UNC-BIOS Theory QUAL Solutions with LaTeX

Students of Biostatistics

Last Updated: July 11, 2015

Remarks:

1. Solutions are written to be as detailed as possible for the reader.
2. Additional approaches/methods to answering questions are welcomed and can be incorporated into the solutions.
3. For questions without solutions, any and all solutions are welcome.

Contents

1 Theory 2009

1 Part 1

1.1.1 Question 1

1. Let A and B be two difference events in a probability space related to a random experiment. Suppose that n independent and identical trials of the experiment are carried out and that we observe the frequencies of occurrence of the events $A \cap B$, $A \cap B^c$, $A^c \cap B$, and $A^c \cap B^c$. The results can be summarized in the following 2×2 contingency table:

	A	A^c	Total
B	X_{11}	X_{12}	n_1
B^c	X_{21}	X_{22}	n_2
Total	m_1	m_2	n

(a) Let $p_{ij} = E\left[X_{ij}\right]/n, i = 1, 2, j = 1, 2$, where $\sum_{ij} p_{ij} = 1$. The distribution of $X = (X_{11}, X_{12}, X_{21}, X_{22})$ is multinomial, with probability function given by

$$f(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{n!}{\prod_i \prod_j x_{ij}!} \prod_i \prod_j p_{ij}^{x_{ij}}$$

Verify that this distribution is in the exponential family of distributions, and write the distribution in its canonical form.

Solution

The goal is to express $f(\cdot)$ as $f(\boldsymbol{x}) = \exp\left\{Q(\boldsymbol{x})^T\boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\boldsymbol{x})\right\}$ where

$$\boldsymbol{x} = (x_{11}, x_{12}, x_{21}, x_{22}) \text{ and } \boldsymbol{\theta} \equiv \boldsymbol{\theta}(p_{11}, p_{12}, p_{21}, p_{22})$$

And so,

$$\begin{aligned}
f(\boldsymbol{x}) &= \frac{n!}{\prod_i \prod_j x_{ij}!} \prod_i \prod_j p_{ij}^{x_{ij}} = \exp\left\{\log\left(\frac{n!}{\prod_i \prod_j x_{ij}!} \prod_i \prod_j p_{ij}^{x_{ij}}\right)\right\} \\
&= \exp\left\{\begin{bmatrix} x_{11} & x_{12} & x_{21} \end{bmatrix} \begin{bmatrix} \log\left(\frac{p_{11}}{p_{22}}\right) \\ \log\left(\frac{p_{12}}{p_{22}}\right) \\ \log\left(\frac{p_{21}}{p_{22}}\right) \end{bmatrix} - n\log\left(1 + \frac{p_{11}}{p_{22}} + \frac{p_{12}}{p_{22}} + \frac{p_{21}}{p_{22}}\right) + \log\left(\frac{n!}{\prod_i \prod_j x_{ij}!}\right)\right\} \\
&\equiv \exp\left\{\begin{bmatrix} x_{11} & x_{12} & x_{21} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} - n\log\left(1 + e^{\theta_1} + e^{\theta_2} + e^{\theta_3}\right) + \log\left(\frac{n!}{\prod_i \prod_j x_{ij}!}\right)\right\} \\
&= \exp\left\{Q(\boldsymbol{x})^T\boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\boldsymbol{x})\right\}
\end{aligned}$$

(b) Show that A and B are independent if and only if $\log\left(\frac{p_{11}}{p_{22}}\right) = \log\left(\frac{p_{12}}{p_{22}}\right) + \log\left(\frac{p_{21}}{p_{22}}\right)$.

Solution

If A and B are independent, then

$$\begin{aligned}
P(A, B) &= P(A) \times P(B) \\
\Leftrightarrow p_{11} &= (p_{11} + p_{21}) \times (p_{11} + p_{12}) \\
&= p_{11}^2 + p_{11}p_{12} + p_{11}p_{21} + p_{12}p_{21} \\
&= p_{11}(p_{11} + p_{12} + p_{21}) + p_{12}p_{21} \\
&= p_{11}(1 - p_{22}) + p_{12}p_{21} \\
\Leftrightarrow p_{11}p_{22} &= p_{12}p_{21} \\
\Leftrightarrow \frac{p_{11}}{p_{22}} &= \frac{p_{12}p_{21}}{p_{22}^2} \\
\Leftrightarrow \log\left(\frac{p_{11}}{p_{22}}\right) &= \log\left(\frac{p_{12}}{p_{22}}\right) + \log\left(\frac{p_{21}}{p_{22}}\right)
\end{aligned}$$

Therefore $A \perp\!\!\!\perp B \Leftrightarrow \log\left(\frac{p_{11}}{p_{22}}\right) = \log\left(\frac{p_{12}}{p_{22}}\right) + \log\left(\frac{p_{21}}{p_{22}}\right)$.

(c) Let $\theta = a_0 \log\left(\frac{p_{11}}{p_{22}}\right) + a_1 \log\left(\frac{p_{12}}{p_{22}}\right) + a_2 \log\left(\frac{p_{21}}{p_{22}}\right)$, where (a_0, a_1, a_2) are given constants. Assuming that $a_0 = 1$ and $a_1 = a_2 = -1$, derive a UMPU size α test for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, and derive the conditional power function of the test. (Hint: Use a theorem for multiparameter exponential families to construct the UMPU test).

Solution

Note: From BIOS 761 notes, the UMPU test corresponding to $H_0 : \theta = \theta_0$ vs. $H_A : \theta \neq \theta_0$ is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } u < c_1(t) \text{ or } u > c_2(t) \\ \gamma_i & \text{if } u = c_i(t) \\ 0 & \text{otherwise} \end{cases}$$

where $u = u(\boldsymbol{x})$ corresponds to the sufficient statistic for θ and $t = t(\boldsymbol{x})$ corresponds to the sufficient statistic(s) for ξ, the vector of nuisance parameters for multiparameter exponential families of the form

$$p_{\theta,\xi}(\boldsymbol{x}) = c(\theta,\xi)\exp\left[\theta u(\boldsymbol{x}) + \sum_{i=1}^{k} \xi_i t_i(\boldsymbol{x})\right] = c(\theta,\xi)\exp\left[\theta u(\boldsymbol{x}) + \xi^T t(\boldsymbol{x})\right]$$

and $E_{\theta_0}\left[\phi(\boldsymbol{x})|T=t\right] = \alpha$ and $E_{\theta_0}\left[u \cdot \phi(\boldsymbol{x})|T=t\right] = \alpha E_{\theta_0}\left[u|T=t\right]$.

The joint likelihood (density) is

$$\begin{aligned}
f(\boldsymbol{x}) &\propto \exp\left\{x_{11}\log\left(\frac{p_{11}}{p_{22}}\right) + x_{12}\log\left(\frac{p_{12}}{p_{22}}\right) + x_{21}\log\left(\frac{p_{21}}{p_{22}}\right)\right\} \\
&= \exp\left\{x_{11}\left[\log\left(\frac{p_{11}}{p_{22}}\right) - \log\left(\frac{p_{12}}{p_{22}}\right) - \log\left(\frac{p_{21}}{p_{22}}\right)\right] + (x_{11}+x_{12})\log\left(\frac{p_{12}}{p_{22}}\right) + (x_{11}+x_{21})\log\left(\frac{p_{21}}{p_{22}}\right)\right\} \\
&= \exp\left\{x_{11}\theta + (x_{11}+x_{12})\xi_1 + (x_{11}+x_{21})\xi_2\right\}
\end{aligned}$$

So $u(\boldsymbol{x}) = x_{11}$ is sufficient for θ and $t_1(\boldsymbol{x}) = x_{11} + x_{12}, t_2(\boldsymbol{x}) = x_{11} + x_{21}$ are sufficient for ξ. At this point, we need to know the conditional distribution of X_{11} given $X_{1\cdot} \equiv X_{11} + X_{12}$ and $X_{\cdot 1} \equiv X_{11} + X_{21}$ or in other words the conditional distribution of $U|T$.

Notice that

$$P\left(X_{11}|X_{1\cdot}, X_{\cdot 1}, n\right) = \frac{P\left(X_{11}, X_{1\cdot}, X_{\cdot 1}|n\right)}{P\left(X_{1\cdot}, X_{\cdot 1}|n\right)} = \frac{f(\boldsymbol{x})}{\sum_{\forall x_{11}} f(\boldsymbol{x})}$$

and that

$$\begin{aligned}
f(\boldsymbol{x}) &= \frac{n!}{\prod_i \prod_j x_{ij}!}\prod_i\prod_j p_{ij}^{x_{ij}} \\
&= \frac{n!}{x_{11}!(x_{1\cdot}-x_{11})!(x_{\cdot 1}-x_{11})!(n-x_{1\cdot}-x_{\cdot 1}-x_{11})!}\left(\frac{p_{11}p_{22}}{p_{12}p_{21}}\right)^{x_{11}} p_{12}^{x_{1\cdot}}p_{21}^{x_{\cdot 1}}p_{22}^{n-x_{1\cdot}-x_{\cdot 1}} \\
&= \frac{n!}{x_{11}!(x_{1\cdot}-x_{11})!(x_{\cdot 1}-x_{11})!(n-x_{1\cdot}-x_{\cdot 1}-x_{11})!}e^{\theta x_{11}} p_{12}^{x_{1\cdot}}p_{21}^{x_{\cdot 1}}p_{22}^{n-x_{1\cdot}-x_{\cdot 1}}
\end{aligned}$$

which is now a function of $X_{11}, X_{\cdot 1}, X_{1\cdot}$, and n.

Regarding the bounds of X_{11} given $X_{1\cdot}$ and $X_{\cdot 1}$, $X_{1\cdot} \geq X_{11}$ and $X_{\cdot 1} \geq X_{11}$, so $X_{11} \leq \min(X_{1\cdot}, X_{\cdot 1})$. In addition, $X_{11} \geq 0$ and $X_{11} = n - X_{12} - X_{21} - X_{22} \geq \max(0, X_{1\cdot} + X_{\cdot 1} - n)$

Therefore

$$\begin{aligned}
P\left(X_{11}|X_{1\cdot}, X_{\cdot 1}, n\right) &= \frac{\frac{n!}{x_{11}!(x_{1\cdot}-x_{11})!(x_{\cdot 1}-x_{11})!(n-x_{1\cdot}-x_{\cdot 1}-x_{11})!}e^{\theta x_{11}}p_{12}^{x_{1\cdot}}p_{21}^{x_{\cdot 1}}p_{22}^{n-x_{1\cdot}-x_{\cdot 1}}}{\sum_{k=\max(0,x_{1\cdot}+x_{\cdot 1}-n)}^{\min(x_{1\cdot},x_{\cdot 1})}\left\{\frac{n!}{k!(x_{1\cdot}-k)!(x_{\cdot 1}-k)!(n-x_{1\cdot}-x_{\cdot 1}-k)!}e^{\theta k}p_{12}^{x_{1\cdot}}p_{21}^{x_{\cdot 1}}p_{22}^{n-x_{1\cdot}-x_{\cdot 1}}\right\}} \\
&= \frac{\frac{1}{x_{11}!(x_{1\cdot}-x_{11})!(x_{\cdot 1}-x_{11})!(n-x_{1\cdot}-x_{\cdot 1}-x_{11})!}e^{\theta x_{11}}}{\sum_{k=\max(0,x_{1\cdot}+x_{\cdot 1}-n)}^{\min(x_{1\cdot},x_{\cdot 1})}\left\{\frac{1}{k!(x_{1\cdot}-k)!(x_{\cdot 1}-k)!(n-x_{1\cdot}-x_{\cdot 1}-k)!}e^{\theta k}\right\}} \\
&= \frac{x_{1\cdot}!(n-x_{1\cdot})!}{x_{1\cdot}!(n-x_{1\cdot})!}\cdot\frac{\frac{1}{x_{11}!(x_{1\cdot}-x_{11})!(x_{\cdot 1}-x_{11})!(n-x_{1\cdot}-x_{\cdot 1}-x_{11})!}e^{\theta x_{11}}}{\sum_{k=\max(0,x_{1\cdot}+x_{\cdot 1}-n)}^{\min(x_{1\cdot},x_{\cdot 1})}\left\{\frac{1}{k!(x_{1\cdot}-k)!(x_{\cdot 1}-k)!(n-x_{1\cdot}-x_{\cdot 1}-k)!}e^{\theta k}\right\}} \\
&= \frac{\binom{x_{1\cdot}}{x_{11}}\binom{n-x_{1\cdot}}{x_{\cdot 1}-x_{11}}e^{\theta x_{11}}}{\sum_{k=\max(0,x_{1\cdot}+x_{\cdot 1}-n)}^{\min(x_{1\cdot},x_{\cdot 1})}\left\{\binom{x_{1\cdot}}{k}\binom{n-x_{1\cdot}}{x_{\cdot 1}-k}e^{\theta k}\right\}}
\end{aligned}$$

And so under $H_0 : \theta = 0$,

$$\begin{aligned}
P\left(X_{11}|X_{1\cdot}, X_{\cdot 1}, n\right) &= \frac{\binom{x_{1\cdot}}{x_{11}}\binom{n-x_{1\cdot}}{x_{\cdot 1}-x_{11}}}{\sum_{k=\max(0,x_{1\cdot}+x_{\cdot 1}-n)}^{\min(x_{1\cdot},x_{\cdot 1})}\left\{\binom{x_{1\cdot}}{k}\binom{n-x_{1\cdot}}{x_{\cdot 1}-k}\right\}} \\
&= \frac{\binom{x_{1\cdot}}{x_{11}}\binom{n-x_{1\cdot}}{x_{\cdot 1}-x_{11}}}{\binom{n}{x_{\cdot 1}}} \\
\Rightarrow X_{11}|X_{1\cdot}, X_{\cdot 1}, n &\sim HG(n, X_{1\cdot}, X_{\cdot 1})
\end{aligned}$$

Since $X_{11}|X_{1\cdot}, X_{\cdot 1}, n$ follows a hypergeometric distribution under H_0, we know that

$$E\left[X_{11}|X_{1\cdot}, X_{\cdot 1}, n\right] = \frac{X_{1\cdot}X_{\cdot 1}}{n}.$$

To derive the conditional power function of the test, first note that the test is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } u < c_1(t) \text{ or } u > c_2(t) \\ \gamma_1 & \text{if } u = c_1(t) \\ \gamma_2 & \text{if } u = c_2(t) \\ 0 & \text{o.w.} \end{cases}$$

where the constants $\gamma_1, \gamma_2, c_1(t)$, and $c_2(t)$ are already derived. Conditional power is the expectation of $\phi(\boldsymbol{x})$ given $T = t$ under H_1. In other words,

$$\begin{aligned}
\text{Cond. Power} &= E_{H_1}\left[\phi(\boldsymbol{x})|T = t\right] \\
&= P\left(u < c_1(t) \text{ or } u > c_2(t)|T = t\right) + \gamma_1 P\left(u = c_1(t)|T = t\right) + \\
&\quad \gamma_2 P\left(u = c_2(t)|T = t\right) \\
&= 1 - P\left(c_1(t) \le u \le c_2(t)|T = t\right) + \gamma_1 P\left(u = c_1(t)|T = t\right) + \\
&\quad \gamma_2 P\left(u = c_2(t)|T = t\right) \\
&= 1 - \sum_{j=c_1(t)}^{c_2(t)}\left\{P\left(x_{11} = j|x_{1\cdot}, x_{\cdot 1}, n\right)\right\} + \gamma_1 P\left(x_{11} = c_1(t)|x_{1\cdot}, x_{\cdot 1}, n\right) + \\
&\quad \gamma_2 P\left(x_{11} = c_2(t)|x_{1\cdot}, x_{\cdot 1}, n\right)
\end{aligned}$$

where $P\left(x_{11} = j|x_{1\cdot}, x_{\cdot 1}, n\right) = \dfrac{\binom{x_{1\cdot}}{j}\binom{n-x_{1\cdot}}{x_{\cdot 1}-j}e^{\theta j}}{\sum_{k=\max(0,x_{1\cdot}+x_{\cdot 1}-n)}^{\min(x_{1\cdot},x_{\cdot 1})}\left\{\binom{x_{1\cdot}}{k}\binom{n-x_{1\cdot}}{x_{\cdot 1}-k}e^{\theta k}\right\}}$

Side Note: In this question, I don't think one needs to show how the constants are derived but I included the steps to do so for anyone who wants a refresher.

