} \exp\left\{-s\left(1 - \textstyle\sum_{j=1}^{k-1} q_j\right)\right\} s^{k-1} \\ &= \left[\prod_{i=1}^{k-1} \Gamma(\alpha_i) q_i^{\alpha_i - 1}\right] \Gamma(\alpha_k)\left(1 - \textstyle\sum_{i=j}^{k-1} q_j\right)^{\alpha_k - 1} s^{\sum_{i=j}^{k} q_j} e^{-s}, \qquad \begin{matrix} 0 < q_1, \dots, q_{k-1} < 1 \\ s > 0 \end{matrix}, \quad \sum_{i=1}^{k-1} q_i < 1 \end{aligned}$$

It follows that

$$
\begin{aligned}
f_{Q_1,\ldots,Q_{k-1}}(q_1,\ldots,q_{k-1}) &= \int_0^\infty f_{Q_1,\ldots,Q_{k-1},S}(q_1,\ldots,q_{k-1},s)\,ds \\
&= \left[\prod_{i=1}^{k-1}\Gamma(\alpha_i)\,q_i^{\alpha_i-1}\right]\Gamma(\alpha_k)\left(1-\sum_{i=j}^{k-1}q_j\right)^{\alpha_k-1}\int_0^\infty s^{\sum_{i=j}^{k}q_j}e^{-s}\,ds \\
&= \left[\prod_{i=1}^{k-1}\Gamma(\alpha_i)\,q_i^{\alpha_i-1}\right]\Gamma(\alpha_k)\left(1-\sum_{i=j}^{k-1}q_j\right)^{\alpha_k-1}\frac{1}{\Gamma\left(\sum_{i=1}^{k}\right)} \\
&= \frac{\prod_{i=1}^{k}\Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{k}\right)}\left[\prod_{i=1}^{k-1}q_i^{\alpha_i-1}\right]\left(1-\sum_{i=j}^{k-1}q_j\right)^{\alpha_k-1}, \qquad 0<q_1,\ldots,q_{k-1}<1,\quad \sum_{i=1}^{k-1}q_i<1
\end{aligned}
$$

or equivalently

$$
f_{Q_1,\ldots,Q_k}(q_1,\ldots,q_k) = \frac{\prod_{i=1}^{k}\Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{k}\right)}\prod_{i=1}^{k}q_i^{\alpha_i-1}, \qquad 0<q_1,\ldots,q_k<1,\quad \sum_{i=1}^{k}q_i=1
$$

3 2011 Theory II Problem #3

3. Consider the linear model

$$
\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}
$$

where $\boldsymbol{X}$ is a $n\times p$ covariate matrix, $\boldsymbol{\beta}$ is a $p\times 1$ vector of regression coefficients, and $\boldsymbol{\epsilon} = (\epsilon_1,\ldots,\epsilon_n)'$, where the ϵ_i's are i.i.d. $N(0,\sigma^2)$, $i=1,\ldots,n$. In this problem, both $\boldsymbol{\beta}$ and σ^2 are unknown.

(a) Suppose that $\text{rank}(\boldsymbol{X}) = r \le p$ and we wish to test

$$
H_0\colon \boldsymbol{\ell}'\boldsymbol{\beta} = \theta_0 \quad \text{vs.} \quad H_1\colon \boldsymbol{\ell}'\boldsymbol{\beta} \neq \theta_0 \tag{4}
$$

where $\boldsymbol{\ell} \in C(\boldsymbol{X}')$ and $C(\boldsymbol{X}')$ denotes the column space of $\boldsymbol{X}$, and θ_0 is a specified constant. Derive a UMPU size α test for the hypothesis in (4). Determine the exact distribution of the test statistic under H_0 and H_1 as well as an explicit expression of the critical value to make the test size α.

Proposition 1. *Let $\boldsymbol{A}$ be a symmetric matrix and let $\boldsymbol{G}$ be a generalized inverse of $\boldsymbol{A}$. Then $\boldsymbol{G}'$ is a generalized inverse of $\boldsymbol{A}$.*

Proof. $\boldsymbol{A} = \boldsymbol{AGA} \implies \boldsymbol{A}' = (\boldsymbol{AGA})' \implies \boldsymbol{A} = \boldsymbol{AG'A}$ □

Corollary 1. *One may use the result from Proposition 1 to justify writing $[(\boldsymbol{X}'\boldsymbol{X})^-]' = (\boldsymbol{X}'\boldsymbol{X})^-$*

Proposition 2. *$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'$ is the perpendicular projection matrix onto $C(\boldsymbol{X})$.*

Proof. See Christensen, page 430 □

Proposition 3. *Let $\mathcal{W} \subset \mathbb{R}^n$ with $\dim(\mathcal{W}) = r$ where $0 < r < n$. Then there exists an orthogonal matrix $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n-r})$ such that $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ form a basis for $\mathcal{W}$.*

Proof. It is a fundamental result from linear algebra that we may construct a basis $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_r^*, \boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_{n-r}^*$ for $\mathbb{R}^n$ such that $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_r^*$ forms a basis for $\mathcal{W}$. Applying the Gram-Schmidt process to $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_r^*, \boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_{n-r}^*$ yields an orthogonal matrix, say, $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n-r})$. Furthermore, since each $\boldsymbol{w}_i$, $i = 1, \ldots, r$ is a linear combination of $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_r^*$ it holds that $C(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r) \subset \mathcal{W}$. But since $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ are orthogonal it follows that $\operatorname{rank}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r) = r$ and hence $C(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r) = \mathcal{W}$. □

Proposition 4. *Let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n-r}$ be an orthogonal basis for $\mathbb{R}^n$. Let $\mathcal{W}$ be the subspace generated by $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ and let $\mathcal{U}$ be the subspace generated by $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n-r}$. Then $\mathcal{U} = \mathcal{W}^\perp$.*

Proof. This fundamental linear algebra result can be found in many introductory linear algebra texts; for one such presentation see Lang's *Introduction to Linear Algebra*, pages 187-188. □

Proposition 5. *Let $\boldsymbol{M}$ be a perpendicular projection matrix onto $C(\boldsymbol{X})$. Then there exists an orthogonal matrix $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2)$ such that $\boldsymbol{M}\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1\ \boldsymbol{0})$.*