To obtain $\gamma_1, \gamma_2, c_1(t), c_2(t)$ we need to solve two equations under the null. First,

$$\begin{aligned}
\alpha &= E\left[\phi(\boldsymbol{x})|T = t\right] \\
&= P\left(u < c_1(t) \text{ or } u > c_2(t)|T = t\right) + \gamma_1 P\left(u = c_1(t)|T = t\right) + \gamma_2 P\left(u = c_2(t)|T = t\right) \\
&= 1 - P\left(c_1(t) \le u \le c_2(t)|T = t\right) + \gamma_1 P\left(u = c_1(t)|T = t\right) + \gamma_2 P\left(u = c_2(t)|T = t\right) \\
&= 1 - \sum_{j=c_1(t)}^{c_2(t)}\left\{\frac{\binom{x_{1\cdot}}{j}\binom{n-x_{1\cdot}}{x_{\cdot 1}-j}}{\binom{n}{x_{\cdot 1}}}\right\} + \gamma_1\frac{\binom{x_{1\cdot}}{c_1(t)}\binom{n-x_{1\cdot}}{x_{\cdot 1}-c_1(t)}}{\binom{n}{x_{\cdot 1}}} + \gamma_2\frac{\binom{x_{1\cdot}}{c_2(t)}\binom{n-x_{1\cdot}}{x_{\cdot 1}-c_2(t)}}{\binom{n}{x_{\cdot 1}}}
\end{aligned}$$

Second, $E\left[u \cdot \phi(\boldsymbol{x}) | T=t\right] = \alpha E\left[u | T=t\right]$. For the LHS,

$$
\begin{aligned}
E\left[u \cdot \phi(\boldsymbol{x}) | T=t\right] &= E\left[u \cdot \phi(\boldsymbol{x}) \cdot \mathbf{1}\left\{u < c_1(t) \text{ or } u > c_2(t)\right\} | T=t\right] + E\left[u \cdot \phi(\boldsymbol{x}) \cdot \mathbf{1}\left\{u = c_1(t)\right\} | T=t\right] + \\
&\quad E\left[u \cdot \phi(\boldsymbol{x}) \cdot \mathbf{1}\left\{u = c_2(t)\right\} | T=t\right] \\
&= E\left[u \cdot \mathbf{1}\left\{u < c_1(t) \text{ or } u > c_2(t)\right\} | T=t\right] + E\left[c_1(t) \cdot \gamma_1 \cdot \mathbf{1}\left\{u = c_1(t)\right\} | T=t\right] + \\
&\quad E\left[c_2(t) \cdot \gamma_2 \cdot \mathbf{1}\left\{u = c_2(t)\right\} | T=t\right] \\
&= \sum_{j=\max(0, x_{1\cdot}+x_{\cdot 1}-n)}^{c_1(t)-1} j \left\{ \frac{\binom{x_{1\cdot}}{j}\binom{n-x_{1\cdot}}{x_{\cdot 1}-j}}{\binom{n}{x_{\cdot 1}}} \right\} + \sum_{j=c_2(t)+1}^{\min(x_{1\cdot}, x_{\cdot 1})} j \left\{ \frac{\binom{x_{1\cdot}}{j}\binom{n-x_{1\cdot}}{x_{\cdot 1}-j}}{\binom{n}{x_{\cdot 1}}} \right\} + \\
&\quad c_1(t) \cdot \gamma_1 \cdot \frac{\binom{x_{1\cdot}}{c_1(t)}\binom{n-x_{1\cdot}}{x_{\cdot 1}-c_1(t)}}{\binom{n}{x_{\cdot 1}}} + c_2(t) \cdot \gamma_2 \cdot \frac{\binom{x_{1\cdot}}{c_2(t)}\binom{n-x_{1\cdot}}{x_{\cdot 1}-c_2(t)}}{\binom{n}{x_{\cdot 1}}}
\end{aligned}
$$

For the RHS,

$$
\alpha E\left[u | T=t\right] = \alpha \cdot \frac{x_{1\cdot} x_{\cdot 1}}{n}
$$

(d) Derive a UMPU size α test for testing $H_0 : P(A) \geq P(B)$ versus $H_1 : P(A) < P(B)$. (Hint: Use the techniques of part (c) in setting up the hypothesis in terms of θ and then constructing the test).

Solution

Notice that

$$\begin{aligned} P(A) \geq P(B) &\Leftrightarrow p_{11} + p_{21} \geq p_{11} + p_{12} \\ &\Leftrightarrow p_{21} \geq p_{12} \\ &\Leftrightarrow \frac{p_{21}}{p_{22}} \geq \frac{p_{12}}{p_{22}} \\ &\Leftrightarrow \log\left(\frac{p_{21}}{p_{22}}\right) - \log\left(\frac{p_{12}}{p_{22}}\right) \geq 0 \end{aligned}$$

So let's define $\theta \equiv \log\left(\frac{p_{21}}{p_{22}}\right) - \log\left(\frac{p_{12}}{p_{22}}\right)$ and hence our hypothesis test can be re-expressed as

$$H_0 : \theta \geq 0 \text{ vs. } H_1 : \theta \leq 0.$$

Defining θ above as such will become apparent later (but one can always redefine θ).

Note: From BIOS 761 notes, the UMPU test for multi-parameter exponential families of the form $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ is $\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } u > c(t) \\ \gamma & \text{if } u = c(t) \\ 0 & \text{o.w.} \end{cases}$ but since the hypotheses for this question have reversed inequalities, simply reverse the inequalities to get

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } u < c(t) \\ \gamma & \text{if } u = c(t) \\ 0 & \text{o.w.} \end{cases}$$

Using part (a) but re-expressing $f(\boldsymbol{x})$ for this question,

$$\begin{aligned} f(\boldsymbol{x}) &\propto \exp\left\{x_{11}\log\left(\frac{p_{11}}{p_{22}}\right) + x_{12}\log\left(\frac{p_{12}}{p_{22}}\right) + x_{21}\log\left(\frac{p_{21}}{p_{22}}\right)\right\} \\ &= \exp\left\{x_{11}\log\left(\frac{p_{11}}{p_{22}}\right) + (x_{12}+x_{21})\log\left(\frac{p_{12}}{p_{22}}\right) + x_{21}\left[\log\left(\frac{p_{21}}{p_{22}}\right) - \log\left(\frac{p_{12}}{p_{22}}\right)\right]\right\} \end{aligned}$$

So $u(\boldsymbol{x}) = x_{21}$ is sufficient for θ and $t_1(\boldsymbol{x}) = x_{11}$, $t_2(\boldsymbol{x}) = x_{12} + x_{21}$ are sufficient for ξ. Now we need to find the conditional distribution $X_{21}|X_{11}, X_{12} + X_{21}$. For the multinomial distribution, X_{21} is not independent of X_{11} but since we're conditioning on X_{11} **AND** $X_{12}+X_{21}$, we have that $X_{21} \perp\!\!\!\perp X_{11}|X_{12} + X_{21}$. The conditional distribution of $U|T$ is binomially distributed. More specifically,

$$\begin{aligned} P\left(X_{21} = j|X_{12} + X_{21} = k\right) &= \binom{k}{j}\left(\frac{p_{21}}{p_{12}+p_{21}}\right)^j\left(1 - \frac{p_{21}}{p_{12}+p_{21}}\right)^{k-j} \\ &= \binom{k}{j}\left(\frac{p_{21}}{p_{12}+p_{21}}\right)^j\left(\frac{p_{12}}{p_{12}+p_{21}}\right)^{k-j} \\ &\equiv \binom{k}{j} p_f^j(1-p_f)^{k-j} \end{aligned}$$

Therefore $E\left[u|T = t\right] = k \cdot p_f$. To find $c(t)$ and γ, solve the equation below under H_0.

$$\begin{aligned} \alpha &= E\left[\phi(\boldsymbol{x})|T = t\right] \\ &= P\left(u < c(t)|T = t\right) + \gamma P\left(u = c(t)|T = t\right) \\ &= 1 - P\left(u \geq c(t)|T = t\right) + \gamma P\left(u = c(t)|T = t\right) \\ &= 1 - \sum_{j=c(t)}^{k}\left\{\binom{k}{j} p_f^j(1-p_f)^{k-j}\right\} + \gamma \cdot \binom{k}{c(t)} p_f^{c(t)}(1-p_f)^{k-c(t)} \end{aligned}$$

(e) Derive the likelihood ratio statistic, denoted by Λ_n, for the hypothesis in part (c) and show that it is asymptotically equivalent to the Pearson chi-square statistic. Specifically,

(i) show that

$$-2\log(\Lambda_n) = \sum_{j=1}^{2}\sum_{i=1}^{2}\frac{(X_{ij}-n\widehat{p}_{ij})^2}{n\widehat{p}_{ij}} + o_p(1),$$

where $\widehat{p}_{ij}$ denotes the maximum likelihood estimate of p_{ij} under H_0.

Solution

The likelihood ratio statistic Λ_n is defined as

$$\Lambda_n = \frac{\sup_{\theta\in\Theta_0} L(\theta|\boldsymbol{x})}{\sup_{\theta\in\Theta_0\cup\Theta_1} L(\theta|\boldsymbol{x})} = \frac{\prod_i\prod_j \widehat{p}_{ij}^{x_{ij}}}{\prod_i\prod_j \widehat{\pi}_{ij}^{x_{ij}}}$$

where $\widehat{\pi}_{ij} = \frac{x_{ij}}{n}$, the MLE under $\Theta_0\cup\Theta_1$, the unrestricted parameter space.
Since the expression on the RHS above contains $o_p(1)$, this suggests that we'll need to expand the LHS. I'll use a Taylor series expansion.

$$\begin{aligned}
-2\log(\Lambda_n) &= -2\sum_i\sum_j x_{ij}\log\left(\frac{\widehat{p}_{ij}}{\widehat{\pi}_{ij}}\right) = -2\sum_i\sum_j n\widehat{\pi}_{ij}\log\left(\frac{\widehat{p}_{ij}}{\widehat{\pi}_{ij}}\right)\\
&= -2n\sum_i\sum_j \widehat{\pi}_{ij}\log\left(\frac{\widehat{p}_{ij}}{\widehat{\pi}_{ij}}\right)\\
&= 2n\sum_i\sum_j \widehat{\pi}_{ij}\log\left(\frac{\widehat{\pi}_{ij}}{\widehat{p}_{ij}}\right)
\end{aligned}$$

At this point, I'll Taylor expand the term $\widehat{\pi}_{ij}\log\left(\frac{\widehat{\pi}_{ij}}{\widehat{p}_{ij}}\right)$ centered around $\widehat{p}_{ij}$.

<u>**Note**</u>: To expand a function $f(x)$ centered around a point a, we have

$$\begin{aligned}
f(x) &= \sum_{j=0}^{\infty}\frac{\partial_x^j f(x)\big|_{x=a}\cdot(x-a)^j}{j!}\\
&= f(a)+f'(a)(x-a)+\frac{f''(a)(x-a)^2}{2!}+o_p((x-a)^2)
\end{aligned}$$

In this case, $f(x) = x\log\left(\frac{x}{a}\right)$. Hence

$$\begin{aligned}
f(x) &= (x-a)+\frac{(x-a)^2}{2a}+o_p\left((x-a)^2\right)\\
\Rightarrow \widehat{\pi}_{ij}\log\left(\frac{\widehat{\pi}_{ij}}{\widehat{p}_{ij}}\right) &= (\widehat{\pi}_{ij}-\widehat{p}_{ij})+\frac{(\widehat{\pi}_{ij}-\widehat{p}_{ij})^2}{2\widehat{p}_{ij}}+o_p\left((\widehat{\pi}_{ij}-\widehat{p}_{ij})^2\right)
\end{aligned}$$

And now, we have that

$$\begin{aligned}
-2\log(\Lambda_n) &= 2n\sum_i\sum_j \widehat{\pi}_{ij}\log\left(\frac{\widehat{\pi}_{ij}}{\widehat{p}_{ij}}\right)\\
&= 2n\sum_i\sum_j\left\{(\widehat{\pi}_{ij}-\widehat{p}_{ij})+\frac{(\widehat{\pi}_{ij}-\widehat{p}_{ij})^2}{2\widehat{p}_{ij}}+o_p\left((\widehat{\pi}_{ij}-\widehat{p}_{ij})^2\right)\right\}\\
&\quad \text{Note: } \sum_i\sum_j(\widehat{\pi}_{ij}-\widehat{p}_{ij}) = 0\\
&= 2n\sum_i\sum_j\left\{\frac{(\widehat{\pi}_{ij}-\widehat{p}_{ij})^2}{2\widehat{p}_{ij}}+o_p\left((\widehat{\pi}_{ij}-\widehat{p}_{ij})^2\right)\right\}\\
&= \sum_i\sum_j\frac{(n\widehat{\pi}_{ij}-n\widehat{p}_{ij})^2}{n\widehat{p}_{ij}}+\frac{1}{2}\sum_i\sum_j\left\{o_p\left(\left(\sqrt{n}(\widehat{\pi}_{ij}-\widehat{p}_{ij})\right)^2\right)\right\}\\
&= \sum_i\sum_j\frac{(X_{ij}-n\widehat{p}_{ij})^2}{n\widehat{p}_{ij}}+\sum_i\sum_j\left\{o_p\left(\frac{1}{2}\left(\sqrt{n}(\widehat{\pi}_{ij}-\widehat{p}_{ij})\right)^2\right)\right\}
\end{aligned}$$

Looking at the term $\sqrt{n}(\widehat{\pi}_{ij}-\widehat{p}_{ij})$, we see that

$$\begin{aligned}
\sqrt{n}(\widehat{\pi}_{ij}-\widehat{p}_{ij}) &= \sqrt{n}(\widehat{\pi}_{ij}-p_{ij}+p_{ij}-\widehat{p}_{ij})\\
&= \sqrt{n}(\widehat{\pi}_{ij}-p_{ij})-\sqrt{n}(\widehat{p}_{ij}-p_{ij})
\end{aligned}$$

For the first term, by the CLT where $\sigma_{ij}^2 = V[\mathbf{1}\{A_i, B_j\}]$

$$\sqrt{n}\left(\widehat{\pi}_{ij} - p_{ij}\right) = \sqrt{n}\left(\frac{X_{ij}}{n} - p_{ij}\right) \to_d \mathcal{N}\left(0, \sigma_{ij}^2\right) \Rightarrow \sqrt{n}\left(\widehat{\pi}_{ij} - p_{ij}\right) = O_p(1)$$

For the second term and using part (c), $p_{ij} = p_{i\cdot}p_{\cdot j}$. Hence $\sqrt{n}\left(\widehat{p}_{ij} - p_{ij}\right) = \sqrt{n}\left(\widehat{p}_{i\cdot}\widehat{p}_{\cdot j} - p_{i\cdot}p_{\cdot j}\right)$. Using MLE theory and the Delta method, we know that

$$\sqrt{n}\left(g(\widehat{p}_{i\cdot}, \widehat{p}_{\cdot j}) - g(p_{i\cdot}, p_{\cdot j})\right) \to_d \nabla g(p_{i\cdot}, p_{\cdot j}) \cdot \mathcal{N}(0, \Sigma) \Rightarrow \sqrt{n}\left(\widehat{p}_{ij} - p_{ij}\right) = O_p(1)$$

We now see that

$$\begin{aligned}
\sqrt{n}(\widehat{\pi}_{ij} - \widehat{p}_{ij}) &= O_p(1) + O_p(1) \\
& \text{Note: } O_p(1) + O_p(1) = O_p(1) \\
&= O_p(1) \\
\Rightarrow \left(\sqrt{n}(\widehat{\pi}_{ij} - \widehat{p}_{ij})\right)^2 &= \left(O_p(1)\right)^2 \\
& \text{Note: } \left(O_p(1)\right)^2 = O_p(1) \\
&= O_p(1) \\
\Rightarrow o_p\left(\frac{1}{2}\left(\sqrt{n}(\widehat{\pi}_{ij} - \widehat{p}_{ij})\right)^2\right) &= o_p\left(O_p(1)\right) \\
& \text{Note: } o_p\left(O_p(1)\right) = o_p(1) \\
&= o_p(1)
\end{aligned}$$

Finally, $\displaystyle\sum_i\sum_j \left\{o_p\left(\frac{1}{2}\left(\sqrt{n}(\widehat{\pi}_{ij} - \widehat{p}_{ij})\right)^2\right)\right\} = \sum_i\sum_j o_p(1)$. Note that $o_p(1) + o_p(1) = o_p(1)$.
We've now shown that

$$-2\log\left(\Lambda_n\right) = \sum_{j=1}^{2}\sum_{i=1}^{2}\frac{(X_{ij} - n\widehat{p}_{ij})^2}{n\widehat{p}_{ij}} + o_p(1),$$

(ii) find the asymptotic distribution of $-2\log\left(\Lambda_n\right)$ under H_0 and H_1.

Solution

We know that under H_0, as $n \to \infty, -2\log(\Lambda_n) \to_d \chi_r^2$ where $r = 1$ in this case.
Under H_1, the test statistic converges in distribution to noncentral χ^2 with 1 degree of freedom and noncentrality parameter $\gamma = \displaystyle\sum_i\sum_j \frac{(np_{ij} - np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}}$.

1.1.2 Question 2, (e) incomplete

2. Let λ have exponential density $\theta e^{-\theta\lambda}$, for $0 < \theta < \infty$. Conditional on λ, let (X, Y) be a pair of independent Poisson random variables with respect p.m.f.'s $\lambda^x e^{-\lambda}/x!, x = 0, 1, \ldots,$ and $(\beta\lambda)^y e^{-\beta\lambda}/y!, y = 0, 1, \ldots,$ for $0 < \beta < \infty$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d sample where (X_1, Y_1) has the same unconditional joint distribution as (X, Y). Do the following:

   (a) Determine the following properties of the unconditional distribution of (X, Y):

      (i) Show that $E[X] = \theta^{-1}, E[Y] = \beta\theta^{-1}, V[X] = \theta^{-1} + \theta^{-2}, V[Y] = \beta\theta^{-1} + \beta^2\theta^{-2}$, and $\text{cov}(X, Y) = \beta\theta^{-2}$.

      Solution

      We are told that

$$\begin{aligned}
\lambda \sim Exp(\theta) &\Rightarrow E[\lambda] = \theta^{-1}, V[\lambda] = \theta^{-2} \\
P(X, Y|\lambda) &= P(X|\lambda)\, P(Y|\lambda) \\
X|\lambda \sim Poisson(\lambda) &\Rightarrow E[X|\lambda] = V[X|\lambda] = \lambda \\
Y|\lambda \sim Poisson(\beta\lambda) &\Rightarrow E[Y|\lambda] = V[Y|\lambda] = \beta\lambda
\end{aligned}$$

      Therefore

$$\begin{aligned}
E[X] &= E_\lambda[E[X|\lambda]] = E_\lambda[\lambda] \\
&= \theta^{-1} \\
E[Y] &= E_\lambda[E[Y|\lambda]] = E_\lambda[\beta\lambda] \\
&= \beta\theta^{-1} \\
V[X] &= E_\lambda[V[X|\lambda]] + V_\lambda[E[X|\lambda]] \\
&= E_\lambda[\lambda] + V_\lambda[\lambda] \\
&= \theta^{-1} + \theta^{-2} \\
V[Y] &= E_\lambda[V[Y|\lambda]] + V_\lambda[E[Y|\lambda]] \\
&= E_\lambda[\lambda] + V_\lambda[\lambda] \\
&= E_\lambda[\beta\lambda] + V_\lambda[\beta\lambda] \\
&= \beta\theta^{-1} + \beta^2\theta^{-2} \\
\text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\
&= E_\lambda[E[XY|\lambda]] - \theta^{-1} \cdot \beta\theta^{-1} \\
&= E_\lambda[E[X|\lambda]\, E[Y|\lambda]] - \beta\theta^{-2} \\
&= E_\lambda[\lambda \cdot \beta\lambda] - \beta\theta^{-2} \\
&= \beta \cdot E_\lambda[\lambda^2] - \beta\theta^{-2} \\
&\quad \text{Note: } E_\lambda[\lambda^2] = \left(V[\lambda] + (E[\lambda])^2\right) = \left(\theta^{-2} + \theta^{-2}\right) = 2\theta^{-2} \\
&= 2\beta\theta^{-2} - \beta\theta^{-2} \\
&= \beta\theta^{-2}
\end{aligned}$$

(ii) Show that the unconditional joint density of (X, Y) is

$$\left(\frac{\theta}{\theta+\beta+1}\right)\frac{(x+y)!}{x!y!}\left(\frac{1}{\theta+\beta+1}\right)^{x}\left(\frac{\beta}{\theta+\beta+1}\right)^{y}$$

Solution
The joint density is

$$\begin{aligned}
f(X,Y) &= \int_{\lambda} f(X,Y,\lambda)d\lambda = \int_{\lambda} f\left(X,Y|\lambda\right)f(\lambda)d\lambda \\
&= \int_{\lambda} f\left(X|\lambda\right)f\left(Y|\lambda\right)f(\lambda)d\lambda \\
&= \int_{\lambda} \frac{\lambda^{x}e^{-\lambda}}{x!}\cdot\frac{(\beta\lambda)^{y}e^{-\beta\lambda}}{y!}\cdot\theta e^{-\theta\lambda}d\lambda \\
&= \frac{\theta\beta^{y}}{x!y!}\cdot\int_{\lambda}\lambda^{x+y}e^{-(1+\beta+\theta)\lambda}d\lambda \\
&= \frac{\theta\beta^{y}}{x!y!}\cdot\int_{\lambda}\frac{\Gamma(x+y+1)(1+\beta+\theta)^{x+y+1}}{\Gamma(x+y+1)(1+\beta+\theta)^{x+y+1}}\cdot\lambda^{x+y}e^{-(1+\beta+\theta)\lambda}d\lambda \\
&= \frac{\theta\beta^{y}}{x!y!}\cdot\frac{\Gamma(x+y+1)}{(1+\beta+\theta)^{x+y+1}}\cdot\int_{0}^{\infty}\frac{(1+\beta+\theta)^{x+y+1}}{\Gamma(x+y+1)}\cdot\lambda^{x+y}e^{-(1+\beta+\theta)\lambda}d\lambda \\
&= \frac{\theta\beta^{y}}{x!y!}\cdot\frac{\Gamma(x+y+1)}{(1+\beta+\theta)^{x+y+1}} \\
&= \left(\frac{\theta}{\theta+\beta+1}\right)\frac{(x+y)!}{x!y!}\left(\frac{1}{\theta+\beta+1}\right)^{x}\left(\frac{\beta}{\theta+\beta+1}\right)^{y}
\end{aligned}$$

(b) Show that the maximum likelihood estimator based on a sample of size n for θ is $\widehat{\theta}_n = \bar{X}_n^{-1}$ and for β is $\widehat{\beta}_n = \bar{Y}_n/\bar{X}_n$, where $\bar{X}_n = n^{-1}\sum_{i=1}^{n}X_i$ and $\bar{Y}_n = n^{-1}\sum_{i=1}^{n}Y_i$.

Solution
Looking at the joint likelihood and log likelihood,

$$\begin{aligned}
L\left(\theta,\beta|\boldsymbol{x},\boldsymbol{y}\right) &= \prod_{i=1}^{n}L\left(x_i,y_i|\theta,\beta\right) = \prod_{i=1}^{n}\frac{\theta\beta^{y_i}}{(1+\beta+\theta)^{x_i+y_i+1}}\cdot\frac{(x_i+y_i)!}{x_i!y_i!} \\
&= \frac{\theta^{n}\beta^{\sum_i y_i}}{(1+\beta+\theta)^{\sum_i x_i+\sum_i y_i+n}}\cdot\prod_{i=1}^{n}\frac{(x_i+y_i)!}{x_i!y_i!} \\
\Rightarrow l_n(\theta,\beta) &= \log\left(L\left(\theta,\beta|\boldsymbol{x},\boldsymbol{y}\right)\right) \\
&= n\log(\theta)+\log(\beta)\sum_i y_i - \left(\sum_i x_i+\sum_i y_i+n\right)\log(1+\beta+\theta)+\log\left(\prod_{i=1}^{n}\frac{(x_i+y_i)!}{x_i!y_i!}\right)
\end{aligned}$$

Maximize the log likelihood, $l_n(\theta,\beta)$, with respect to θ and β by setting the first derivatives equal to 0. After some algebraic manipulation of the two equations, we'll get

$$\widehat{\theta}_n = \bar{X}_n^{-1} \text{ and } \widehat{\beta}_n = \bar{Y}_n/\bar{X}_n$$

(c) Show that

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}_n-\theta\\ \widehat{\beta}_n-\beta\end{pmatrix}\to_d N\left(0,\begin{bmatrix}\theta^2(\theta+1) & \beta\theta^2\\ \beta\theta^2 & \theta\beta(\beta+1)\end{bmatrix}\right).$$

Solution

There are two ways to verify the convergence.

(1) Use **CLT** and **Delta Method** (we can incorporate what we know from part (a))

(2) Use **MLE theory** assuming regularity conditions hold and invert Fisher's Information matrix.

To save time, I'll use approach (1).

By CLT,

$$\sqrt{n}\left(\begin{bmatrix}\bar{X}_n\\ \bar{Y}_n\end{bmatrix}-\begin{bmatrix}E[X]\\ E[Y]\end{bmatrix}\right)\to_d \mathcal{N}\left(\mathbf{0},\begin{bmatrix}V[X] & \mathrm{Cov}(X,Y)\\ \mathrm{Cov}(X,Y) & V[Y]\end{bmatrix}\right)$$

In this case,

$$\sqrt{n}\left(\begin{bmatrix}\bar{X}_n\\ \bar{Y}_n\end{bmatrix}-\begin{bmatrix}\theta^{-1}\\ \beta\theta^{-1}\end{bmatrix}\right)\to_d \mathcal{N}\left(\mathbf{0},\begin{bmatrix}\theta^{-1}+\theta^{-2} & \beta\theta^{-2}\\ \beta\theta^{-2} & \beta\theta^{-1}+\beta^2\theta^{-2}\end{bmatrix}\right)$$

By Delta Method, if $\sqrt{n}\left(\widehat{\xi}-\xi\right)\to_d \mathcal{N}(0,\Sigma)$ then

$$\sqrt{n}\left(g(\widehat{\xi})-g(\xi)\right)\to_d \nabla g(\xi)\cdot\mathcal{N}(0,\Sigma)$$

Let us define

$$\begin{aligned}
g(x,y) &\equiv \begin{bmatrix}\frac{1}{x} & \frac{y}{x}\end{bmatrix}\\
\Rightarrow \nabla g(x,y) &= \begin{bmatrix}-\frac{1}{x^2} & 0\\ -\frac{y}{x^2} & \frac{1}{x}\end{bmatrix}\\
&= x^{-1}\begin{bmatrix}-x^{-1} & 0\\ -yx^{-1} & 1\end{bmatrix}\\
\Rightarrow \nabla g(\theta^{-1},\beta\theta^{-1}) &= \left(\theta^{-1}\right)^{-1}\begin{bmatrix}-\left(\theta^{-1}\right)^{-1} & 0\\ -\left(\beta\theta^{-1}\right)\left(\theta^{-1}\right)^{-1} & 1\end{bmatrix}\\
&= \theta\begin{bmatrix}-\theta & 0\\ -\beta & 1\end{bmatrix}
\end{aligned}$$

So we have that

$$\begin{aligned}
\sqrt{n}\begin{pmatrix}\widehat{\theta}_n-\theta\\ \widehat{\beta}_n-\beta\end{pmatrix} &= \sqrt{n}\left(g\left(\bar{X}_n,\bar{Y}_n\right)-g\left(\theta^{-1},\beta\theta^{-1}\right)\right)\\
&\to_d \theta\begin{bmatrix}-\theta & 0\\ -\beta & 1\end{bmatrix}\cdot\mathcal{N}\left(\mathbf{0},\begin{bmatrix}\theta^{-1}+\theta^{-2} & \beta\theta^{-2}\\ \beta\theta^{-2} & \beta\theta^{-1}+\beta^2\theta^{-2}\end{bmatrix}\right)\\
&= \mathcal{N}\left(\mathbf{0},\theta\begin{bmatrix}-\theta & 0\\ -\beta & 1\end{bmatrix}\cdot\begin{bmatrix}\theta^{-1}+\theta^{-2} & \beta\theta^{-2}\\ \beta\theta^{-2} & \beta\theta^{-1}+\beta^2\theta^{-2}\end{bmatrix}\cdot\theta\begin{bmatrix}-\theta & 0\\ -\beta & 1\end{bmatrix}^T\right)\\
&= \mathcal{N}\left(\mathbf{0},\theta^2\begin{bmatrix}\theta(1+\theta^{-1}) & \beta\\ \beta & \beta\theta^{-1}(\beta+1)\end{bmatrix}\cdot\begin{bmatrix}-\theta & -\beta\\ 0 & 1\end{bmatrix}\right)\\
&= \mathcal{N}\left(\mathbf{0},\begin{bmatrix}\theta^2(\theta+1) & \beta\theta^2\\ \beta\theta^2 & \theta\beta(\beta+1)\end{bmatrix}\right)
\end{aligned}$$

(d) Let $T_1 = \sqrt{n\bar{X}_n/2}\left(\bar{Y}_n/\bar{X}_n - 1\right)$ and $T_2 = \sqrt{n\bar{X}_n/2}\ln\left(\bar{Y}_n/\bar{X}_n\right)$, and show that under the null hypothesis $H_0 : \beta = 1$, that

(i) $T_1 \to_d N(0,1)$,

Solution

Re-express T_1 in terms of $\widehat{\theta}_n$ and $\widehat{\beta}_n$.

$$\begin{aligned} T_1 &= \sqrt{\frac{n\bar{X}_n}{2}}\left(\frac{\bar{Y}_n}{\bar{X}_n} - 1\right) \\ &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{\widehat{\theta}_n}}\sqrt{n}\left(\widehat{\beta}_n - 1\right) \end{aligned}$$

Using the results of part (c), $\sqrt{n}\left(\widehat{\beta}_n - \beta\right) \to_d \mathcal{N}\left(0, \theta\beta(\beta+1)\right)$. So under $H_0 : \beta = 1$,

$$\begin{aligned} \sqrt{n}\left(\widehat{\beta}_n - 1\right) &\to_d \mathcal{N}(0, 2\theta) \\ \Rightarrow \frac{1}{\sqrt{2}}\sqrt{n}\left(\widehat{\beta}_n - 1\right) &\to_d \mathcal{N}(0, \theta) \end{aligned}$$

By the Weak Law of Large Numbers and Continuous Mapping Theorem,

$$\begin{aligned} \bar{X}_n &\to_p \theta \\ \Rightarrow \frac{1}{\sqrt{\bar{X}_n}} &\to_p \frac{1}{\sqrt{\theta}} \end{aligned}$$

Finally, by Slutsky's Theorem,

$$\frac{1}{\sqrt{2}}\frac{1}{\sqrt{\widehat{\theta}_n}}\sqrt{n}\left(\widehat{\beta}_n - 1\right) \to_d \mathcal{N}(0,1)$$

(ii) $T_1 - T_2 \to_p 0$,

Solution

Re-express $T_1 - T_2$ in terms of $\widehat{\theta}_n$ and $\widehat{\beta}_n$.

$$\begin{aligned} T_1 - T_2 &= \sqrt{\frac{n\bar{X}_n}{2}}\left(\frac{\bar{Y}_n}{\bar{X}_n} - 1\right) - \sqrt{\frac{n\bar{X}_n}{2}}\ln\left(\frac{\bar{Y}_n}{\bar{X}_n}\right) \\ &= \sqrt{\frac{n\bar{X}_n}{2}}\left(\frac{\bar{Y}_n}{\bar{X}_n} - \ln\left(\frac{\bar{Y}_n}{\bar{X}_n}\right) - 1\right) \\ &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{\widehat{\theta}_n}}\sqrt{n}\left(\widehat{\beta}_n - \ln\left(\widehat{\beta}_n\right) - 1\right) \end{aligned}$$

Using the results of part (i), $\sqrt{n}\left(\widehat{\beta}_n - 1\right) \to_d \mathcal{N}\left(0, 2\theta\right)$. From the Delta Method, we have

$$\sqrt{n}\left(\widehat{\beta}_n - \ln\left(\widehat{\beta}_n\right) - 1\right) = \sqrt{n}\left(g\left(\widehat{\beta}_n\right) - g\left(1\right)\right) \to_d \nabla g(1) \cdot \mathcal{N}(0, 2\theta)$$

where $g(x) = x - \ln(x)$. This implies that $\nabla g(x) = 1 - \frac{1}{x}$. And so, $\nabla g(1) = 0$. We then have that

$$\sqrt{n}\left(\widehat{\beta}_n - \ln\left(\widehat{\beta}_n\right) - 1\right) \to_d \mathcal{N}(0,0) = 0$$

We know that if $X \to_d c$ where c is a constant, then $X \to_p c$. So

$$\sqrt{n}\left(\widehat{\beta}_n - \ln\left(\widehat{\beta}_n\right) - 1\right) \to_p 0$$

Also from part (i), $\frac{1}{\sqrt{2\widehat{\theta}_n}} \to_p \frac{1}{\sqrt{2\theta}}$. Finally, through the Continuous Mapping Theorem, if $(X_n, Y_n) \to_p (X, Y)$ and let $g(x,y) = xy$, then $g(X_n, Y_n) \to_p g(X, Y)$. In this case,

$$T_1 - T_2 = \frac{1}{\sqrt{2\widehat{\theta}_n}}\sqrt{n}\left(\widehat{\beta}_n - \ln\left(\widehat{\beta}_n\right) - 1\right) \to_p 0$$

(iii) $T_2 \to_d N(0,1)$

Solution

Re-express T_2 in terms of $\widehat{\beta}_n$.

$$T_2 = \sqrt{\frac{n\bar{X}_n}{2}}\ln\left(\frac{\bar{Y}_n}{\bar{X}_n}\right) = \frac{1}{\sqrt{2\widehat{\theta}_n}}\sqrt{n}\left(\ln(\widehat{\beta}_n) - \ln(1)\right)$$

Using the results of part (i), $\sqrt{n}\left(\widehat{\beta}_n - 1\right) \to_d \mathcal{N}\left(0, 2\theta\right)$. From the Delta Method, we have

$$\sqrt{n}\left(\ln\left(\widehat{\beta}_n\right) - \ln(1)\right) = \sqrt{n}\left(g\left(\widehat{\beta}_n\right) - g\left(1\right)\right) \to_d \nabla g(1) \cdot \mathcal{N}(0, 2\theta)$$

where $g(x) = \ln(x)$ and hence $\nabla g(x) = \dfrac{1}{x}$. And so, $\nabla g(1) = 1$. We then have that

$$\sqrt{n}\left(\ln(\widehat{\beta}_n) - \ln(1)\right) \to_d \mathcal{N}(0, 2\theta)$$

Again from part (i), $\dfrac{1}{\sqrt{2\widehat{\theta}_n}} \to_p \dfrac{1}{\sqrt{2\theta}}$. Through Slutsky's Theorem, we have that

$$\frac{1}{\sqrt{2\widehat{\theta}_n}}\sqrt{n}\left(\ln(\widehat{\beta}_n) - \ln(1)\right) \to_d \mathcal{N}(0, 1)$$

(e) Suppose $\beta = 1$ and we wish to make inference on $\tau = \theta/(\theta+2)$. Using result (ii) of part (a), derive the Bayes estimator of τ under squared error loss with prior density

$$\pi(\tau) \propto \tau^{a_0-1}(1-\tau)^{b_0-1},$$

where the scalars $a_0 > 0$ and $b_0 > 0$ are specified hyperparameters. Show that this Bayes estimator is admissible.

Solution

First, notice that

$$\begin{aligned} \tau &= \frac{\theta}{\theta+2} \Leftrightarrow \theta = \frac{2\tau}{1-\tau} \\ \Rightarrow \frac{1}{\theta+2} &= \frac{1-\tau}{2} \end{aligned}$$

Second, note from part (a) (ii) that

$$P(x, y|\theta, \beta) = \left(\frac{\theta}{\theta+\beta+1}\right) \frac{(x+y)!}{x!y!} \left(\frac{1}{\theta+\beta+1}\right)^x \left(\frac{\beta}{\theta+\beta+1}\right)^y$$

and with $\beta = 1$, we have

$$\begin{aligned} P(x, y|\theta) &= \left(\frac{\theta}{\theta+2}\right) \frac{(x+y)!}{x!y!} \left(\frac{1}{\theta+2}\right)^x \left(\frac{1}{\theta+2}\right)^y \\ &= \left(\frac{\theta}{\theta+2}\right) \frac{(x+y)!}{x!y!} \left(\frac{1}{\theta+2}\right)^{x+y} \\ &= \tau \left(\frac{1-\tau}{2}\right)^{x+y} \frac{(x+y)!}{x!y!} = P(x, y|\tau) \end{aligned}$$

For a single observation (X, Y), the posterior distribution of τ is

$$\begin{aligned} P(\tau|a_0, b_0, x, y) &\propto P(x, y|\tau) \cdot \pi(\tau) \\ &\propto \tau(1-\tau)^{x+y}\tau^{a_0-1}(1-\tau)^{b_0-1} \\ &= \tau^{a_0}(1-\tau)^{x+y+b_0-1} \\ \Rightarrow (\tau|a_0, b_0, x, y) &\sim Beta(a_0+1, b_0+x+y) \\ \Rightarrow E[\tau|a_0, b_0, x, y] &= \frac{a_0+1}{a_0+1+b_0+x+y} \end{aligned}$$

Note: If we observed n iid observations, then

$$E[\tau|a_0, b_0, \boldsymbol{x}, \boldsymbol{y}] = \frac{a_0+n}{a_0+n+b_0+\sum_i x_i + \sum_i y_i}$$

As we know, under squared error loss, the Bayes estimator is unique and equals the posterior mean. So the unique Bayes estimator of τ is

$$d_\tau = \frac{a_0+n}{a_0+n+b_0+\sum_i x_i + \sum_i y_i}$$

Not sure how to show this is admissible. Maybe calculate risk and show it's finite?? There's a theorem in 761 notes stating that if a Bayes estimator is unique and has finite risk then it's admissible.