Proof. It follows from Proposition 3 and Proposition 4 that there exists an orthogonal matrix $(\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2)$ such that $C(\boldsymbol{\Gamma}_1) = \boldsymbol{X}$ and $C(\boldsymbol{\Gamma}_2) = [C(\boldsymbol{\Gamma}_1)]^\perp$. The stated result then follows directly from the definition of a perpendicular projection matrix. □

Our model is $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{X} \sim n \times p$ with $\operatorname{rank}(\boldsymbol{X}) = r \leq p$. Then $\boldsymbol{M} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^- \boldsymbol{X}'$ is a projection matrix of rank r so that there exists an $n \times n$ orthogonal matrix $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2)$ such that $\boldsymbol{M}\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1\ \boldsymbol{0})$. Thus we can express our model as

$$\boldsymbol{Z} \equiv \boldsymbol{\Gamma}'\boldsymbol{Y} = \boldsymbol{\Gamma}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}'\boldsymbol{\epsilon}$$

Since $\boldsymbol{\Gamma}$ is orthogonal it follows $\boldsymbol{\Gamma}$ is full rank. Thus this transformation is bijective and any testing based on the model is equivalent to testing based on the original model. Let $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{bmatrix}$, then we observe that

$$\mathbb{E}\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\mathbb{E}Y \\ \boldsymbol{\Gamma}_2'\mathbb{E}Y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{\Gamma}_2'\boldsymbol{X}\boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{\Gamma}_2'\boldsymbol{M}\boldsymbol{X}\boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{0} \end{bmatrix}$$

$$\operatorname{Var}[\boldsymbol{Z}] = \boldsymbol{\Gamma}'\operatorname{Var}[\boldsymbol{Y}]\boldsymbol{\Gamma} = \sigma^2\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \sigma^2\boldsymbol{I}$$

Then

$$
\begin{aligned}
p\left(\boldsymbol{z};\, \boldsymbol{\eta}, \sigma^2\right) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\begin{bmatrix}\boldsymbol{Z}_1\\ \boldsymbol{Z}_2\end{bmatrix}-\begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{0}\end{bmatrix}\right)'\left(\begin{bmatrix}\boldsymbol{Z}_1\\ \boldsymbol{Z}_2\end{bmatrix}-\begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{0}\end{bmatrix}\right)\right\}\\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[(\boldsymbol{Z}_1-\boldsymbol{\eta})'(\boldsymbol{Z}_1-\boldsymbol{\eta})+\boldsymbol{Z}_2'\boldsymbol{Z}_2\right]\right\}\\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{Z}_1'\boldsymbol{Z}_1-2\boldsymbol{Z}_1'\boldsymbol{\eta}+\boldsymbol{\eta}'\boldsymbol{\eta}+\boldsymbol{Z}_2'\boldsymbol{Z}_2\right]\right\}\\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[||\boldsymbol{Z}_1||^2+||\boldsymbol{Z}_2||^2-2\boldsymbol{\eta}'\boldsymbol{Z}_1+||\boldsymbol{\eta}||^2\right]\right\}\\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{||\boldsymbol{Z}_1||^2+||\boldsymbol{Z}_2||^2}{2\sigma^2}+\frac{\boldsymbol{\eta}'\boldsymbol{Z}_1}{\sigma^2}-\frac{||\boldsymbol{\eta}||^2}{2\sigma^2}\right\}
\end{aligned}
$$

Next, define

$$\boldsymbol{\lambda}' = \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$$

Since $\boldsymbol{\ell} \in C(\boldsymbol{X}')$ it follows that $\boldsymbol{\lambda}'\boldsymbol{X} = \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{\ell}'$. Let $\widehat{\boldsymbol{\beta}}$ the LSE of $\boldsymbol{\beta}$, then

$$
\begin{aligned}
\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}} &= \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\lambda}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\lambda}'\boldsymbol{M}\boldsymbol{y} = \boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{M}\boldsymbol{y} = \boldsymbol{\lambda}'[\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2]\begin{bmatrix}\boldsymbol{\Gamma}_1'\\ \boldsymbol{\Gamma}_2'\end{bmatrix}\boldsymbol{M}\boldsymbol{y}\\
&= \boldsymbol{\lambda}'[\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2]\begin{bmatrix}\boldsymbol{\Gamma}_1'\boldsymbol{M}\\ \boldsymbol{\Gamma}_2\boldsymbol{M}\end{bmatrix}\boldsymbol{y} = \boldsymbol{\lambda}'[\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2]\begin{bmatrix}\boldsymbol{\Gamma}_1'\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{y} = \boldsymbol{\lambda}'\left(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1'+\boldsymbol{\Gamma}_2\cdot\boldsymbol{0}\right)\boldsymbol{y} = \boldsymbol{\lambda}'\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1'\boldsymbol{y}\\
&= \boldsymbol{\lambda}'\boldsymbol{\Gamma}_1\boldsymbol{z}_1 \equiv \boldsymbol{a}'\boldsymbol{z}_1
\end{aligned}
$$

where we have defined $\boldsymbol{a}' = \boldsymbol{\lambda}'\boldsymbol{\Gamma}_1 = \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Gamma}_1$. It follows that

$$\boldsymbol{\ell}'\boldsymbol{\beta} = \mathbb{E}\left[\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}\right] = \mathbb{E}\left[\boldsymbol{a}'\boldsymbol{Z}_1\right] = \boldsymbol{a}'\boldsymbol{\eta}$$

Now we wish to express $p\left(\boldsymbol{z};\, \boldsymbol{\eta}, \sigma^2\right)$ in such a form that we may invoke Theorem 2.7 in the class text (slides 328-331) to test the hypothesis in question. Assume that $\boldsymbol{\ell} \neq \boldsymbol{0}$; otherwise the hypothesis is nonsensical. Then

$$\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} = \boldsymbol{\ell} \neq \boldsymbol{0} \quad\Longrightarrow\quad \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} \neq \boldsymbol{0} \quad\Longrightarrow\quad \boldsymbol{a} = \boldsymbol{\Gamma}_1'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} \neq \boldsymbol{0}$$