1.1.3 Question 3

3. Suppose that there is a random variable A having discrete probability distribution with support on the non-negative integers. Assume that $P(A=j)=p_j, j=0,1,2,\ldots$, such that $p_j>0$ and $\sum_j p_j=1$.

   Define the random variables $B_n, n=1,2,\ldots$ recursively:

$$B_{n+1}=\sum_{k=1}^{B_n} A_{nk},$$

   where $A_{nk}, n=1,2,\ldots,k=1,\ldots,B_n$ are iid with the same distribution as A. Assume that B_0 is a known positive integer and let $P_{ij}=P\left(B_{n+1}=j|B_n=i\right)=P\left(\sum_{k=1}^{i} A_{nk}=j\right)$.

   (a) Show that, in general, $E\left(B_{n+1}\right)=E\left(B_n\right)E\left(A_{nk}\right)$ and $\operatorname{Var}\left(B_{n+1}\right)=E\left(B_n\right)\operatorname{Var}\left(A_{nk}\right)+\operatorname{Var}\left(B_n\right)\left\{E\left(A_{nk}\right)\right\}^2$, $n\geq 1$.

   Solution

   To prove the expectation,

$$\begin{aligned}
E\left[B_{n+1}\right] &= E\left[\sum_{k=1}^{B_n} A_{nk}\right] \\
&= E\left[E\left[\sum_{k=1}^{B_n} A_{nk}\middle| B_n\right]\right] \\
&= E\left[\sum_{k=1}^{B_n} E\left[A_{nk}|B_n\right]\right]=E\left[B_n\cdot E\left[A_{nk}|B_n\right]\right] \\
&= E\left[B_n\cdot E\left[A_{nk}\right]\right] \\
&= E\left[B_n\right]E\left[A_{nk}\right]
\end{aligned}$$

   To prove the variance,

$$\begin{aligned}
V\left[B_{n+1}\right] &= V\left[\sum_{k=1}^{B_n} A_{nk}\right] \\
&= E\left[V\left[\sum_{k=1}^{B_n} A_{nk}\middle| B_n\right]\right]+V\left[E\left[\sum_{k=1}^{B_n} A_{nk}\middle| B_n\right]\right] \\
&= E\left[\sum_{k=1}^{B_n} V\left[A_{nk}|B_n\right]\right]+V\left[\sum_{k=1}^{B_n} E\left[A_{nk}|B_n\right]\right] \\
&= E\left[B_n\cdot V\left[A_{nk}|B_n\right]\right]+V\left[B_n\cdot E\left[A_{nk}|B_n\right]\right] \\
&= E\left[B_n\cdot V\left[A_{nk}\right]\right]+V\left[B_n\cdot E\left[A_{nk}\right]\right] \\
&= E\left[B_n\right]V\left[A_{nk}\right]+V\left[B_n\right]\left\{E\left[A_{nk}\right]\right\}^2
\end{aligned}$$

In the sequel, suppose that $B_0 = 1$ and that $E(A_{nk}) = \mu < \infty$ and $\operatorname{Var}(A_{nk}) = \sigma^2 < \infty, n = 1, 2, \ldots, k = 1, \ldots, B_n$.

(b) Show that for $n \geq 1, E(B_n) = \mu^n$ and $\operatorname{Var}(B_n) = \sigma^2 \mu^{n-1}(1-\mu^n)\{1-\mu\}^{-1}$ if $\mu \neq 1$ and $n\sigma^2$ if $\mu = 1$.

Solution

To prove the expectation using the result from part (a),

$$\begin{aligned}
E[B_n] &= E[B_{n-1}]\,E[A_{n-1,k}] \\
&\quad \text{Note: We're told that } E[A_{nk}] = \mu \\
&= E[B_{n-1}] \cdot \mu \\
&= E[B_{n-2}]\,E[A_{n-2,k}] \cdot \mu \\
&= E[B_{n-2}] \cdot \mu^2 \\
&= \vdots \\
&= E[B_0] \cdot \mu^n \\
&\quad \text{Note: } B_0 = 1 \\
&= \mu^n
\end{aligned}$$

To prove the variance using the result from part (a),

$$\begin{aligned}
V[B_{n+1}] &= E[B_n]\,V[A_{nk}] + V[B_n]\{E[A_{nk}]\}^2 \\
&\quad \text{Note: We've just shown that } E[B_n] = \mu^n \\
&\quad \text{Note: We're told } E[A_{nk}] = \mu, V[A_{nk}] = \sigma^2 \\
&= \mu^n\sigma^2 + V[B_n]\,\mu^2 \\
&= \mu^n\sigma^2 + \left(\mu^{n-1}\sigma^2 + V[B_{n-1}]\,\mu^2\right)\mu^2 \\
&= \mu^n\sigma^2 + \mu^{n+1}\sigma^2 + V[B_{n-1}]\,\mu^4 \\
&= \mu^n\sigma^2 + \mu^{n+1}\sigma^2 + \left(\mu^{n-2}\sigma^2 + V[B_{n-2}]\,\mu^2\right)\mu^4 \\
&= \mu^n\sigma^2 + \mu^{n+1}\sigma^2 + \mu^{n+2}\sigma^2 + V[B_{n-2}]\,\mu^6 \\
&= \vdots \\
&= \mu^n\sigma^2 + \mu^{n+1}\sigma^2 + \cdots + \mu^{n+(n-1)}\sigma^2 + V\left[B_{n-(n-1)}\right]\mu^{2n} \\
&= \sigma^2\mu^{n-1}\sum_{k=1}^{n}\mu^k + V[B_1]\,\mu^{2n} \\
&\quad \text{Note: } B_1 \equiv \sum_{k=1}^{B_0} A_{0k} = A_{0k} \Rightarrow V[B_1] = V[A_{0k}] = \sigma^2 \\
&= \sigma^2\mu^{n-1}\sum_{k=1}^{n}\mu^k + \sigma^2\mu^{2n} = \sigma^2\mu^{n-1}\left(\sum_{k=1}^{n}\mu^k + \mu^{n+1}\right) \\
&= \sigma^2\mu^{n-1}\sum_{k=1}^{n+1}\mu^k
\end{aligned}$$

Regarding partial sums, let $\sum_{k=1}^{a} b^k \equiv S$. So

$$\begin{aligned}
S &= b + b^2 + \cdots + b^a \\
&= b\left(1 + b + b^2 + \cdots + b^{a-1}\right) \\
&= b\left(1 + b + b^2 + \cdots + b^{a-1} + b^a - b^a\right) \\
&= b\left(1 + S - b^a\right) \\
\Leftrightarrow S &= b(1-b^a)(1-b)^{-1}
\end{aligned}$$

Using this, we see that

$$\sum_{k=1}^{n+1}\mu^k = \mu(1-\mu^{n+1})(1-\mu)^{-1}$$

Hence

$$\begin{aligned}
V[B_{n+1}] &= \sigma^2\mu^{n-1} \cdot \mu(1-\mu^{n+1})(1-\mu)^{-1} \\
&= \sigma^2\mu^n(1-\mu^{n+1})(1-\mu)^{-1} \\
\Rightarrow V[B_n] &= \sigma^2\mu^{n-1}(1-\mu^n)(1-\mu)^{-1}
\end{aligned}$$

For $\mu = 1$, simply apply L'Hopital's rule. In the end, we see that for $n \geq 1$,

$$V[B_n] = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \sigma^2\mu^{n-1}(1-\mu^n)(1-\mu)^{-1} & \text{if } \mu \neq 1 \end{cases}$$

(c) Under the conditions in (b), show that if $\mu < 1$ then $P(B_n = 0) \to 1$ as $n \to \infty$.

Solution

Assuming $\mu < 1$ and using part (b),

$$E[B_n] = \mu^n \text{ and } V[B_n] = \sigma^2 \mu^{n-1}(1-\mu^n)(1-\mu)^{-1}.$$

Hence as $n \to \infty$, $E[B_n] \to 0$.

Notice that if $B_n = 0$, this can expressed as $|B_n| < \epsilon$ for a given $\epsilon > 0$. So we can express $P(B_n = 0) = a$ for some unknown a between 0 and 1 as (for a given $\epsilon > 0$)

$$\begin{aligned}
a &= P(|B_n| < \epsilon) \\
\Leftrightarrow 1 - a &= P(|B_n| \geq \epsilon) \\
&\quad \text{Note: Since } B_n \text{ is nonnegative, } |B_n| = B_n \\
\Rightarrow 1 - a &= P(B_n \geq \epsilon)
\end{aligned}$$

Note: The **Markov inequality** states that for nonnegative random variable X and $k > 0$,

$$P(X \geq k) \leq \frac{E[X]}{k}$$

And so

$$\begin{aligned}
1 - a &= P(B_n \geq \epsilon) \\
&\leq \frac{E[B_n]}{\epsilon} \\
&= \frac{\mu^n}{\epsilon} \to 0 \text{ as } n \to \infty
\end{aligned}$$

This implies that $1 - a \leq 0$ or that $1 \leq a$ but $0 \leq a \leq 1$. Therefore $a = 1$. Finally we see that for $\epsilon > 0$,

$$P(B_n < \epsilon) = P(B_n = 0) = 1$$

(d) Define $P_{1j}^{(n)} = P(B_n = j | B_0 = 1)$. Show that $E\left(z^{B_n}\right)$, denoted by $\phi_n(z)$, may be expressed as $\sum_{j=0}^{\infty} P_{1j}^{(n)} z^j$, for a scalar z such that $|z| \leq 1$.

Demonstrate that $\phi_n(0) = P(B_n = 0 | B_0 = 1)$.

Solution

To prove the expectation,

$$\begin{aligned}
\phi_n(z) &\equiv E\left[z^{B_n}\right] \\
&= E\left[z^{B_n} | B_0 = 1\right] \\
&= \sum_{j=0}^{\infty} \left\{ z^j P(B_n = j | B_0 = 1) \right\} \\
&\equiv \sum_{j=0}^{\infty} \left\{ P_{1j}^{(n)} z^j \right\} \\
&= P_{10}^{(n)} + P_{11}^{(n)} z + P_{12}^{(n)} z^2 + \cdots
\end{aligned}$$

When $z = 0$, we have

$$\phi_n(0) = P_{10}^{(n)} = P(B_n = 0 | B_0 = 1)$$

(e) Define $\phi(z) = E\left(z^{A_{nk}}\right) = \sum_{j=0}^{\infty} p_j z^j$, for a scalar z, with $|z| \leq 1$, for $k = 1, \dots, B_n$. Show that $\sum_{j=0}^{\infty} P\left(A_{n1} + A_{n2} + \cdots + A_{nk} = j | B_n = k\right) z^j = \{\phi(z)\}^k$.

Solution

We can see that

$$\begin{aligned}
\phi(z) &= E\left[z^{A_{nk}}\right] = \sum_{j=0}^{\infty} z^j P\left(A_{nk} = j\right) \\
&= \sum_{j=0}^{\infty} P\left(A = j\right) z^j \\
&= \sum_{j=0}^{\infty} p_j z^j
\end{aligned}$$

And so,

$$\begin{aligned}
\sum_{j=0}^{\infty} P\left(A_{n1} + \cdots + A_{nk} = j | B_n = k\right) z^j &= \sum_{j=0}^{\infty} P\left(\sum_{k'=0}^{B_n} A_{nk'} = j \middle| B_n = k\right) z^j \\
&= E\left[z^{\sum_{k'=0}^{B_n} A_{nk'}} \middle| B_n = k\right] \\
&= E\left[z^{A_{n1}} \middle| B_n = k\right] E\left[z^{A_{n2}} \middle| B_n = k\right] \cdots E\left[z^{A_{nk}} \middle| B_n = k\right] \\
&= E\left[z^{A_{n1}}\right] E\left[z^{A_{n2}}\right] \cdots E\left[z^{A_{nk}}\right] \\
&= \{\phi(z)\}^k
\end{aligned}$$

(f) Establish that the following recursive relationship holds:

$$\phi_n(z) = \phi_{n-1}\left\{\phi(z)\right\}, n \geq 1$$

(Hint: condition on B_{n-1}).

Solution

Starting with $\phi_n(z)$, we have

$$\begin{aligned}
\phi_n(z) &= E\left[z^{B_n}\right] = \sum_{j=0}^{\infty} P\left(B_n = j\right) z^j \\
&= \sum_{j=0}^{\infty}\left\{\sum_{k=0}^{\infty} P\left(B_n = j, B_{n-1} = k\right)\right\} z^j \\
&= \sum_{j=0}^{\infty}\left\{\sum_{k=0}^{\infty} P\left(B_n = j | B_{n-1} = k\right) P\left(B_{n-1} = k\right) z^j\right\} \\
&= \sum_{k=0}^{\infty} P\left(B_{n-1} = k\right)\left\{\sum_{j=0}^{\infty} P\left(B_n = j | B_{n-1} = k\right) z^j\right\}
\end{aligned}$$

Using part (e),

$$\begin{aligned}
\sum_{j=0}^{\infty} P\left(B_n = j | B_{n-1} = k\right) z^j &= \sum_{j=0}^{\infty} P\left(\sum_{m=0}^{B_{n-1}} A_{n-1,m} = j \middle| B_{n-1} = k\right) z^j \\
&= E\left[z^{\sum_{m=0}^{B_{n-1}} A_{n-1,m}} \middle| B_{n-1} = k\right] \\
&= \left\{\phi(z)\right\}^k
\end{aligned}$$

Using this intermediate result, we have

$$\begin{aligned}
\phi_n(z) &= \sum_{k=0}^{\infty} P\left(B_{n-1} = k\right)\left\{\phi(z)\right\}^k \\
&= E\left[\left\{\phi(z)\right\}^{B_{n-1}}\right] \\
&= \phi_{n-1}\left\{\phi(z)\right\} \text{ for } n \geq 1
\end{aligned}$$

1 Part 2

Table 1: n pairs of binary observations

		Y_2		
		0	1	Total
Y_1	0	n_{00}	n_{01}	n_{0+}
	1	n_{10}	n_{11}	n_{1+}
Total		n_{+0}	n_{+1}	n

1.2.1 Question 1

1. Table 1, a 2×2 contingency table, is based on n independent pairs of binary observations $(y_{i1}, y_{i2}), i = 1, \cdots, n$ from a cross-sectional study, where $Y_{ik} = 1$ denotes 'success' and 0 denotes 'failure' for $k = 1, 2$.

   (a) Assume $P(Y_{i1} = j, Y_{i2} = k) = \pi_{jk}$ for all i. Derive the maximum likelihood estimates of π_{jk}, denoted by $\widehat{\pi}_{jk}$, based on Table 1 and show that

   $$\sqrt{n}\left(\widehat{\pi}_{00} - \pi_{00}, \widehat{\pi}_{01} - \pi_{01}, \widehat{\pi}_{10} - \pi_{10}, \widehat{\pi}_{11} - \pi_{11}\right)^T$$

   converges in distribution to a multivariate normal random vector with zero-mean and covariance $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T$, where $\boldsymbol{\pi} = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$.

   Solution

   Starting with the density function and letting $\mathbf{n} = (n_{00}, n_{01}, n_{10}, n_{11})$, we have

   $$\begin{aligned} P(\mathbf{n}|\boldsymbol{\pi}) &= \frac{n!}{\prod_j \prod_k n_{ik}!} \prod_j \prod_k \pi_{jk}^{n_{jk}} \text{ with } \sum_j \sum_k \pi_{jk} = 1 \text{ and } \sum_j \sum_k n_{jk} = n \\ \Leftrightarrow l_n(\boldsymbol{\pi}) &\equiv \log\left(P(\mathbf{n}|\boldsymbol{\pi})\right) = \log\left(\frac{n!}{\prod_j \prod_k n_{jk}!}\right) + \sum_j \sum_k n_{jk} \log(\pi_{jk}) \\ & \quad \text{Note: Substitute } \pi_{11} = 1 - \pi_{00} - \pi_{01} - \pi_{10} \\ &= C + n_{00}\log(\pi_{00}) + n_{01}\log(\pi_{01}) + n_{10}\log(\pi_{10}) + n_{11}\log(1 - \pi_{00} - \pi_{01} - \pi_{10}) \end{aligned}$$

   Maximize the log likelihood with respect to the three parameters.

   $$\begin{aligned} \frac{\partial l_n(\boldsymbol{\pi})}{\partial \pi_{jk}} &= \frac{n_{jk}}{\pi_{jk}} - \frac{n_{11}}{1 - \pi_{00} - \pi_{01} - \pi_{10}} = 0 \\ \Leftrightarrow \frac{n_{jk}}{\pi_{jk}} &= \frac{n_{11}}{1 - \pi_{00} - \pi_{01} - \pi_{10}} \\ \Leftrightarrow \frac{n_{00}}{\pi_{00}} = \frac{n_{01}}{\pi_{01}} = \frac{n_{10}}{\pi_{10}} &= \frac{n_{11}}{1 - \pi_{00} - \pi_{01} - \pi_{10}} \\ \Leftrightarrow \frac{n}{1} &= \frac{n_{00}}{\pi_{00}} \\ \Rightarrow \widehat{\pi}_{00} &= \frac{n_{00}}{n} \\ \Rightarrow \widehat{\pi}_{jk} &= \frac{n_{jk}}{n} \text{ for } j = 0, 1; k = 0, 1 \end{aligned}$$

   By the Weak Law of Large Numbers, we have that

   $$\widehat{\pi}_{jk} = \frac{n_{jk}}{n} = \frac{\sum_{i=1}^n \mathbf{1}_{\{Y_{i1}=j, Y_{i2}=k\}}}{n} \rightarrow_p E\left[\mathbf{1}_{\{Y_{i1}=j, Y_{i2}=k\}}\right] = \pi_{jk}$$

   Using the Central Limit Theorem, we have that

   $$\sqrt{n}\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}\right) \rightarrow_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

   To calculate the jth and kth elements of $\boldsymbol{\Sigma}$, looking at $[\boldsymbol{\Sigma}]_{jk} = \text{Cov}\left(\mathbf{1}_{\{Y_{i1}=j\}}\mathbf{1}_{\{Y_{i2}=k\}}, \mathbf{1}_{\{Y_{i1}=j'\}}\mathbf{1}_{\{Y_{i2}=k'\}}\right)$

   $$\begin{aligned} &= E\left[\mathbf{1}_{\{Y_{i1}=j\}}\mathbf{1}_{\{Y_{i2}=k\}}\mathbf{1}_{\{Y_{i1}=j'\}}\mathbf{1}_{\{Y_{i2}=k'\}}\right] - E\left[\mathbf{1}_{\{Y_{i1}=j\}}\mathbf{1}_{\{Y_{i2}=k\}}\right] E\left[\mathbf{1}_{\{Y_{i1}=j'\}}\mathbf{1}_{\{Y_{i2}=k'\}}\right] \\ &= \begin{cases} \pi_{jk} - \pi_{jk}^2 & \text{if } j = j', k = k' \\ -\pi_{jk}\pi_{j'k'} & \text{o.w.} \end{cases} \end{aligned}$$

   So the covariance is

   $$\boldsymbol{\Sigma} = \begin{bmatrix} \pi_{00} - \pi_{00}^2 & -\pi_{00}\pi_{01} & -\pi_{00}\pi_{10} & -\pi_{00}\pi_{11} \\ -\pi_{00}\pi_{01} & \pi_{01} - \pi_{01}^2 & -\pi_{01}\pi_{10} & -\pi_{01}\pi_{11} \\ -\pi_{00}\pi_{10} & -\pi_{01}\pi_{10} & \pi_{10} - \pi_{10}^2 & -\pi_{10}\pi_{11} \\ -\pi_{00}\pi_{11} & -\pi_{01}\pi_{11} & -\pi_{10}\pi_{11} & \pi_{11} - \pi_{11}^2 \end{bmatrix} = \text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T$$

(b) Under the assumptions in (a), further assume that $\pi_{1+} = \pi_{11}+\pi_{10} = \exp(\alpha)/[1+\exp(\alpha)]$ and $\pi_{+1} = \pi_{11}+\pi_{01} = \exp(\alpha+\beta)/[1+\exp(\alpha+\beta)]$. Using the results from (a), construct an estimator for $\alpha, \beta)$, denoted by $(\widehat{\alpha}_M, \widehat{\beta}_M)$. Use delta method to show that the asymptotic variance of $\sqrt{n}(\widehat{\beta}_M - \beta)$ is

$$(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} - 2(\pi_{11}\pi_{00} - \pi_{10}\pi_{01})/(\pi_{+1}\pi_{+0}\pi_{1+}\pi_{0+}).$$

Solution

We're told that $\pi_{1+} = \dfrac{\exp(\alpha)}{1+\exp(\alpha)}$ which equals $P(Y_{i1} = 1)$ and $\pi_{+1} = \dfrac{\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)}$ which equals $P(Y_{i2} = 1)$. Therefore

$$P(Y_1 = y_{i1}) = \frac{\exp(\alpha y_{i1})}{1+\exp(\alpha)} \text{ and } P(Y_2 = y_{i2}) = \frac{\exp((\alpha+\beta)y_{i2})}{1+\exp(\alpha+\beta)}$$

To calculate the MLEs, we look at the joint likelihood and log likelihood.

$$\begin{aligned}
L(\boldsymbol{y}_1, \boldsymbol{y}_2|\alpha,\beta) &= \prod_{i=1}^{n} \{P(Y_1 = y_{i1}) \cdot P(Y_2 = y_{i2})\} \\
&= \prod_{i=1}^{n} \left\{ \frac{\exp(\alpha y_{i1})}{1+\exp(\alpha)} \cdot \frac{\exp((\alpha+\beta)y_{i2})}{1+\exp(\alpha+\beta)} \right\} \\
&= \frac{\exp\left(\alpha \sum_i y_{i1}\right)}{[1+\exp(\alpha)]^n} \cdot \frac{\exp\left((\alpha+\beta)\sum_i y_{i2}\right)}{[1+\exp(\alpha+\beta)]^n} \\
\Rightarrow l_n(\alpha,\beta) &\equiv \log\left(L(\boldsymbol{y}_1, \boldsymbol{y}_2|\alpha,\beta)\right) \\
&= -n\log[1+\exp(\alpha)] + \alpha\sum_i y_{i1} - n\log[1+\exp(\alpha+\beta)] + (\alpha+\beta)\sum_i y_{i2}
\end{aligned}$$

Taking the first derivatives of the log likelihood and solving for α and β, we have

$$\begin{aligned}
\frac{\partial l_n(\alpha,\beta)}{\partial\beta} &= -\frac{n\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} + \sum_i y_{i2} = 0 \\
\Leftrightarrow \bar{y}_2 &= \frac{\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} \\
\Leftrightarrow \beta &= \log\left(\frac{\bar{y}_2}{1-\bar{y}_2}\right) - \alpha \\
\frac{\partial l_n(\alpha,\beta)}{\partial\alpha} &= -\frac{n\exp(\alpha)}{1+\exp(\alpha)} + \sum_i y_{i1} - \frac{n\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} + \sum_i y_{i2} = 0 \\
\Leftrightarrow \bar{y}_1 + \bar{y}_2 &= \frac{\exp(\alpha)}{1+\exp(\alpha)} + \frac{\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} \\
\Leftrightarrow \bar{y}_1 + \bar{y}_2 &= \frac{\exp(\alpha)}{1+\exp(\alpha)} + \bar{y}_2 \\
\Rightarrow \widehat{\alpha}_M &= \log\left(\frac{\bar{y}_1}{1-\bar{y}_1}\right) \\
\Rightarrow \widehat{\beta}_M &= \log\left(\frac{\bar{y}_2}{1-\bar{y}_2}\right) - \log\left(\frac{\bar{y}_1}{1-\bar{y}_1}\right)
\end{aligned}$$

Noticing that the MLEs are functions of $\bar{y}_1$ and $\bar{y}_2$, we can use CLT combined with Delta Method to obtain the asymptotic variance of $\widehat{\beta}_M$. By CLT,

$$\sqrt{n}\left(\begin{bmatrix}\bar{y}_1\\ \bar{y}_2\end{bmatrix} - \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}\right) \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_b)$$

We need to find the expectations and covariance matrix $\boldsymbol{\Sigma}_b$.

$$\begin{aligned}
\mu_1 &= E[y_{i1}] = P(Y_1 = 1) = \pi_{1+} \\
\mu_2 &= E[y_{i2}] = P(Y_2 = 1) = \pi_{+1} \\
V[y_{i1}] &= \pi_{1+}(1-\pi_{1+}) = \pi_{1+}\pi_{0+} \\
V[y_{i2}] &= \pi_{+1}(1-\pi_{+1}) = \pi_{+1}\pi_{+0} \\
\mathrm{Cov}(y_{i1}, y_{i2}) &= E[y_{i1}y_{i2}] - E[y_{i1}]E[y_{i2}] \\
&= \pi_{11} - \pi_{1+}\pi_{+1} \\
&= \pi_{11}\pi_{00} - \pi_{10}\pi_{01} \\
\Rightarrow \boldsymbol{\Sigma}_b &= \begin{bmatrix} \pi_{1+}\pi_{0+} & \pi_{11}\pi_{00} - \pi_{10}\pi_{01} \\ \pi_{11}\pi_{00} - \pi_{10}\pi_{01} & \pi_{+1}\pi_{+0} \end{bmatrix} \\
&\equiv \begin{bmatrix} a(1-a) & z \\ z & c(1-c) \end{bmatrix}
\end{aligned}$$

where $a = \pi_{1+}, z = \pi_{11}\pi_{00} - \pi_{10}\pi_{01}, c = \pi_{+1}$.

To use the Delta Method on $\sqrt{n}\left(\widehat{\beta}_M - \beta\right) = \sqrt{n}\left(g\left(\begin{bmatrix}\bar{y}_1\\ \bar{y}_2\end{bmatrix}\right) - g\left(\begin{bmatrix}\pi_{1+}\\ \pi_{+1}\end{bmatrix}\right)\right)$, define

$$\begin{aligned}
g(a,c) &= \log\left(\frac{c}{1-c}\right) - \log\left(\frac{a}{1-a}\right) \\
&= \log(c) - \log(1-c) - \log(a) + \log(1-a) \\
\Rightarrow \nabla g(a,c) &= \begin{bmatrix}-\dfrac{1}{a(1-a)} & \dfrac{1}{c(1-c)}\end{bmatrix}
\end{aligned}$$

And so we have,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_M - \beta\right) &= \sqrt{n}\left(g\left(\begin{bmatrix}\bar{y}_1\\ \bar{y}_2\end{bmatrix}\right) - g\left(\begin{bmatrix}\pi_{1+}\\ \pi_{+1}\end{bmatrix}\right)\right) \to_d \nabla g(a,c)\cdot \mathcal{N}(0,\mathbf{\Sigma}_b) \\
&= \mathcal{N}\left(0, \nabla g(a,c)\cdot \mathbf{\Sigma}_b \cdot \nabla g(a,c)^T\right) \\
&= \mathcal{N}\left(0, \begin{bmatrix}-\dfrac{1}{a(1-a)} & \dfrac{1}{c(1-c)}\end{bmatrix}\begin{bmatrix}a(1-a) & z\\ z & c(1-c)\end{bmatrix}\begin{bmatrix}-\dfrac{1}{a(1-a)}\\ \dfrac{1}{c(1-c)}\end{bmatrix}\right) \\
&= \mathcal{N}\left(0, \begin{bmatrix}-1+\dfrac{z}{c(1-c)} & -\dfrac{z}{a(1-a)}+1\end{bmatrix}\begin{bmatrix}-\dfrac{1}{a(1-a)}\\ \dfrac{1}{c(1-c)}\end{bmatrix}\right) \\
&= \mathcal{N}\left(0, \frac{1}{a(1-a)} - \frac{z}{ac(1-a)(1-c)} - \frac{z}{ac(1-a)(1-c)} + \frac{1}{c(1-c)}\right) \\
&= \mathcal{N}\left(0, \frac{1}{a(1-a)} - \frac{2z}{ac(1-a)(1-c)} + \frac{1}{c(1-c)}\right) \\
&= \mathcal{N}\left(0, (\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} - \frac{2\left(\pi_{11}\pi_{00} - \pi_{10}\pi_{01}\right)}{\pi_{1+}\pi_{0+}\pi_{+1}\pi_{+0}}\right)
\end{aligned}$$

(c) Consider the subject-specific model

$$P\left(Y_{i1}=1\right)=\frac{\exp(\alpha_i)}{1+\exp(\alpha_i)}, P\left(Y_{i2}=1\right)=\frac{\exp(\alpha_i+\beta)}{1+\exp(\alpha_i+\beta)}.$$

Assume independence of (Y_{i1}, Y_{i2}) across subjects, that is, across i. Show that $s_i = y_{i1} + y_{i2}$ is a sufficient statistic for α_i and that the conditional maximum likelihood estimate for β given $s_i, i = 1, \ldots, n$ is $\widehat{\beta}_C = \log\left(n_{01}/(n_* - n_{01})\right)$, where $n_* = n_{01} + n_{10}$. Derive that the asymptotic variance of $\sqrt{n}\left(\widehat{\beta}_C - \beta\right)$ is $1/\pi_{01} + 1/\pi_{10}$.

Solution

Given the subject-specific parameterization, we have

$$P\left(Y_1=y_{i1}\right)=\frac{\exp(\alpha_i y_{i1})}{1+\exp(\alpha_i)} \text{ and } P\left(Y_2=y_{i2}\right)=\frac{\exp((\alpha_i+\beta)y_{i2})}{1+\exp(\alpha_i+\beta)}$$

Looking at the density or likelihood for individual i,

$$\begin{aligned}
P\left(Y_1=y_{i1}\right)\cdot P\left(Y_2=y_{i2}\right) &= \frac{\exp(\alpha_i y_{i1})}{1+\exp(\alpha_i)}\cdot\frac{\exp((\alpha_i+\beta)y_{i2})}{1+\exp(\alpha_i+\beta)}\\
&= \frac{\exp(\alpha_i(y_{i1}+y_{i2})+\beta y_{i2})}{[1+\exp(\alpha_i)][1+\exp(\alpha_i+\beta)]}
\end{aligned}$$

And hence $S_i \equiv Y_{i1}+Y_{i2}$ is the sufficient statistic for α_i. Let's derive $P\left(Y_1=y_{i1}, Y_2=y_{i2}|S=s_i\right)$. Notice that for $S=0,2$, the conditional probability is deterministic aka $P\left(Y_1=y_{i1}, Y_2=y_{i2}|S=0,2\right)=1$. But for $S_i=1$,

$$\begin{aligned}
P\left(Y_1=y_{i1}, Y_2=y_{i2}|S_i=1\right) &= \frac{P\left(Y_1=y_{i1}, Y_2=y_{i2}, S_i=1\right)}{P\left(S_i=1\right)}\\
&= \frac{P\left(Y_1=1-y_{i2}, Y_2=y_{i2}\right)}{P\left(Y_1=1, Y_2=0\right)+P\left(Y_1=0, Y_2=1\right)}\\
&= \frac{\dfrac{\exp(\alpha_i+\beta y_{i2})}{[1+\exp(\alpha_i)][1+\exp(\alpha_i+\beta)]}}{\dfrac{\exp(\alpha_i)}{[1+\exp(\alpha_i)][1+\exp(\alpha_i+\beta)]}+\dfrac{\exp(\alpha_i+\beta)}{[1+\exp(\alpha_i)][1+\exp(\alpha_i+\beta)]}}\\
&= \frac{\exp(\alpha_i+\beta y_{i2})}{\exp(\alpha_i)+\exp(\alpha_i+\beta)}\\
&= \frac{\exp(\beta y_{i2})}{1+\exp(\beta)}
\end{aligned}$$

Look at the joint conditional likelihood and log likelihood to calculate the conditional MLE for β. But first, let's define the set $\mathcal{S} = \{i : y_{i1}+y_{i2}=1\}$.

$$\begin{aligned}
L\left(\boldsymbol{y}_1, \boldsymbol{y}_2|\mathbf{s}=1\right) &= \prod_{\forall i\in\mathcal{S}} P\left(Y_1=y_{i1}, Y_2=y_{i2}|S_i=1\right)\\
&= \prod_{\forall i\in\mathcal{S}} \frac{\exp(\beta y_{i2})}{1+\exp(\beta)}\\
&= \frac{\exp\left(\beta\sum_{\forall i\in\mathcal{S}} y_{i2}\right)}{[1+\exp(\beta)]^{n_R}}\\
&\quad \text{Note: } \sum_{\forall i\in\mathcal{S}} y_{i2} = n_{01} \text{ and } n_R = n_{10}+n_{01}\\
&= \frac{\exp\left(\beta n_{01}\right)}{[1+\exp(\beta)]^{n_R}}\\
\Rightarrow l_n(\beta) &\equiv \log\left(L\left(\boldsymbol{y}_1, \boldsymbol{y}_2|\mathbf{s}=1\right)\right)\\
&= \beta n_{01} - n_R\log\left[1+\exp(\beta)\right]
\end{aligned}$$

Maximize the conditional log likelihood with respect to β.

$$\begin{aligned}
\frac{\partial l_n(\beta)}{\partial\beta} &= n_{01} - n_R\frac{\exp(\beta)}{1+\exp(\beta)} = 0\\
\Leftrightarrow \frac{n_{01}}{n_R} &= \frac{\exp(\beta)}{1+\exp(\beta)}\\
\Leftrightarrow \exp(\beta) &= \frac{\dfrac{n_{01}}{n_R}}{1-\dfrac{n_{01}}{n_R}} = \frac{n_{01}}{n_R-n_{01}} = \frac{n_{01}}{n_{10}}\\
\Rightarrow \widehat{\beta}_C &= \log\left(\frac{n_{01}}{n_{10}}\right) = \log\left(\frac{n_{01}/n}{n_{10}/n}\right)\\
&= \log\left(\frac{\widehat{\pi}_{01}}{\widehat{\pi}_{10}}\right)
\end{aligned}$$

For the asymptotic variance of $\widehat{\beta}_C$, notice that the estimator is a function of $\widehat{\pi}_{01}$ and $\widehat{\pi}_{10}$. Recall from part (a) that

$$\begin{aligned}
\sqrt{n}\left(\begin{bmatrix}\widehat{\pi}_{01}\\ \widehat{\pi}_{10}\end{bmatrix} - \begin{bmatrix}\pi_{01}\\ \pi_{10}\end{bmatrix}\right) &\to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_c) \\
&= \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\pi_{01} - \pi_{01}^2 & -\pi_{01}\pi_{10}\\ -\pi_{01}\pi_{10} & \pi_{10} - \pi_{10}^2\end{bmatrix}\right) \\
&\equiv \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}a(1-a) & -ab\\ -ab & b(1-b)\end{bmatrix}\right)
\end{aligned}$$