Thus without loss of generality, assume that $a_1 \neq 0$. We may do this since if a_1 was in fact 0 then we can just reindex the following algebraic manipulations using some i such that $a_i \neq 0$. Denote $\boldsymbol{z_1} = (z_{11}, \ldots, z_{1r})$. Then

$$
\begin{aligned}
p\left(\boldsymbol{z};\, \boldsymbol{\eta}, \sigma^2\right) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{||\boldsymbol{Z}_1||^2 + ||\boldsymbol{Z}_2||^2}{2\sigma^2} + \frac{\boldsymbol{\eta}'\boldsymbol{Z}_1}{\sigma^2} - \frac{||\boldsymbol{\eta}||^2}{2\sigma^2}\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{\boldsymbol{a}'\boldsymbol{\eta}}{a_1\sigma^2} z_{11} - \frac{\boldsymbol{a}'\boldsymbol{\eta}}{a_1\sigma^2} z_{11} - \frac{||\boldsymbol{Z}_1||^2 + ||\boldsymbol{Z}_2||^2}{2\sigma^2} + \frac{\boldsymbol{\eta}'\boldsymbol{Z}_1}{\sigma^2} - \frac{||\boldsymbol{\eta}||^2}{2\sigma^2}\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{\boldsymbol{a}'\boldsymbol{\eta}}{a_1\sigma^2} z_{11} - \left(\frac{\eta_1}{\sigma^2} z_{11} + \sum_{k=2}^{r} \frac{a_k\eta_k}{a_1\sigma^2} z_{11}\right) - \frac{||\boldsymbol{Z}_1||^2 + ||\boldsymbol{Z}_2||^2}{2\sigma^2} + \sum_{k=1}^{r} \frac{\eta_k}{\sigma^2} z_{1k} - \frac{||\boldsymbol{\eta}||^2}{2\sigma^2}\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{\boldsymbol{a}'\boldsymbol{\eta}}{a_1\sigma^2} z_{11} + \sum_{k=2}^{r} \left(z_{1k} - \frac{a_k z_{11}}{a_1}\right) \frac{\eta_k}{\sigma^2} - \frac{||\boldsymbol{Z}_1||^2 + ||\boldsymbol{Z}_2||^2}{2\sigma^2} - \frac{||\boldsymbol{\eta}||^2}{2\sigma^2}\right\} \\
&= \exp\left\{\theta\, U(\boldsymbol{Z}) + \left[\boldsymbol{T}(\boldsymbol{Z})\right]'\xi + c\,(\theta, \xi)\right\}
\end{aligned}
$$

where

$$
(\theta, \xi) = \left(\frac{\boldsymbol{a}'\boldsymbol{\eta}}{a_1\sigma^2},\ \frac{\eta_2}{\sigma^2}, \ldots, \frac{\eta_r}{\sigma^2},\ \frac{1}{2\sigma^2}\right)
$$

$$
\left(U(\boldsymbol{z}), \boldsymbol{T}(\boldsymbol{z})\right) = \left(z_{11},\ z_{12} - \frac{a_2 z_{11}}{a_1}, \ldots, z_{1r} - \frac{a_r z_{11}}{a_1},\ ||\boldsymbol{z}_1||^2 + ||\boldsymbol{z}_2||^2\right)
$$

Next, consider

$$
V = h(U, \boldsymbol{T}) \equiv \frac{\sqrt{n-r}\,(\boldsymbol{a}'\boldsymbol{Z}_1 - \theta_0)}{||\boldsymbol{Z}_1||\ ||\boldsymbol{a}||} = \frac{(\boldsymbol{\ell}'\hat{\boldsymbol{\beta}} - \theta_0)\,/\sqrt{\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell}}}{\sqrt{||\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2/(n-r)}} \overset{H_0}{\sim} t_{n-r}
$$

where the last equality is a consequence of the following identities:

① $\boldsymbol{a}'\boldsymbol{Z}_1 = \boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}$

②
$$
\begin{aligned}
||\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2 &= ||(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}||^2 = ||\boldsymbol{\Gamma}'(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}||^2 = ||\boldsymbol{\Gamma}'\boldsymbol{Y} - \boldsymbol{\Gamma}'\boldsymbol{M}\boldsymbol{Y}||^2 \\
&= \left|\left| \begin{bmatrix}\boldsymbol{\Gamma}_1' \\ \boldsymbol{\Gamma}_2'\end{bmatrix}\boldsymbol{Y} - \begin{bmatrix}\boldsymbol{\Gamma}_1'\boldsymbol{M} \\ \boldsymbol{\Gamma}_2'\boldsymbol{M}\end{bmatrix}\boldsymbol{Y} \right|\right|^2 = \left|\left| \begin{bmatrix}\boldsymbol{\Gamma}_1' \\ \boldsymbol{\Gamma}_2'\end{bmatrix}\boldsymbol{Y} - \begin{bmatrix}\boldsymbol{\Gamma}_1' \\ \boldsymbol{0}\end{bmatrix}\boldsymbol{Y} \right|\right|^2 \\
&= \left|\left| \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{\Gamma}_2'\boldsymbol{Y}\end{bmatrix} \right|\right|^2 = ||\boldsymbol{0}||^2 + ||\boldsymbol{\Gamma}_2'\boldsymbol{Y}||^2 = ||\boldsymbol{Z}_2||^2
\end{aligned}
$$

③
$$
\begin{aligned}
||\boldsymbol{a}||^2 &= ||\boldsymbol{\Gamma}_1'\boldsymbol{\lambda}||^2 = \left|\left| \begin{bmatrix}\boldsymbol{\Gamma}_1' \\ \boldsymbol{0}\end{bmatrix}\boldsymbol{\lambda} \right|\right|^2 = \left|\left| \begin{bmatrix}\boldsymbol{\Gamma}_1'\boldsymbol{M} \\ \boldsymbol{\Gamma}_2'\boldsymbol{M}\end{bmatrix}\boldsymbol{\lambda} \right|\right|^2 = ||\boldsymbol{\Gamma}'\boldsymbol{M}\boldsymbol{\lambda}||^2 \\
&= \boldsymbol{\lambda}'\boldsymbol{M}\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{M}\boldsymbol{\lambda} = \boldsymbol{\lambda}'\boldsymbol{M}\boldsymbol{\lambda} = \boldsymbol{\lambda}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\lambda} \\
&= \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} = \boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell}
\end{aligned}
$$

Next we want to verify the conditions of Remark 2.6b in the class text (slide 332). It is clear that V can be expressed as $V = a(\boldsymbol{t})U + b(\boldsymbol{t})$ for $a(\boldsymbol{t}) > 0$. Then we wish to establish that V is independent of $\boldsymbol{T}$ on the boundary. As a property of exponential families, $(U, \boldsymbol{T})$ is complete sufficient for the family of distributions $\mathscr{P} = \left\{N(\boldsymbol{X\beta}, \sigma^2) \colon \boldsymbol{\beta} \in \mathbb{R}^p,\ \sigma^2 > 0\right\}$. But since V follows a t distribution on the boundary it is ancillary of $\mathscr{P}$ so that by Basu's theorem $V \perp\!\!\!\perp (U, \boldsymbol{T})$ and hence of $\boldsymbol{T}$. Thus the conditions of Remark 2.6b are satisfied and our UMPU test becomes

$$\phi(v) = \begin{cases} 1, & v < c_1 \text{ or } v > c_2 \\ 0, & \text{else} \end{cases}$$

where c_1, c_2 satisfy

$$\begin{cases} \mathbb{E}_{\theta_0}\left[\phi(V)\right] = \alpha \\ \mathbb{E}_{\theta_0}\left[V\phi(V)\right] = \alpha\,\mathbb{E}_{\theta_0}[V] \end{cases}$$

Notice that V is symmetric on the boundary so that we also reject for

$$-V < c_1 \quad \text{or} \quad -V > c_2 \qquad \Longleftrightarrow \qquad V < -c_2 \quad \text{or} \quad V > -c_1$$

But UMPU tests are unique a.s. so it must be the case that $c_1 = -c_2$. It follows that

$$\begin{aligned} \alpha \overset{set}{=} \mathbb{E}_{\theta_0}\left[\phi(V)\right] &= \mathbb{P}(V < c_1) + \mathbb{P}(V > c_2) = \mathbb{P}(V < c_1) + \mathbb{P}(V > -c_1) = 2\,\mathbb{P}(V < c_1) \\ \Longrightarrow \quad & c_1 = t_{n-r}\left(\tfrac{\alpha}{2}\right) \\ \Longrightarrow \quad & c_2 = -t_{n-r}\left(\tfrac{\alpha}{2}\right) = t_{n-r}\left(1 - \tfrac{\alpha}{2}\right) \end{aligned}$$

(b) Consider the model in (3) and the hypothesis in (4).

(i) Derive an explicit closed-form expression for the asymptotic power function of the UMPU test in part (a).

We follow the approach of DuPont and Plummer (1998). Let $\widehat{\theta} = \boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}$ and define $\sigma_\theta^2 = \mathrm{Var}\left[\sqrt{n}\widehat{\theta}\right]$. Also define

$$S^2 \equiv \frac{||\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2}{n-r} \sim \sigma^2 \chi^2_{n-r}$$

We see that

$$\begin{aligned} \sigma_\theta^2 &\equiv \mathrm{Var}\left[\sqrt{n}\widehat{\theta}\right] = \mathrm{Var}\left[\sqrt{n}\,\boldsymbol{\ell}'\widehat{\boldsymbol{\beta}}\right] = \mathrm{Var}\left[\sqrt{n}\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{Y}\right] \\ &= n\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\mathrm{Var}\left[\boldsymbol{Y}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} = n\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\sigma^2\boldsymbol{I}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} \\ &= n\sigma^2\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell} \end{aligned}$$

and

$$\frac{\sqrt{n}(\widehat{\theta} - \theta_0)}{\sigma_\theta} \sim N\left(\frac{\sqrt{n}(\theta - \theta_0)}{\sigma_\theta},\ 1\right)$$

so that since $\widehat{\theta} \perp\!\!\!\perp S^2$

$$R_n \equiv \frac{\sqrt{n}(\widehat{\theta} - \theta_0)\,/\sqrt{n\boldsymbol{\ell}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\ell}}}{S} = \frac{\sqrt{n}(\widehat{\theta} - \theta_0)\,/\sigma_\theta}{S/\sigma} \overset{H_0}{\sim} t_{n-r}$$

Denote T_ν as the cdf of the t-distribution with ν degrees of freedom and let θ_a be the true value of θ under the alternative hypothesis. Define

$$\delta_n \equiv \frac{\theta_a - \theta_0}{\sigma_\theta}$$

Then

$$\begin{aligned}
&\text{Power}(\theta_a, \theta_0, \sigma^2, n) = \mathbb{P}\Big(|R_n| > T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right)\Big) \\
&\quad = \mathbb{P}\Big(R_n < -T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right)\Big) + \mathbb{P}\Big(R_n > T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right)\Big) \\
&\quad = \mathbb{P}\left(R_n - \frac{\sqrt{n}\delta_n}{S/\sigma} < -T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \frac{\sqrt{n}\delta_n}{S/\sigma}\right) + \mathbb{P}\left(R_n - \frac{\sqrt{n}\delta_n}{S/\sigma} > T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \frac{\sqrt{n}\delta_n}{S/\sigma}\right) \\
&\quad \approx \mathbb{P}\left(R_n - \frac{\sqrt{n}\delta_n}{S/\sigma} < -T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \sqrt{n}\delta_n\right) + \mathbb{P}\left(R_n - \frac{\sqrt{n}\delta_n}{S/\sigma} > T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \sqrt{n}\delta_n\right) \\
&\quad = T_{n-r}\left\{ -T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \sqrt{n}\delta_n \right\} + T_{n-r}\left\{ \sqrt{n}\delta_n - T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) \right\}
\end{aligned}$$