To apply the Delta Method on $\sqrt{n}\left(\log\left(\frac{\widehat{\pi}_{01}}{\widehat{\pi}_{10}}\right) - \log\left(\frac{\pi_{01}}{\pi_{10}}\right)\right) = \sqrt{n}\left(g\left(\begin{bmatrix}\widehat{\pi}_{01}\\ \widehat{\pi}_{10}\end{bmatrix}\right) - g\left(\begin{bmatrix}\pi_{01}\\ \pi_{10}\end{bmatrix}\right)\right)$, define

$$\begin{aligned}
g(a,b) &= \log\left(\frac{a}{b}\right) = \log(a) - \log(b) \\
\Rightarrow \nabla g(a,b) &= \begin{bmatrix}\frac{1}{a} & -\frac{1}{b}\end{bmatrix}
\end{aligned}$$

And so we have,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_C - \beta\right) &= \sqrt{n}\left(g\left(\begin{bmatrix}\widehat{\pi}_{01}\\ \widehat{\pi}_{10}\end{bmatrix}\right) - g\left(\begin{bmatrix}\pi_{01}\\ \pi_{10}\end{bmatrix}\right)\right) \to_d \nabla g(a,b) \cdot \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_c) \\
&= \mathcal{N}\left(0, \nabla g(a,b) \cdot \mathbf{\Sigma}_c \cdot \nabla g(a,b)^T\right) \\
&= \mathcal{N}\left(0, \begin{bmatrix}\frac{1}{a} & -\frac{1}{b}\end{bmatrix} \cdot \begin{bmatrix}a(1-a) & -ab\\ -ab & b(1-b)\end{bmatrix} \cdot \begin{bmatrix}\frac{1}{a} & -\frac{1}{b}\end{bmatrix}^T\right) \\
&= \mathcal{N}\left(0, \begin{bmatrix}1-a+a & -b-(1-b)\end{bmatrix} \cdot \begin{bmatrix}\frac{1}{a}\\ -\frac{1}{b}\end{bmatrix}\right) = \mathcal{N}\left(0, \begin{bmatrix}1 & -1\end{bmatrix} \cdot \begin{bmatrix}\frac{1}{a}\\ -\frac{1}{b}\end{bmatrix}\right) \\
&= \mathcal{N}\left(0, \frac{1}{a} + \frac{1}{b}\right) \\
&= \mathcal{N}\left(0, \frac{1}{\pi_{01}} + \frac{1}{\pi_{10}}\right)
\end{aligned}$$

(d) Next, consider the unconditional maximum likelihood estimation of $(\alpha_i, i = 1, \ldots, n, \beta)$ under the subject-specific model in (c). Show that the unconditional maximum likelihood estimator for β, denoted by $\widehat{\beta}_{MLE}$, is inconsistent.

Solution

Let $\alpha = (\alpha_1, \ldots, \alpha_n)$. The joint probability and log likelihood are

$$
\begin{aligned}
P(\boldsymbol{y}_1, \boldsymbol{y}_2 | \alpha, \beta) &= \prod_{i=1}^{n} \frac{\exp(\alpha_i(y_{i1} + y_{i2}) + \beta y_{i2})}{[1 + \exp(\alpha_i)][1 + \exp(\alpha_i + \beta)]} \\
\Leftrightarrow l_n(\alpha, \beta) &\equiv \log(P(\boldsymbol{y}_1, \boldsymbol{y}_2 | \alpha, \beta)) \\
&= \sum_{i=1}^{n} [\alpha_i(y_{i1} + y_{i2}) + \beta y_{i2} - \log(1 + \exp(\alpha_i)) - \log(1 + \exp(\alpha_i + \beta))]
\end{aligned}
$$

Maximizing the log likelihood,

$$
\begin{aligned}
\frac{\partial l_n(\alpha, \beta)}{\partial \alpha_i} &= y_{i1} + y_{i2} - \frac{\exp(\alpha_i)}{1 + \exp(\alpha_i)} - \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} = 0 \\
\Leftrightarrow y_{i1} + y_{i2} &= \frac{\exp(\alpha_i)}{1 + \exp(\alpha_i)} + \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)}
\end{aligned}
$$

Notice that if

- $y_{i1} + y_{i2} = 0 \Rightarrow$ Both terms on the RHS are nonnegative. So we need them both to individually equal 0. This is possible if $\widehat{\alpha}_i = -\infty$ for $i = 1, \ldots, n$.
- $y_{i1} + y_{i2} = 2 \Rightarrow$ Again both terms on the RHS are nonnegative. So we need them both to individually equal 1. This is possible if $\widehat{\alpha}_i = \infty$ for $i = 1, \ldots, n$.
- $y_{i1} + y_{i2} = 1$ leads to

$$
\begin{aligned}
1 &= \frac{\exp(\alpha_i)}{1 + \exp(\alpha_i)} + \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \\
\Leftrightarrow \frac{1}{1 + \exp(\alpha_i)} &= \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \\
\Leftrightarrow \frac{1}{1 + \exp(\alpha_i)} &= \frac{\exp(\alpha_i)\exp(\beta)}{1 + \exp(\alpha_i)\exp(\beta)} \\
\Leftrightarrow 1 + \exp(\alpha_i)\exp(\beta) &= (1 + \exp(\alpha_i))\exp(\alpha_i)\exp(\beta) \\
\Leftrightarrow 1 &= \exp(2\alpha_i)\exp(\beta) \\
\Rightarrow \widehat{\beta} &= -2\widehat{\alpha}_i
\end{aligned}
$$

Continuing we maximizing the log likelihood with respect to β,

$$
\begin{aligned}
\frac{\partial l_n(\alpha, \beta)}{\partial \beta} &= \sum_{i=1}^{n} \left[y_{i2} - \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \right] = 0 \\
\Leftrightarrow \sum_{i=1}^{n} y_{i2} &= \sum_{i=1}^{n} \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)}
\end{aligned}
$$

At this point, expand both the LHS and RHS. For the LHS,

$$
\begin{aligned}
\sum_{i=1}^{n} y_{i2} &= \sum_{i=1}^{n} y_{i2} \mathbf{1}\{s_i = 0\} + \sum_{i=1}^{n} y_{i2} \mathbf{1}\{s_i = 2\} + \sum_{i=1}^{n} y_{i2} \mathbf{1}\{s_i = 1\} \\
&= 0 + n_{11} + n_{01} \\
&= n_{\cdot 1}
\end{aligned}
$$

For the RHS,

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} &= \sum_{i=1}^{n} \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \mathbf{1}\{s_i = 0\} + \sum_{i=1}^{n} \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \mathbf{1}\{s_i = 2\} \\
&\quad + \sum_{i=1}^{n} \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \mathbf{1}\{s_i = 1\} \\
&\quad \text{Note: Apply what we found from calculating } \frac{\partial l_n(\alpha, \beta)}{\partial \alpha_i}. \\
&= 0 + n_{11} + \sum_{i=1}^{n} \frac{\exp(-\frac{1}{2}\beta + \beta)}{1 + \exp(-\frac{1}{2}\beta + \beta)} \mathbf{1}\{s_i = 1\} \\
&= n_{11} + (n_{01} + n_{10}) \cdot \frac{\exp(\frac{1}{2}\beta)}{1 + \exp(\frac{1}{2}\beta)}
\end{aligned}
$$

And now to equate the LHS and RHS,

$$\begin{aligned}
n_{11} + n_{01} &= n_{11} + (n_{01} + n_{10}) \cdot \frac{\exp(\frac{1}{2}\beta)}{1 + \exp(\frac{1}{2}\beta)} \\
\Rightarrow \widehat{\beta}_{MLE} &= 2 \log\left(\frac{n_{01}}{n_{10}}\right) \\
&= 2 \log\left(\frac{n_{01}/n}{n_{10}/n}\right) \\
&= 2 \log\left(\frac{\widehat{\pi}_{01}}{\widehat{\pi}_{10}}\right) \\
&\to_p 2 \log\left(\frac{\pi_{01}}{\pi_{10}}\right)
\end{aligned}$$

However in part (b), we found that $\beta = \log\left(\frac{\pi_{+1}}{1 - \pi_{+1}}\right) - \log\left(\frac{\pi_{1+}}{1 - \pi_{1+}}\right) = \log\left(\frac{\pi_{+1}\pi_{0+}}{\pi_{1+}\pi_{+0}}\right)$. So $\widehat{\beta}_{MLE}$ is not consistent.

(e) Show that $(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} \leq (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$. Argue that $\text{var}\left[\sqrt{n}\left(\widehat{\beta}_M\right)\right] \leq \text{var}\left[\sqrt{n}\left(\widehat{\beta}_C\right)\right]$ when Y_{i1} and Y_{i2} are independent for each $i = 1, \ldots, n$, and $\alpha_i = \alpha$ are identical for $i = 1, \ldots, n$.

Solution

To prove the inequality, we'll start with

$$(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1}\, vs.\, (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$$

To save time spent writing and reduce chances of typos, I strongly recommend redefining expressions.

$$\begin{aligned}
\Rightarrow (a(1-a))^{-1} + (b(1-b))^{-1} \quad &vs. \quad (b(1-a))^{-1} + (a(1-b))^{-1} \\
\Leftrightarrow 1 + \frac{a(1-a)}{b(1-b)} \quad &vs. \quad \frac{1-a}{1-b} + \frac{a}{b} \\
\Leftrightarrow b(1-b) + a(1-a) \quad &vs. \quad b(1-a) + a(1-b) \\
\Leftrightarrow b - b^2 + a - a^2 \quad &vs. \quad b - 2ab + a \\
\Leftrightarrow 0 \quad &vs. \quad a^2 - 2ab + b^2 \\
\Leftrightarrow 0 \quad &\leq \quad (a-b)^2
\end{aligned}$$

Hence

$$(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} \leq (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$$

Using results from part (b),

$$\begin{aligned}
\text{var}\left[\sqrt{n}\left(\widehat{\beta}_M\right)\right] \quad &= \quad \text{var}\left[\sqrt{n}\left(\widehat{\beta}_M - \beta\right)\right] \\
&\rightarrow_p \quad (\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} - \frac{2(\pi_{11}\pi_{00} - \pi_{10}\pi_{01})}{\pi_{+1}\pi_{+0}\pi_{1+}\pi_{0+}} \\
& \quad \text{Note: Since } Y_{i1} \perp Y_{i2} \Rightarrow \pi_{11}\pi_{00} - \pi_{10}\pi_{01} = 0 \\
&= \quad (\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1}
\end{aligned}$$

Using results from part (c) along with $\alpha_i = \alpha$ for all i,

$$\begin{aligned}
\text{var}\left[\sqrt{n}\left(\widehat{\beta}_C\right)\right] \quad &= \quad \text{var}\left[\sqrt{n}\left(\widehat{\beta}_C - \beta\right)\right] \\
&\rightarrow_p \quad \frac{1}{\pi_{01}} + \frac{1}{\pi_{10}} \\
& \quad \text{Note: Since } Y_{i1} \perp Y_{i2} \Rightarrow \frac{1}{\pi_{ij}} = \frac{1}{\pi_{i+}\pi_{+j}} \\
&= \quad \frac{1}{\pi_{0+}\pi_{+1}} + \frac{1}{\pi_{1+}\pi_{+0}} \\
&= \quad (\pi_{0+}\pi_{+1})^{-1} + (\pi_{1+}\pi_{+0})^{-1}
\end{aligned}$$

And so we see that asymptotically, $\text{var}\left[\sqrt{n}\left(\widehat{\beta}_M\right)\right] \leq \text{var}\left[\sqrt{n}\left(\widehat{\beta}_C\right)\right]$ when Y_{i1} and Y_{i2} are independent for each $i = 1, \ldots, n$, and $\alpha_i = \alpha$ are identical for $i = 1, \ldots, n$.

1.2.2 Question 2

2. We consider the model $Y_{ij} = \mu_i + (x_{ij} - \bar{x}_i)\gamma_i + \epsilon_{ij}, i = 1, 2, j = 1, \ldots, n_i$, where $n_i > 0, \bar{x}_i \equiv (1/n_i)\sum_j x_{ij}$, and μ_i and γ_i are scalar parameters. Suppose that x_{ij} are known scalars which are not all equal for each $i = 1, 2$. Further, suppose that $\{\epsilon_{ij}, i = 1, 2, j = 1, \ldots, n_i\}$ are assumed to be independent and identically distributed such that $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is a scalar parameter.

   (a) Let $\boldsymbol{\beta} = (\mu_1, \gamma_1, \mu_2, \gamma_2)^T$. We wish to write this model as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2})^T$, $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{1n_1}, \epsilon_{21}, \ldots, \epsilon_{2n_2})^T$, and $\boldsymbol{X}$ is an appropriately defined matrix.

      (i) Specify $\boldsymbol{X}$ and the distribution of $\boldsymbol{\epsilon}$.

      Solution

      We can write out the model for the i, jth's model in vector form as

$$\begin{aligned} Y_{ij} &= \mu_i + (x_{ij} - \bar{x}_i)\gamma_i + \epsilon_{ij} \\ &= \begin{bmatrix} 1 & x_{ij} - \bar{x}_i \end{bmatrix} \begin{bmatrix} \mu_i \\ \gamma_i \end{bmatrix} + \epsilon_{ij} \end{aligned}$$

      And so for the ith's group model, we have

$$\begin{aligned} \boldsymbol{Y}_i \equiv \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{bmatrix} &= \begin{bmatrix} 1 & x_{i1} - \bar{x}_i \\ \vdots & \vdots \\ 1 & x_{in_i} - \bar{x}_i \end{bmatrix} \begin{bmatrix} \mu_i \\ \gamma_i \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix} \\ &\equiv \begin{bmatrix} \boldsymbol{J}_{n_i} & \boldsymbol{X}_i - \bar{\boldsymbol{X}}_i \end{bmatrix} \begin{bmatrix} \mu_i \\ \gamma_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_i \end{bmatrix} \end{aligned}$$

      Finally we can express the model.

$$\boldsymbol{Y} \equiv \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{J}_{n_1} & \boldsymbol{X}_1 - \bar{\boldsymbol{X}}_1 & \boldsymbol{0}_{n_1} & \boldsymbol{0}_{n_1} \\ \boldsymbol{0}_{n_2} & \boldsymbol{0}_{n_2} & \boldsymbol{J}_{n_2} & \boldsymbol{X}_2 - \bar{\boldsymbol{X}}_2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \gamma_1 \\ \mu_2 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{bmatrix}$$

      Hence, we have our definition and distribution.

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} - \bar{x}_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n_1} - \bar{x}_1 & 0 & 0 \\ 0 & 0 & 1 & x_{21} - \bar{x}_2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{2n_2} - \bar{x}_2 \end{bmatrix} \text{ and } \boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{0}_{n_1+n_2}, \sigma^2 \boldsymbol{I}_{(n_1+n_2)\times(n_1+n_2)}\right)$$

(ii) Give the estimate (call it $\widehat{\boldsymbol{\beta}}$) for $\boldsymbol{\beta}$ which has minimum variance among the class of linear (in $\boldsymbol{Y}$) unbiased estimates.

Solution

Since we have HILE Gauss, by **Markov's Theorem**, we know that the BLUE estimate of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is $\boldsymbol{\rho}^T\boldsymbol{M}\boldsymbol{Y}$. In this case

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

where $\boldsymbol{X}$ is full rank. Let's define $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{X}_{1,c} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}_2 & \boldsymbol{X}_{2,c} \end{bmatrix}$ where $\boldsymbol{X}_{i,c} = \boldsymbol{X}_i - \bar{\boldsymbol{X}}_i$ and $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix}$. Then

$$\begin{aligned}
\boldsymbol{X}^T\boldsymbol{X} &= \begin{bmatrix} \boldsymbol{J}_1^T & \boldsymbol{0}^T \\ \boldsymbol{X}_{1,c}^T & \boldsymbol{0}^T \\ \boldsymbol{0}^T & \boldsymbol{J}_2^T \\ \boldsymbol{0}^T & \boldsymbol{X}_{2,c}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{X}_{1,c} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}_2 & \boldsymbol{X}_{2,c} \end{bmatrix} = \begin{bmatrix} n_1 & 0 & 0 & 0 \\ 0 & \sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2 & 0 & 0 \\ 0 & 0 & n_2 & 0 \\ 0 & 0 & 0 & \sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2 \end{bmatrix} \\
\Rightarrow \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} &= \begin{bmatrix} \frac{1}{n_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2} \end{bmatrix} \\
\boldsymbol{X}^T\boldsymbol{Y} &= \begin{bmatrix} \boldsymbol{J}_1^T & \boldsymbol{0}^T \\ \boldsymbol{X}_{1,c}^T & \boldsymbol{0}^T \\ \boldsymbol{0}^T & \boldsymbol{J}_2^T \\ \boldsymbol{0}^T & \boldsymbol{X}_{2,c}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n_1} Y_{1j} \\ \sum_{j=1}^{n_1} Y_{1j}(x_{1j}-\bar{x}_1) \\ \sum_{j=1}^{n_2} Y_{2j} \\ \sum_{j=1}^{n_2} Y_{2j}(x_{2j}-\bar{x}_2) \end{bmatrix} \\
\Rightarrow \widehat{\boldsymbol{\beta}} &= \begin{bmatrix} \bar{Y}_1 \\ \frac{\sum_{j=1}^{n_1} Y_{1j}(x_{1j}-\bar{x}_1)}{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2} \\ \bar{Y}_2 \\ \frac{\sum_{j=1}^{n_2} Y_{2j}(x_{2j}-\bar{x}_2)}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2} \end{bmatrix} = \begin{bmatrix} \widehat{\mu}_1 \\ \widehat{\gamma}_1 \\ \widehat{\mu}_2 \\ \widehat{\gamma}_2 \end{bmatrix}
\end{aligned}$$

(b) (i) Specify a column vector $\boldsymbol{a}$ such that $\boldsymbol{a}^T\boldsymbol{\beta} = (\gamma_1 - \gamma_2)$. Is $\boldsymbol{a}^T\boldsymbol{\beta}$ estimable? Explain why or why not.