where the justification for the approximation stems from the fact that $S^2/\sigma^2 \xrightarrow{a.s.} 1$. Denote Power$(\theta_a,$ $\theta_0,\ \sigma^2, n) = \Gamma$. Usually one of the terms in the above result is negligible so approximating it as zero we have for a chosen level of power either (when $\theta_a - \theta_0 > 0$)

$$\begin{aligned}
&\Gamma \approx 0 + T_{n-r}\left\{ \sqrt{n}\delta_n - T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) \right\} \quad \Longrightarrow \quad T_{n-r}^{-1}(\Gamma) = \sqrt{n}\delta_n - T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) \\
&\quad \Longrightarrow \quad n = \frac{\left(T_{n-r}^{-1}(\Gamma) + T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right)\right)^2}{\delta_n^2}
\end{aligned}$$

or alternatively (when $\theta_a - \theta_0 \le 0$)

$$\begin{aligned}
&\Gamma \approx T_{n-r}\left\{ -T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \sqrt{n}\delta_n \right\} + 0 \quad \Longrightarrow \quad T_{n-r}^{-1}(\Gamma) = -T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right) - \sqrt{n}\delta_n \\
&\quad \Longrightarrow \quad n = \frac{\left(T_{n-r}^{-1}(\Gamma) + T_{n-r}^{-1}\left(1-\tfrac{\alpha}{2}\right)\right)^2}{\delta_n^2}
\end{aligned}$$

Thus the result is the same regardless of which term is dropped. We note that numerical techniques must be applied to actually solve the above equation since both the magnitude of δ_n as well as the degrees of freedom are dependent on n.

(ii) Suppose that $p = 2$, $\boldsymbol{X}$ is $n \times 2$ where the first column consists of a vector of ones, $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, $\boldsymbol{\ell} = (0, 1)$, rank$(\boldsymbol{X}) = 2$, and $\sum_{i=1}^n x_i = n/2$, where $(x_1, \ldots, x_n)'$ denotes the second column of $\boldsymbol{X}$. Use the asymptotic power function of part (i) to derive an explicit closed form sample size formula for an α level test with prespecified power.

Let

$$\sigma_{\beta_1}^2 \equiv \text{Var}\left[\sqrt{n}\widehat{\beta}_1\right] = n\,\text{Var}\left[\widehat{\beta}_1\right] = \frac{n\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sigma^2}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} = \frac{\sigma^2}{\frac{1}{n}\sum_{i=1}^n x_i^2 - (n^2/4)}$$

Then plugging

$$\delta_n = \frac{(\beta_1^{(H_1)} - 0)}{\sigma_{\beta_1}}$$

into the result from (i) yields a formula by which a sample size may be calculated.

(c) Consider the model in (3) and suppose that $\text{rank}(\boldsymbol{X}) = p$. We wish to test

$$H_0\colon \boldsymbol{R\beta} = \boldsymbol{b}_0 \quad \text{vs.} \quad H_1\colon \boldsymbol{R\beta} \neq \boldsymbol{b}_0 \tag{5}$$

where $\boldsymbol{R}$ is an $s \times p$ specified matrix of constants of rank $s \leq p$ and $\boldsymbol{b}_0$ is a specified $s \times 1$ vector. Derive the size α likelihood ratio test for this hypothesis and determine the exact distribution of the likelihood ratio statistic (or a monotonic function of it) under H_0 and H_1. In carrying out this derivation, you need to derive all relevant estimates under H_0 and H_1.

Proposition 6. *Let $\boldsymbol{A}$ be a $p \times p$ positive-definite matrix and let $\boldsymbol{C}$ be a $s \times p$ matrix with rank s. Then $\boldsymbol{CAC}'$ is positive-definite.*

Proof. $\boldsymbol{x}'\boldsymbol{CAC}'\boldsymbol{x} \geq 0$ with equality if and only if $\boldsymbol{x}'\boldsymbol{C} = 0$. Since the columns of $\boldsymbol{C}$ are linearly independent this occurs if and only if $\boldsymbol{x} = 0$. Thus $\boldsymbol{x}'\boldsymbol{CAC}'\boldsymbol{x} > 0$ for $x \neq \boldsymbol{0}$. □

Proposition 7. *Let $\boldsymbol{Y}$ be a random vector that follows a $N(\mu, \boldsymbol{I})$ distribution. If $\boldsymbol{M}$ is any perpendicular matrix, then $\boldsymbol{Y}'\boldsymbol{MY} \sim \chi^2\Big(\text{rank}(\boldsymbol{M}),\ \boldsymbol{\mu}'\boldsymbol{M}\boldsymbol{\mu}/2\Big)$*

Proof. See Christensen, page 9 □

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})\right\}$$

$$\begin{aligned}\ell_n(\boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\Big[\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}\Big]\end{aligned}$$

We begin by finding the unrestricted MLE of $(\boldsymbol{\beta}, \sigma^2)$. Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ denote the unrestriced MLEs of $\boldsymbol{\beta}$ and σ^2, respectively. Then since we are assuming that $\boldsymbol{X}$ is full rank we have

$$\frac{d}{d\boldsymbol{\beta}}\ell_n(\boldsymbol{\beta}, \sigma^2) = \frac{1}{2\sigma^2}\Big[-2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X\beta}\Big] \overset{set}{=} \boldsymbol{0} \implies \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y} \implies \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\frac{d}{d\sigma^2}\ell_n(\widehat{\boldsymbol{\beta}}, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \overset{set}{=} 0 \implies \widehat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

Next we wish to maximize $\ell_n(\boldsymbol{\beta}, \sigma^2)$ subject to H_0. Notice that maximizing $\ell_n(\boldsymbol{\beta}, \sigma^2)$ with respect to $\boldsymbol{\beta}$ subject to the constraint is equivalent to minimizing $(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$ subject to the constraint. We have

$$\begin{aligned}
&\frac{d}{d\boldsymbol{\beta}}\Big[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})+\boldsymbol{\lambda}'(\boldsymbol{R}\boldsymbol{\beta}-\boldsymbol{b}_0)\Big]\\
&\quad=\frac{d}{d\boldsymbol{\beta}}\Big[\boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\lambda}'(\boldsymbol{R}\boldsymbol{\beta}-\boldsymbol{b}_0)\Big]\\
&\quad=-2\boldsymbol{X}'\boldsymbol{y}+2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{R}'\boldsymbol{\lambda}\overset{set}{=}\boldsymbol{0}
\end{aligned}$$