Solution

One approach: Notice that $\gamma_1 - \gamma_2 = \boldsymbol{a}^T \begin{bmatrix} \mu_1 \\ \gamma_1 \\ \mu_2 \\ \gamma_2 \end{bmatrix} \Rightarrow \boldsymbol{a}^T = \begin{bmatrix} 0 & 1 & 0 - 1 \end{bmatrix}$. $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable if there exists a vector $\boldsymbol{\rho}$ such that $\boldsymbol{a}^T = \boldsymbol{\rho}^T\boldsymbol{X}$.

And so we see that

$$\begin{aligned} \boldsymbol{a}^T &= \boldsymbol{\rho}^T\boldsymbol{X} \\ \Leftrightarrow \begin{bmatrix} 0 & 1 & 0 - 1 \end{bmatrix} &= \begin{bmatrix} \rho_{11} & \cdots & \rho_{1n_1} & \rho_{21} & \cdots & \rho_{2n_2} \end{bmatrix} \boldsymbol{X} \\ &\equiv \begin{bmatrix} \boldsymbol{\rho}_1^T & \boldsymbol{\rho}_2^T \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{X}_{1,c} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}_2 & \boldsymbol{X}_{2,c} \end{bmatrix} \end{aligned}$$

We can let $\rho_{1j} = \dfrac{x_{1j} - \bar{x}_1}{\sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2}$ and $\rho_{2j} = -\dfrac{(x_{2j} - \bar{x}_2)}{\sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2}$ to satisfy $\boldsymbol{a}^T = \boldsymbol{\rho}^T\boldsymbol{X}$.

Another approach: Notice that the four columns of $\boldsymbol{X}$ are linearly independent and hence $\boldsymbol{X}$ is full rank, hence any linear combination of the elements of $\boldsymbol{\beta}$ or $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable. This is because centered covariates are orthogonal to the intercept (easy to prove).

(ii) Suppose σ^2 is known. Derive the distribution of $\boldsymbol{a}^T\widehat{\boldsymbol{\beta}}$ and give a $(1-\alpha)$-level confidence interval for $\gamma_1 - \gamma_2$.

Solution

To calculate the distribution of $\boldsymbol{a}^T\widehat{\boldsymbol{\beta}}$, we have

$$\begin{aligned} \boldsymbol{a}^T\widehat{\boldsymbol{\beta}} &= \boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y} \\ &\quad \text{Note: } \boldsymbol{Y}|\boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_{(n_1+n_2)\times(n_1+n_2)}\right) \\ &\sim \mathcal{N}\left(\boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T \cdot \sigma^2\boldsymbol{I} \cdot \left(\boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right)^T\right) \\ &= \mathcal{N}\left(\boldsymbol{a}^T\boldsymbol{\beta}, \sigma^2\boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{a}\right) \\ &= \mathcal{N}\left(\gamma_1 - \gamma_2, \sigma^2\left(\frac{1}{\sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2} + \frac{1}{\sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2}\right)\right) \end{aligned}$$

Using the distribution of $\boldsymbol{a}^T\widehat{\boldsymbol{\beta}} = \widehat{\gamma}_1 - \widehat{\gamma}_2$, we can derive the $(1-\alpha)$ confidence interval for $\gamma_1 - \gamma_2$.

$$\widehat{\gamma}_1 - \widehat{\gamma}_2 \pm z_{1-\frac{\alpha}{2}}\sqrt{\sigma^2\left(\frac{1}{\sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2} + \frac{1}{\sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2}\right)}$$

(iii) Suppose σ^2 is unknown. Give a statistic to test $H_0 : \gamma_1 = \gamma_2$ and indicate its distribution under H_0.

Solution

We can re-express the hypothesis $H_0 : \gamma_1 = \gamma_2$ as $H_0 : \boldsymbol{a}^T\boldsymbol{\beta} = 0$ using what we know from previous questions. The corresponding statistic is

$$\begin{aligned} F^* &\equiv \frac{\|\boldsymbol{M}_{MP}\boldsymbol{Y}\|/rank(\boldsymbol{a})}{\|(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Y}\|/rank(\boldsymbol{I}-\boldsymbol{M})} \\ &= \frac{\boldsymbol{Y}^T\boldsymbol{M}_{MP}\boldsymbol{Y}/1}{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Y}/(n_1+n_2-4)} \end{aligned}$$

To simplify the numerator,

$$\begin{aligned} \boldsymbol{Y}^T\boldsymbol{M}_{MP}\boldsymbol{Y} &= \boldsymbol{Y}^T\boldsymbol{M}\boldsymbol{\rho}(\boldsymbol{\rho}^T\boldsymbol{M}\boldsymbol{\rho})^{-1}\boldsymbol{\rho}^T\boldsymbol{M}\boldsymbol{Y} \\ &= (\boldsymbol{a}^T\widehat{\boldsymbol{\beta}})^T(\boldsymbol{\rho}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\rho})^{-1}(\boldsymbol{a}^T\widehat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{a}^T\widehat{\boldsymbol{\beta}})^2(\boldsymbol{a}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a})^{-1} \\ &= \frac{\left(\dfrac{\sum_{j=1}^{n_1} Y_{1j}(x_{1j}-\bar{x}_1)}{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2} - \dfrac{\sum_{j=1}^{n_2} Y_{2j}(x_{2j}-\bar{x}_2)}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2}\right)^2}{\left(\dfrac{1}{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2} + \dfrac{1}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2}\right)} \end{aligned}$$

To simplify the denominator, I'll denote it as $\widehat{\sigma}^2$ or the MSE.
So the test statistic and distribution under the null is

$$F^* = \frac{\left(\dfrac{\sum_{j=1}^{n_1} Y_{1j}(x_{1j}-\bar{x}_1)}{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2} - \dfrac{\sum_{j=1}^{n_2} Y_{2j}(x_{2j}-\bar{x}_2)}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2}\right)^2}{\widehat{\sigma}^2\left(\dfrac{1}{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2} + \dfrac{1}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2}\right)} \sim F(1, n_1+n_2-4)$$

We'd reject H_0 if $F^* > F(1, n_1+n_2-4, 1-\alpha)$.

(c) (i) Do the results in (b) (ii) and (b)(iii) change if we fit the model under the restriction that $\mu_1 = \mu_2 = 0$? Justify your answer.

Solution

No the confidence interval will remain the same hence (b) (ii) will remain unchanged.

The results of (b) (iii) will change, more specifically the degrees of freedom will be reduced to $n_1 + n_2 - 2$. The reason why the restriction doesn't change much is because the four original columns of $\boldsymbol{X}$ are orthogonal to each other.

(ii) Derive the least squares estimate of $\mu_i, i = 1, 2$ under the constraint that $\boldsymbol{a}^T\boldsymbol{\beta} = 0$ [for $\boldsymbol{a}$ defined in (b)(i)]. Write the estimate in the scalar form explicitly, as opposed to the matrix representation.

Solution

Since we have a linear constraint when minimizing the squared error, we can use Lagrange multipliers. Begin with

$$\begin{aligned}
f(\boldsymbol{\beta}, \lambda) &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda \cdot \boldsymbol{a}^T\boldsymbol{\beta} \\
\Rightarrow \frac{\partial f}{\partial \boldsymbol{\beta}} &= -2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + \lambda\boldsymbol{a} = 0 \\
\Leftrightarrow \boldsymbol{\beta} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - \frac{1}{2}\lambda(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a} \\
\Rightarrow \frac{\partial f}{\partial \lambda} &= \boldsymbol{a}^T\boldsymbol{\beta} = 0 \\
\Leftrightarrow 0 &= \boldsymbol{a}^T\left((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - \frac{1}{2}\lambda(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a}\right) \\
\Rightarrow \lambda &= \frac{2\boldsymbol{a}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}}{\boldsymbol{a}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a}}
\end{aligned}$$

Plugging this last expression back into the expression for $\boldsymbol{\beta}$, we have

$$\begin{aligned}
\tilde{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - \frac{1}{2}\left(\frac{2\boldsymbol{a}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}}{\boldsymbol{a}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a}}\right)(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a} \\
&= \widehat{\boldsymbol{\beta}} - \frac{\boldsymbol{a}^T\widehat{\boldsymbol{\beta}}}{\boldsymbol{a}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{a}}\begin{bmatrix} 0 \\ \frac{1}{\sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2} \\ 0 \\ -\frac{1}{\sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2} \end{bmatrix}
\end{aligned}$$

Looking at the vector we notice that the linear constraint doesn't effect the estimates of μ_1 and μ_2.

(d) Consider the linear model in the matrix form as in (a). Let Y_l denote the lth row of $\boldsymbol{Y}$ and $\boldsymbol{x}_l^T$ denote the lth row of $\boldsymbol{X}$, $l = 1, \ldots, n_1, n_1 + 1, \ldots, n_1 + n_2$. We are interested in deriving the F-test for the hypothesis that the observation in the kth row, Y_k, is not an outlier. Suppose we leave out the kth row from $\boldsymbol{Y}$ and $\boldsymbol{X}$ and we compute the least squares estimate, denoted by $\widehat{\boldsymbol{\beta}}_{(k)}$. Let $\boldsymbol{X}_{(k)}$ denote the design matrix obtained by deleting the kth row from $\boldsymbol{X}$ and let $\boldsymbol{Y}_{(k)}$ denote the vector obtained by deleting the kth row from $\boldsymbol{Y}$. Assume $\boldsymbol{X}_{(k)}$ is full rank. Define $D_k \equiv Y_k - \boldsymbol{x}_k^T\widehat{\boldsymbol{\beta}}_{(k)}$.

(i) Give the matrix formulation for $\widehat{\boldsymbol{\beta}}_{(k)}$.

Solution

We have that $\widehat{\boldsymbol{\beta}}_{(k)} = \left(\boldsymbol{X}_{(k)}^T\boldsymbol{X}_{(k)}\right)^{-1}\boldsymbol{X}_{(k)}^T\boldsymbol{Y}_{(k)}$ as the least squares estimate.

$$\boldsymbol{X}_{(k)}^T\boldsymbol{X}_{(k)} = \boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{x}_k\boldsymbol{x}_k^T$$

$$\boldsymbol{X}_{(k)}^T\boldsymbol{Y}_{(k)} = \boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{x}_k y_k$$

(ii) Derive the distribution of D_k.

Solution

We have that $D_k = Y_k - \boldsymbol{x}_k^T \widehat{\boldsymbol{\beta}}_{(k)}$. Hence D_k is normally distributed, so we just need to find the expectation and variance. For the expectation,

$$\begin{aligned} E\left[D_k\right] &= E\left[Y_k - \boldsymbol{x}_k^T \widehat{\boldsymbol{\beta}}_{(k)}\right] \\ &= \boldsymbol{x}_k^T \boldsymbol{\beta} - \boldsymbol{x}_k^T \boldsymbol{\beta}_{(k)} \\ &= \boldsymbol{x}_k^T \left(\boldsymbol{\beta} - \boldsymbol{\beta}_{(k)}\right) \end{aligned}$$

For the variance,

$$\begin{aligned} V\left[D_k\right] &= V\left[Y_k - \boldsymbol{x}_k^T \widehat{\boldsymbol{\beta}}_{(k)}\right] \\ &= V\left[Y_k\right] + \boldsymbol{x}_k^T V\left[\widehat{\boldsymbol{\beta}}_{(k)}\right] \boldsymbol{x}_k \\ &= \sigma^2 + \boldsymbol{x}_k^T V\left[\left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{X}_{(k)}^T \boldsymbol{Y}_{(k)}\right] \boldsymbol{x}_k \\ &= \sigma^2 + \boldsymbol{x}_k^T \left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{X}_{(k)}^T V\left[\boldsymbol{Y}_{(k)}\right] \boldsymbol{X}_{(k)} \left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{x}_k \\ &= \sigma^2 + \boldsymbol{x}_k^T \left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{X}_{(k)}^T \sigma^2 \boldsymbol{I}_{n-1} \boldsymbol{X}_{(k)} \left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{x}_k \\ &= \sigma^2 \left(1 + \boldsymbol{x}_k^T \left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{x}_k\right) \end{aligned}$$

And so we have that

$$D_k \sim \mathcal{N}\left(\boldsymbol{x}_k^T \left(\boldsymbol{\beta} - \boldsymbol{\beta}_{(k)}\right), \sigma^2 \left(1 + \boldsymbol{x}_k^T \left(\boldsymbol{X}_{(k)}^T \boldsymbol{X}_{(k)}\right)^{-1} \boldsymbol{x}_k\right)\right)$$

(iii) Based on the distribution of D_k in (ii), provide an F-test for the hypothesis that Y_k is not an outlier (i.e. for the hypothesis $H_0 : E(Y_k) = \boldsymbol{x}_k^T \boldsymbol{\beta}$).

Solution

Notice that

- under H_0: Y_k is not an outlier $\Leftrightarrow E[Y_k] = \boldsymbol{x}_k^T \boldsymbol{\beta}$.
- under H_1: Y_k is an outlier $\Leftrightarrow E[Y_k] = \boldsymbol{x}_k^T \boldsymbol{\beta}_k$ and $E\left[\boldsymbol{Y}_{(k)}\right] = \boldsymbol{X}_{(k)} \boldsymbol{\beta}_{(k)}$

So the full model is

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{Y}_{(k)} \\ Y_k \end{bmatrix} &= \begin{bmatrix} \boldsymbol{X}_{(k)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{x}_k^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{(k)} \\ \boldsymbol{\beta}_k \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{(k)} \\ \epsilon_k \end{bmatrix} \\
&= \boldsymbol{X}_* \boldsymbol{\beta}_* + \boldsymbol{\epsilon}_*
\end{aligned}$$

And so our hypotheses can be restated as

- $H_0 : \boldsymbol{\beta}_{(k)} = \boldsymbol{\beta}_k$
- $H_1 : \boldsymbol{\beta}_{(k)} \neq \boldsymbol{\beta}_k$

Let's define $\boldsymbol{M}_* = \boldsymbol{X}_* \left(\boldsymbol{X}_*^T \boldsymbol{X}_*\right)^{-1} \boldsymbol{X}_*^T$ and $\boldsymbol{M} = \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T$ where $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_k^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix}$. The F-test is of the form

$$\begin{aligned}
F^* &= \frac{\|(\boldsymbol{M}_* - \boldsymbol{M})\boldsymbol{Y}\| / r(\boldsymbol{M}_* - \boldsymbol{M})}{\|(\boldsymbol{I} - \boldsymbol{M}_*)\boldsymbol{Y}\| / r(\boldsymbol{I} - \boldsymbol{M}_*)} \\
&\sim^{H_0} F\left(r(\boldsymbol{M}_* - \boldsymbol{M}), r(\boldsymbol{I} - \boldsymbol{M}_*)\right) \\
&= F(1, n-5)
\end{aligned}$$

We reject H_0 when $F^* > F(1, n-5, 1-\alpha)$. (I think. Might need to double check this.)

1.2.3 Question 3

3. To study the effect of a risk factor X on a count variable Y, data are collected from two clinical centers. For center $k = 1, 2$, the data, (Y_{ik}, X_{ik}), $i = 1, \dots, n$, are i.i.d from the distribution: $X_{ik} \sim \mathcal{N}(0, \sigma_k^2)$ and given X_{ik}, Y_{ik} follows a Poisson distribution having p.m.f. $\lambda_{ik}^y \exp(-\lambda_{ik})/y!, y = 0, 1, \dots,$ with $\lambda_{ik} = \exp(\alpha_k + \beta X_{ik})$, where $\sigma_k^2 > 0$. Both (σ_1^2, σ_2^2) and $(\alpha_1, \alpha_2, \beta)$ are unknown parameters.