Since we are assuming that $\boldsymbol{X}$ is full rank we obtain

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}_{H_0}=\boldsymbol{X}'\boldsymbol{y}-\tfrac{1}{2}\boldsymbol{R}'\boldsymbol{\lambda}\quad\Longrightarrow\quad\boldsymbol{\beta}_{H_0}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}-\tfrac{1}{2}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}=\widetilde{\boldsymbol{\beta}}-\tfrac{1}{2}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}$$

Since $\boldsymbol{R}$ is full rank it follows that $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'$ is full rank and hence is invertible. Thus,

$$\begin{aligned}
&\frac{d}{d\boldsymbol{\lambda}}\Big[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{H_0})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}_{H_0})+\boldsymbol{\lambda}'(\boldsymbol{R}\boldsymbol{\beta}_{H_0}-\boldsymbol{b}_0)\Big]=\boldsymbol{R}\boldsymbol{\beta}_{H_0}-\boldsymbol{b}_0\overset{set}{=}\boldsymbol{0}\\
&\quad\Longrightarrow\quad \boldsymbol{b}_0=\boldsymbol{R}\boldsymbol{\beta}_{H_0}=\boldsymbol{R}\widetilde{\boldsymbol{\beta}}-\tfrac{1}{2}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}_{H_0}\\
&\quad\Longrightarrow\quad -\tfrac{1}{2}\boldsymbol{\lambda}_{H_0}=\big[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\big]^{-1}(\boldsymbol{b}_0-\boldsymbol{R}\widetilde{\boldsymbol{\beta}})
\end{aligned}$$

Plugging this result into the previous equality we obtain

$$\boldsymbol{\beta}_{H_0}=\widetilde{\boldsymbol{\beta}}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}_{H_0}=\widetilde{\boldsymbol{\beta}}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\big[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\big]^{-1}(\boldsymbol{b}_0-\boldsymbol{R}\widetilde{\boldsymbol{\beta}})$$

It is immediately apparent that

$$\widetilde{\sigma}^2=\frac{1}{n}(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})$$

Let $\Theta_0=\Big\{\boldsymbol{\beta},\sigma^2\colon \boldsymbol{R}\boldsymbol{\beta}=\boldsymbol{b}_0,\ \sigma^2>0\Big\}$ and let $\Theta_1=\Big\{\boldsymbol{\beta},\sigma^2\colon \boldsymbol{R}\boldsymbol{\beta}\neq\boldsymbol{b}_0,\ \sigma^2>0\Big\}$. Then we reject the null hypothesis when

$$\lambda<\frac{\sup\limits_{\boldsymbol{\beta},\sigma^2\in\theta_{H_0}}L(\boldsymbol{\beta},\sigma^2)}{\sup\limits_{\boldsymbol{\beta},\sigma^2\in\theta_{H_0}\cup\theta_{H_1}}L(\boldsymbol{\beta},\sigma^2)}=\frac{(2\pi\widetilde{\sigma}^2)^{-n/2}\exp\left\{-\frac{1}{2\widetilde{\sigma}^2}(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})\right\}}{(2\pi\widehat{\sigma}^2)^{-n/2}\exp\left\{-\frac{1}{2\widehat{\sigma}^2}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right\}}=\left(\frac{\widehat{\sigma}^2}{\widetilde{\sigma}^2}\right)^{n/2}$$

or equivalently we reject when

$$\lambda_1>\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}$$

or when

$$
\begin{aligned}
\lambda_2 > \frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} - 1 &= \frac{\widetilde{\sigma}^2 - \widehat{\sigma}^2}{\widehat{\sigma}^2} = \frac{\frac{1}{n}||\boldsymbol{y} - \widetilde{\boldsymbol{y}}||^2 - \frac{1}{n}||\boldsymbol{y} - \widehat{\boldsymbol{y}}||^2}{\frac{1}{n}||\boldsymbol{y} - \widehat{\boldsymbol{y}}||^2} = \frac{||\widehat{\boldsymbol{y}} - \widetilde{\boldsymbol{y}}||^2}{||\boldsymbol{y} - \widehat{\boldsymbol{y}}||^2} = \frac{||\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}||^2}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2} \\
&= \frac{\left|\left|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\left[\widehat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{b}_0 - \boldsymbol{R}\widehat{\boldsymbol{\beta}})\right]\right|\right|^2}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2} \\
&= \frac{\left|\left|\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)\right]\right|\right|^2}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2} \\
&= \frac{(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2} \\
&= \frac{(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2}
\end{aligned}
$$

or when

$$
\lambda_3 > \frac{(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2/(n-p)}
$$

where we have used

$$
\begin{aligned}
||\boldsymbol{y} - \widetilde{\boldsymbol{y}}||^2 &= ||\boldsymbol{y} - \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{y}} - \widetilde{\boldsymbol{y}}||^2 \\
&= ||\boldsymbol{y} - \widehat{\boldsymbol{y}}||^2 + ||\widehat{\boldsymbol{y}} - \widetilde{\boldsymbol{y}}||^2 + 2(\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \widetilde{\boldsymbol{y}}) \\
&= ||\boldsymbol{y} - \widehat{\boldsymbol{y}}||^2 + ||\widehat{\boldsymbol{y}} - \widetilde{\boldsymbol{y}}||^2
\end{aligned}
$$

since

$$
\begin{aligned}
(\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \widetilde{\boldsymbol{y}}) &= (\boldsymbol{y} - \boldsymbol{M}\boldsymbol{y})'(\boldsymbol{M}\boldsymbol{y} - \boldsymbol{M}_0\boldsymbol{y}) = [(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{y}]'[(\boldsymbol{M} - \boldsymbol{M}_0)\boldsymbol{y}] \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{M})(\boldsymbol{M} - \boldsymbol{M}_0)\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{M} - \boldsymbol{M}_0 - \boldsymbol{M}\boldsymbol{M} + \boldsymbol{M}\boldsymbol{M}_0)\boldsymbol{y} \\
&= \boldsymbol{y}'(\boldsymbol{M} - \boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{M}_0)\boldsymbol{y} = 0
\end{aligned}
$$

Next we wish to derive the distribution of the test statistic. We see that

$$
\mathbb{E}[\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0] = \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{b}_0
$$

$$
\begin{aligned}
\mathrm{Var}[\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0] &= \boldsymbol{R}\,\mathrm{Var}[\widehat{\boldsymbol{\beta}}]\boldsymbol{R}' = \boldsymbol{R}\,\mathrm{Var}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}]\boldsymbol{R}' = \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{Var}[\boldsymbol{Y}]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' \\
&= \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' = \sigma^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'
\end{aligned}
$$

By Proposition 6 $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'$ is positive-definite. Thus $[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1/2}$ exists and

$$
[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1/2}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0) \sim N\Big(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1/2}(\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{b}_0),\ \boldsymbol{I}\Big)
$$

Then by Proposition 7

$$(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)' \left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1} (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0) \sim \chi^2\left(s,\ \tfrac{1}{2}\tau\right)$$

where

$$\tau = (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)' \left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1} (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)$$

Next

$$||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}||^2 = ||\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{Y}||^2 = ||(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}||^2 = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{I} - \boldsymbol{M})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

Now $(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. Since $(\boldsymbol{I} - \boldsymbol{M})$ is a perpendicular projection matrix with rank $n - r$, by Proposition 7 we see that

$$||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}||^2 \sim \sigma^2 \chi^2_{n-r}$$

Furthermore,

$$\begin{aligned}\operatorname{Cov}\left[\widehat{\boldsymbol{\beta}}, (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}\right] &= \operatorname{Cov}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y},\ (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}\right] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \operatorname{Var}[\boldsymbol{Y}](\boldsymbol{I} - \boldsymbol{M}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}(\boldsymbol{I} - \boldsymbol{M}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{0}\end{aligned}$$

Thus $\widehat{\boldsymbol{\beta}} \perp\!\!\!\perp (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}$ so that

$$(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0) \ \perp\!\!\!\perp\ ||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}||^2/(n - r)$$

and hence

$$\frac{(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2/(n - p)} \overset{H_1}{\sim} F(s, n - r, \tau/2)$$

Under H_0 $\tau = 0$ so we reject when the test statistic is greater than $F_{s,n-r}(1 - \alpha)$.

(d) Derive the score test for the hypothesis and setup of part (c), and state its asymptotic distribution under H_0.

We wish to obtain $\boldsymbol{I}_n(\boldsymbol{\beta}, \sigma^2)$. Using previous calculations from (3c) we obtain

$$\frac{d}{d\boldsymbol{\beta}\, d\boldsymbol{\beta}'}\ell_n(\boldsymbol{\beta}, \sigma^2) = \frac{d}{d\boldsymbol{\beta}'}\frac{1}{\sigma^2}\left[\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right] = -\boldsymbol{X}'\boldsymbol{X}$$

$$\begin{aligned}\frac{d}{d(\sigma^2)^2}\ell_n(\boldsymbol{\beta}, \sigma^2) &= \frac{d}{d\sigma^2}\left\{-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\} \\ &= \frac{n}{2\sigma^4} - \frac{1}{2\sigma^6}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\end{aligned}$$

$$\frac{d}{d\boldsymbol{\beta}\, d\sigma^2}\ell_n(\boldsymbol{\beta}, \sigma^2) = \frac{d}{d\sigma^2}\frac{1}{\sigma^2}\left[\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right] = \frac{1}{\sigma^4}\left[\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right]$$

Let $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I})$. Then

$$\mathbb{E}\left[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right] = \mathbb{E}[\sigma^2\boldsymbol{Z}'\boldsymbol{Z}] = \sigma^2\,\mathbb{E}[\chi^2_n] = n\sigma^2$$

so that

$$\boldsymbol{I}_n(\boldsymbol{\beta},\sigma^2) = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \mathbf{0} \\ \mathbf{0} & \dfrac{n}{2\sigma^4} \end{bmatrix}$$

Let $s_n(\boldsymbol{\beta},\sigma^2) = \dot{\boldsymbol{\ell}}_n(\boldsymbol{\beta},\sigma^2)$. Then Rao's score test is given by

$$R_n = \left[s_n(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2)\right]'\left[\boldsymbol{I}_n(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2)\right]^{-1} s_n(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2)$$

where $\dot{\boldsymbol{\ell}}_n(\boldsymbol{\beta},\sigma^2)$ and $(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2)$ are as calculated in (3c).

(e) Consider the model in 4 and suppose that $\text{rank}(\boldsymbol{X}) = r \le p$. Derive an exact 95% confidence region for $\boldsymbol{R\beta}$, where $\boldsymbol{R}$ is an $s \times p$ matrix of constants of rank $s \le r$, and all rows of $\boldsymbol{R}$ are contained in $C(\boldsymbol{X})$.

A 95% confidence region is given by

$$\left\{\boldsymbol{R\beta}\colon\ \frac{(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{b}_0)}{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2/(n-p)} < F_{s,n-r}(1-\alpha)\right\}$$

(f) Consider the model in (4) and suppose that $\text{rank}(\boldsymbol{X}) = r \le p$. Derive a UMPU size α test for testing $H_0\colon \sigma^2 \le \sigma_0^2$ vs. $H_1\colon \sigma^2 > \sigma_0^2$, where σ_0^2 is a specified constant. Determine the exact distribution of the test statistic under H_0 and determine an explicit expression of the critical value to make the test size α.

Our model is $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{X} \sim n \times p$ with $\text{rank}(\boldsymbol{X}) = r \le p$. Then $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is a projection matrix of rank r so that there exists an $n \times n$ orthogonal matrix $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2)$ such that $\boldsymbol{M\Gamma} = (\boldsymbol{\Gamma}_1\ \mathbf{0})$. Thus we can express our model as

$$\boldsymbol{Z} \equiv \boldsymbol{\Gamma}'\boldsymbol{Y} = \boldsymbol{\Gamma}'\boldsymbol{X\beta} + \boldsymbol{\Gamma}'\boldsymbol{\epsilon}$$

Since $\boldsymbol{\Gamma}$ is orthogonal it follows $\boldsymbol{\Gamma}$ is full rank. Thus this transformation is bijective and any testing based on the model is equivalent to testing based on the original model. Let $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{bmatrix}$, then we observe that

$$\mathbb{E}\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\mathbb{E}Y \\ \boldsymbol{\Gamma}_2'\mathbb{E}Y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\boldsymbol{X\beta} \\ \boldsymbol{\Gamma}_2'\boldsymbol{X\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\boldsymbol{X\beta} \\ \boldsymbol{\Gamma}_2'\boldsymbol{MX\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_1'\boldsymbol{X\beta} \\ \mathbf{0} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{0} \end{bmatrix}$$

$$\text{Var}[\boldsymbol{Z}] = \boldsymbol{\Gamma}'\text{Var}[\boldsymbol{Y}]\boldsymbol{\Gamma} = \sigma^2\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \sigma^2\boldsymbol{I}$$

Then

$$
\begin{aligned}
p\left(\boldsymbol{z};\,\boldsymbol{\eta},\sigma^2\right) &= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\begin{bmatrix}\boldsymbol{Z}_1\\\boldsymbol{Z}_2\end{bmatrix}-\begin{bmatrix}\boldsymbol{\eta}\\\boldsymbol{0}\end{bmatrix}\right)'\left(\begin{bmatrix}\boldsymbol{Z}_1\\\boldsymbol{Z}_2\end{bmatrix}-\begin{bmatrix}\boldsymbol{\eta}\\\boldsymbol{0}\end{bmatrix}\right)\right\}\\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[(\boldsymbol{Z}_1-\boldsymbol{\eta})'(\boldsymbol{Z}_1-\boldsymbol{\eta})+\boldsymbol{Z}_2'\boldsymbol{Z}_2\right]\right\}\\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{Z}_1'\boldsymbol{Z}_1-2\boldsymbol{Z}_1'\boldsymbol{\eta}+\boldsymbol{\eta}'\boldsymbol{\eta}+\boldsymbol{Z}_2'\boldsymbol{Z}_2\right]\right\}\\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[||\boldsymbol{Z}_1||^2+||\boldsymbol{Z}_2||^2-2\boldsymbol{\eta}'\boldsymbol{Z}_1+||\boldsymbol{\eta}||^2\right]\right\}\\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{||\boldsymbol{Z}_1||^2+||\boldsymbol{Z}_2||^2}{2\sigma^2}+\frac{\boldsymbol{\eta}'\boldsymbol{Z}_1}{\sigma^2}-\frac{||\boldsymbol{\eta}||^2}{2\sigma^2}\right\}
\end{aligned}
$$

Thus by Theorem 2.7 in the class text (slides 328-331), a UMPU test is based on $U(\boldsymbol{Z}) \equiv ||\boldsymbol{Z}_1||^2+||\boldsymbol{Z}_2||^2$ and $\boldsymbol{T}(\boldsymbol{Z}) \equiv \boldsymbol{Z}_1$ for $\theta \equiv -1/(2\sigma^2)$. Let $V = ||\boldsymbol{Z}_2||^2$. Then $V = U - ||\boldsymbol{T}||^2$ which satisfies the first condition of Remark 2.6a (slide 331). Since $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are uncorrelated and normally distributed it follows that $\boldsymbol{Z}_1 \perp\!\!\!\perp \boldsymbol{Z}_2$. Thus $V \perp\!\!\!\perp \boldsymbol{T}$ and the second condition of Remark 2.6a is satisfied.

We wish to test H_0: $\sigma^2 \le \sigma_0^2$ vs. H_1: $\sigma^2 > \sigma_0^2$. Then a UMPU test is given by

$$
\phi(v) = \begin{cases} 1, & V > c\\ 0, & \text{else}\end{cases}
$$

where c satisfies $\mathbb{P}_{\sigma_0^2}(V > c) = \alpha$. Since on the boundary $V \sim \sigma_0^2\chi^2_{n-r}$ we have

$$
\begin{aligned}
& \alpha = \mathbb{P}_{\sigma_0^2}\left(V > c\right) = \mathbb{P}\left(\sigma_0^2\chi^2_{n-r} > c\right)\\
\iff\quad & 1-a = \mathbb{P}\left(\sigma_0^2\chi^2_{n-r} \le c\right) = \mathbb{P}\left(\chi^2_{n-r} \le c/\sigma_0^2\right)\\
\iff\quad & \chi^2_{n-r}(1-\alpha) = c/\sigma_0^2\\
\iff\quad & c = \sigma_0^2\chi^2_{n-r}(1-\alpha)
\end{aligned}
$$

Finally, we would like to be able to express V in a form that does not require us to actually calculate a $\boldsymbol{\Gamma}$ with the desired properties. We observe that

$$
||\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta}||^2 = \left|\left|\boldsymbol{\Gamma}'(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})\right|\right|^2 = \left|\left|\boldsymbol{Z}-\begin{bmatrix}\boldsymbol{\eta}\\\boldsymbol{0}\end{bmatrix}\right|\right|^2 = ||\boldsymbol{Z}_1-\boldsymbol{\eta}||^2+||\boldsymbol{Z}_2||^2
$$

so that

$$
\min_{\boldsymbol{\beta}}||\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}||^2 = \min_{\boldsymbol{\eta}}||\boldsymbol{Z}_1-\boldsymbol{\eta}||^2+||\boldsymbol{Z}_2||^2
$$

Therefore,

$$
V = ||\boldsymbol{Z}_2||^2 = \min_{\boldsymbol{\beta}}||\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}||^2 - \min_{\boldsymbol{\eta}}||\boldsymbol{Z}_1-\boldsymbol{\eta}||^2 = ||\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2
$$

since

$$\min_{\boldsymbol{\eta}} ||\boldsymbol{Z}_1 - \boldsymbol{\eta}||^2 = ||\boldsymbol{Z}_1 - \boldsymbol{Z}_1||^2 = 0$$