   (a) Let $(\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\beta})$ be the maximum likelihood estimator using all of the data from the two centers. Find the asymptotic distribution of $\sqrt{2n}(\widehat{\beta} - \beta)$ in terms of the true parameters. Hint:

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} = \begin{pmatrix} (A - BC^{-1}B^T)^{-1} & -A^{-1}B(C - B^TA^{-1}B)^{-1} \\ -C^{-1}B^T(A - BC^{-1}B^T)^{-1} & (C - B^TA^{-1}B)^{-1} \end{pmatrix}$$

Solution

Let us define $\xi = (\sigma_1^2, \sigma_2^2, \alpha_1, \alpha_2, \beta)$. The joint likelihood and log-likelihood are

$$\begin{aligned}
P(\boldsymbol{Y}, \boldsymbol{X}|\xi) &= \prod_{k=1}^{2}\prod_{i=1}^{n} P(Y_{ik}, X_{ik}|\xi) \\
&= \prod_{k=1}^{2}\prod_{i=1}^{n} P\left(X_{ik}|\sigma_k^2\right) P\left(Y_{ik}|X_{ik}; \alpha_k, \beta\right) \\
&= \prod_{k=1}^{2}\prod_{i=1}^{n} P\left(X_{ik}|\sigma_k^2\right) \times \prod_{k=1}^{2}\prod_{i=1}^{n} P\left(Y_{ik}|X_{ik}; \alpha_k, \beta\right) \\
\Rightarrow l_{2n}(\xi) \equiv \log\left(P(\boldsymbol{Y}, \boldsymbol{X}|\xi)\right) &= \sum_k\sum_i \log\left(P\left(X_{ik}|\sigma_k^2\right)\right) + \sum_k\sum_i \log\left(P\left(Y_{ik}|X_{ik}; \alpha_k, \beta\right)\right)
\end{aligned}$$

Now obtain the 1st and 2nd derivatives to calculate Fisher's Information, $I_{2n}(\xi)$. Starting with σ_k^2, we have

$$\begin{aligned}
\frac{\partial l_{2n}(\xi)}{\partial \sigma_k^2} &= \sum_i \partial_{\sigma_k^2}\left[\log\left(P\left(X_{ik}|\sigma_k^2\right)\right)\right] \\
&\quad \text{Note: } P\left(X_{ik}|\sigma_k^2\right) = (2\pi\sigma_k^2)^{-1/2}\exp\left\{-\frac{x_{ik}^2}{2\sigma_k^2}\right\} = \exp\left\{-\frac{1}{2}\log(2\pi\sigma_k^2) - \frac{x_{ik}^2}{2\sigma_k^2}\right\} \\
&= \sum_i \partial_{\sigma_k^2}\left[-\frac{1}{2}\log(2\pi\sigma_k^2) - \frac{x_{ik}^2}{2\sigma_k^2}\right] \\
&= -\frac{1}{2}\sum_i\left\{\frac{1}{\sigma_k^2} - \frac{x_{ik}^2}{\sigma_k^4}\right\} \\
&= \frac{\sum_i x_{ik}^2}{2\sigma_k^4} - \frac{n}{2\sigma_k^2} \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial(\sigma_k^2)^2} &= -\frac{\sum_i x_{ik}^2}{\sigma_k^6} + \frac{n}{2\sigma_k^4} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial(\sigma_k^2)^2}\right] &= -\frac{n}{2\sigma_k^4} + \frac{\sum_i E\left[\left(\frac{x_{ik}}{\sigma_k}\right)^2\right]}{\sigma_k^4} \\
&= -\frac{n}{2\sigma_k^4} + \frac{n}{\sigma_k^4} \\
&= \frac{n}{2\sigma_k^4} \text{ for } k = 1, 2
\end{aligned}$$

Notice that $\dfrac{\partial l_{2n}(\xi)}{\partial \sigma_k^2}$ is not a function of $\sigma_{k'}$, β, α_k, or $\alpha_{k'}$ so

$$E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\sigma_k\partial\sigma_{k'}}\right] = E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\sigma_k\partial\alpha_k}\right] = E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\sigma_k\partial\alpha_{k'}}\right] = E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\sigma_k\partial\beta}\right] = 0$$

For α_k, we have

$$\begin{aligned}
\frac{\partial l_{2n}(\xi)}{\partial \alpha_k} &= \sum_i \partial_{\alpha_k}\left[\log\left(P\left(Y_{ik}|X_{ik};\alpha_k,\beta\right)\right)\right] \\
&\quad \text{Note: } P\left(Y_{ik}|X_{ik};\alpha_k,\beta\right) = \frac{\exp(-\lambda_{ik})\lambda_{ik}^{y_{ik}}}{y_{ik}!} \propto \exp\left\{-\lambda_{ik} + y_{ik}\log(\lambda_{ik})\right\} \\
&= \sum_i \partial_{\alpha_k}\left[-\lambda_{ik} + y_{ik}\log(\lambda_{ik})\right] \\
&\quad \text{Note: } \lambda_{ik} = \exp(\alpha_k + \beta x_{ik}) \\
&= \sum_i \partial_{\alpha_k}\left[-\exp(\alpha_k + \beta x_{ik}) + y_{ik}(\alpha_k + \beta x_{ik})\right] \\
&= \sum_i \left\{-\exp(\alpha_k + \beta x_{ik}) + y_{ik}\right\} \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k^2} &= -\sum_i \exp(\alpha_k + \beta x_{ik}) \\
&= -e^{\alpha_k}\sum_i e^{\beta x_{ik}} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k^2}\right] &= e^{\alpha_k}\sum_i E\left[e^{\beta x_{ik}}\right] \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k \partial \beta} &= \sum_i \left\{-\exp(\alpha_k + \beta x_{ik}) x_{ik}\right\} \\
&= -e^{\alpha_k}\sum_i \left\{x_{ik}\cdot e^{\beta x_{ik}}\right\} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k \partial \beta}\right] &= e^{\alpha_k}\sum_i E\left[x_{ik}\cdot e^{\beta x_{ik}}\right]
\end{aligned}$$

Notice that $\dfrac{\partial l_{2n}(\xi)}{\partial \alpha_k}$ is not a function of $\alpha_{k'}$ so

$$E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k \partial \alpha_{k'}}\right] = 0$$

Finally for β, we have

$$\begin{aligned}
\frac{\partial l_{2n}(\xi)}{\partial \beta} &= \sum_k\sum_i \left\{-x_{ik}\exp(\alpha_k + \beta x_{ik}) + y_{ik}x_{ik}\right\} \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial \beta^2} &= \sum_k\sum_i - x_{ik}^2\exp(\alpha_k + \beta x_{ik}) \\
&= -\sum_k e^{\alpha_k}\sum_i x_{ik}^2 e^{\beta x_{ik}} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \beta^2}\right] &= \sum_k e^{\alpha_k}\sum_i E\left[x_{ik}^2 e^{\beta x_{ik}}\right]
\end{aligned}$$

Regarding the inner expectations, we can expand them with moment generating function properties. Recall that if $X \sim \mathcal{N}(0,\sigma^2)$, then

$$\begin{aligned}
E\left[e^{tX}\right] &= \exp\left\{\frac{1}{2}\sigma^2 t^2\right\} \\
\Rightarrow E\left[Xe^{tX}\right] &= E\left[\partial_t\left(e^{tX}\right)\right] = \partial_t\left(E\left[e^{tX}\right]\right) = \partial_t\left(\exp\left\{\frac{1}{2}\sigma^2 t^2\right\}\right) = \exp\left\{\frac{1}{2}\sigma^2 t^2\right\}\cdot\sigma^2 t \\
\Rightarrow E\left[X^2 e^{tX}\right] &= \partial_t^2\left(E\left[e^{tX}\right]\right) = \partial_t\left(\partial_t\left(E\left[e^{tX}\right]\right)\right) = \partial_t\left(\exp\left\{\frac{1}{2}\sigma^2 t^2\right\}\cdot\sigma^2 t\right) \\
&= \exp\left\{\frac{1}{2}\sigma^2 t^2\right\}\cdot\sigma^2 + \sigma^2 t\cdot\exp\left\{\frac{1}{2}\sigma^2 t^2\right\}\cdot\sigma^2 t \\
&= \sigma^2\exp\left\{\frac{1}{2}\sigma^2 t^2\right\}\left(1 + \sigma^2 t^2\right)
\end{aligned}$$

Therefore

$$
\begin{aligned}
E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k^2}\right] &= e^{\alpha_k}\sum_i E\left[e^{\beta x_{ik}}\right] = e^{\alpha_k}\sum_i \exp\left\{\frac{1}{2}\sigma_k^2\beta^2\right\} \\
&= n\cdot\exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} \\
E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \alpha_k \partial\beta}\right] &= e^{\alpha_k}\sum_i E\left[x_{ik}\cdot e^{\beta x_{ik}}\right] = e^{\alpha_k}\sum_i \sigma_k^2\beta\cdot\exp\left\{\frac{1}{2}\sigma_k^2\beta^2\right\} \\
&= n\cdot\sigma_k^2\beta\exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} \\
E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial \beta^2}\right] &= \sum_k e^{\alpha_k}\sum_i E\left[x_{ik}^2 e^{\beta x_{ik}}\right] = \sum_k e^{\alpha_k}\sum_i \sigma_k^2\exp\left\{\frac{1}{2}\sigma_k^2\beta^2\right\}(1+\sigma_k^2\beta^2) \\
&= n\cdot\sum_k \sigma_k^2\exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\}(1+\sigma_k^2\beta^2)
\end{aligned}
$$

The Fisher Information Matrix is

$$
\begin{aligned}
I_{2n}(\xi) &= \begin{bmatrix}
\frac{n}{2\sigma_1^4} & 0 & 0 & 0 & 0 \\
0 & \frac{n}{2\sigma_2^4} & 0 & 0 & 0 \\
0 & 0 & n\cdot\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} & 0 & n\cdot\sigma_1^2\beta\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} \\
0 & 0 & 0 & n\cdot\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} & n\cdot\sigma_2^2\beta\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} \\
0 & 0 & n\cdot\sigma_1^2\beta\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} & n\cdot\sigma_2^2\beta\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} & n\cdot\sum_k \sigma_k^2\exp\left\{\alpha_k+\frac{1}{2}\sigma_k^2\beta^2\right\}(1+\sigma_k^2\beta^2)
\end{bmatrix} \\
&= 2n\begin{bmatrix}
\frac{1}{4\sigma_1^4} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{4\sigma_2^4} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2}\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} & 0 & \frac{1}{2}\sigma_1^2\beta\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} \\
0 & 0 & 0 & \frac{1}{2}\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} & \frac{1}{2}\sigma_2^2\beta\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} \\
0 & 0 & \frac{1}{2}\sigma_1^2\beta\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} & \frac{1}{2}\sigma_2^2\beta\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} & \frac{1}{2}\sum_k \sigma_k^2\exp\left\{\alpha_k+\frac{1}{2}\sigma_k^2\beta^2\right\}(1+\sigma_k^2\beta^2)
\end{bmatrix} \\
&= 2n\cdot I(\xi)
\end{aligned}
$$

By the Weak Law of Large Numbers, $\frac{1}{2n}I_{2n}(\xi)\to_p I(\xi)$. From MLE theory with the regularity conditions satisfied, we know that the MLE is asymptotically efficient and that

$$\sqrt{2n}\left(\widehat{\xi}_{MLE}-\xi\right)\to_d \mathcal{N}(\mathbf{0}, I^{-1}(\xi))$$

and by the Delta Method or continuous mapping theorem

$$
\begin{aligned}
\sqrt{2n}\left(\widehat{\beta}-\beta\right) &= \sqrt{2n}\left(g(\widehat{\xi})-g(\xi)\right) \\
&\to_d \nabla g(\xi)\cdot\mathcal{N}(\mathbf{0}, I^{-1}(\xi)) \\
&= \mathcal{N}(0, \nabla g(\xi)\cdot I^{-1}(\xi)\cdot(\nabla g(\xi))^T) \\
&= \mathcal{N}(0, I^{-1}(\beta))
\end{aligned}
$$

where $g(\boldsymbol{x})=\boldsymbol{a}^T\boldsymbol{x}$, a linear transformation of $\boldsymbol{x}$ with $\boldsymbol{a}=(0,0,0,0,1)^T$. Hence $I^{-1}(\beta)\equiv\boldsymbol{a}^T I^{-1}(\xi)\boldsymbol{a}$ corresponds to the fifth row and fifth column of the inverted matrix $I(\xi)$. If we denote $I(\xi)$ as

$$I(\xi)=\begin{bmatrix} A & B \\ B^T & C\end{bmatrix}$$

where $A=\begin{bmatrix}\frac{1}{4\sigma_1^4} & 0 \\ 0 & \frac{1}{4\sigma_2^4}\end{bmatrix}$, $B=\begin{bmatrix}0&0&0\\0&0&0\end{bmatrix}$, and

$$C=\begin{bmatrix}
\frac{1}{2}\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} & 0 & \frac{1}{2}\sigma_1^2\beta\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} \\
0 & \frac{1}{2}\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} & \frac{1}{2}\sigma_2^2\beta\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} \\
\frac{1}{2}\sigma_1^2\beta\exp\left\{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2\right\} & \frac{1}{2}\sigma_2^2\beta\exp\left\{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2\right\} & \frac{1}{2}\sum_k \sigma_k^2\exp\left\{\alpha_k+\frac{1}{2}\sigma_k^2\beta^2\right\}(1+\sigma_k^2\beta^2)
\end{bmatrix}$$

Using the hint, we'll see that the off-diagonals of $I^{-1}(\xi)$ will equal 0 because B is a matrix of 0s. Also the bottom right matrix notated $\left(C-B^TA^{-1}B\right)^{-1}$ will just equal C^{-1}. So we only need to properly invert our matrix C by re-applying the hint and extract the bottom right element. Let C be partitioned into

$$\begin{bmatrix} A_* & B_* \\ B_*^T & C_*\end{bmatrix}$$

where $A_* = \begin{bmatrix} \frac{1}{2}\exp\left\{\alpha_1 + \frac{1}{2}\sigma_1^2\beta^2\right\} & 0 \\ 0 & \frac{1}{2}\exp\left\{\alpha_2 + \frac{1}{2}\sigma_2^2\beta^2\right\} \end{bmatrix}$, $B_* = \begin{bmatrix} \frac{1}{2}\sigma_1^2\beta\exp\left\{\alpha_1 + \frac{1}{2}\sigma_1^2\beta^2\right\} \\ \frac{1}{2}\sigma_2^2\beta\exp\left\{\alpha_2 + \frac{1}{2}\sigma_2^2\beta^2\right\} \end{bmatrix}$, and

$$C_* = \left[\frac{1}{2}\sum_k \sigma_k^2 \exp\left\{\alpha_k + \tfrac{1}{2}\sigma_k^2\beta^2\right\}\left(1 + \sigma_k^2\beta^2\right)\right].$$

The bottom right element will equal

$$\left(C_* - B_*^T A_*^{-1} B_*\right)^{-1} = I^{-1}(\beta) = \frac{2}{\sigma_1^2 e^{\alpha_1 + \frac{1}{2}\sigma_1^2\beta^2} + \sigma_2^2 e^{\alpha_2 + \frac{1}{2}\sigma_2^2\beta^2}}.$$

(b) In some practical situations, the individual level data may not be available and obtaining the maximum likelihood estimator in (a) is impossible. However, researchers from the two centers may report separate maximum likelihood estimators. Suppose $(\widehat{\sigma}_k^2, \widehat{\alpha}_k, \widehat{\beta}_k)$ is the maximum likelihood estimator for $(\sigma_k^2, \alpha_k, \beta)$ using ONLY data from center $k = 1, 2$. In such situations, ONLY $(\widehat{\sigma}_k^2, \widehat{\alpha}_k, \widehat{\beta}_k), k = 1, 2$, are available.

(i) Find the asymptotic distribution of $\sqrt{n}(\widehat{\beta}_k - \beta)$ in terms of the parameter $(\sigma_k^2, \alpha_k, \beta)$. Suggest a consistent estimator $\widehat{V}_k$ of the asymptotic variance V_k using ONLY $(\widehat{\sigma}_k^2, \widehat{\alpha}_k, \widehat{\beta}_k)$.

Solution

Let's define $\xi_k = (\sigma_k^2, \alpha_k, \beta)$ and $\widehat{\xi}_k = (\widehat{\sigma}_k^2, \widehat{\alpha}_k, \widehat{\beta}_k)$. The joint likelihood and log-likelihood are

$$\begin{aligned}
P(\boldsymbol{Y}_k, \boldsymbol{X}_k | \xi_k) &= \prod_{i=1}^{n} P(Y_{ik}, X_{ik} | \xi_k) \\
\Rightarrow l_n(\xi_k) \equiv \log(P(\boldsymbol{Y}_k, \boldsymbol{X}_k | \xi_k)) &= \sum_{i=1}^{n} \log(P(Y_{ik}, X_{ik} | \xi_k)) \\
&= \sum_{i=1}^{n} \left[\log\left(P\left(X_{ik} | \sigma_k^2\right)\right) + \log\left(P\left(Y_{ik} | X_{ik}; \alpha_k, \beta\right)\right)\right] \\
&= \sum_{i=1}^{n} \log\left(P\left(X_{ik} | \sigma_k^2\right)\right) + \sum_{i=1}^{n} \log\left(P\left(Y_{ik} | X_{ik}; \alpha_k, \beta\right)\right)
\end{aligned}$$

Using what we calculated from part (a), with some similarities except that $\widehat{\beta}$ is now $\widehat{\beta}_k$, we have

$$\begin{aligned}
E\left[-\frac{\partial^2 l_n(\xi_k)}{\partial(\sigma_k^2)^2}\right] &= \frac{n}{2\sigma_k^4} \\
E\left[-\frac{\partial^2 l_n(\xi_k)}{\partial \alpha_k^2}\right] &= n \cdot \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} \\
E\left[-\frac{\partial^2 l_n(\xi_k)}{\partial \alpha_k \partial \beta}\right] &= n \cdot \sigma_k^2 \beta \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} \\
E\left[-\frac{\partial^2 l_n(\xi_k)}{\partial \beta^2}\right] &= n \cdot \sigma_k^2 \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\}(1 + \sigma_k^2\beta^2)
\end{aligned}$$

with all other elements of $I_n(\xi_k)$ equaling 0. So for $k = 1, 2$, we have that

$$\begin{aligned}
I_n(\xi_k) &= \begin{bmatrix} \frac{n}{2\sigma_k^4} & 0 & 0 \\ 0 & n \cdot \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} & n \cdot \sigma_k^2\beta \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} \\ 0 & n \cdot \sigma_k^2\beta \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\} & n \cdot \sigma_k^2 \exp\left\{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2\right\}(1 + \sigma_k^2\beta^2) \end{bmatrix} \\
&= n \cdot I(\xi_k)
\end{aligned}$$

By Weak Law of Large Numbers, $\frac{1}{n} I_n(\xi_k) \to_p I(\xi_k)$. From MLE theory with the regularity conditions satisfied, we know that the MLE is asymptotically efficient and that

$$\sqrt{n}\left(\widehat{\xi}_{k,MLE} - \xi_k\right) \to_d \mathcal{N}(\mathbf{0}, I^{-1}(\xi_k))$$

and by the Delta Method or continuous mapping theorem

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_k - \beta\right) &= \sqrt{n}\left(g(\widehat{\xi}_k) - g(\xi_k)\right) \\
&\to_d \nabla g(\xi_k) \cdot \mathcal{N}(\mathbf{0}, I^{-1}(\xi_k)) \\
&= \mathcal{N}(0, \nabla g(\xi_k) \cdot I^{-1}(\xi_k) \cdot (\nabla g(\xi_k))^T) \\
&= \mathcal{N}(0, I^{-1}(\beta))
\end{aligned}$$

where $g(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}$, a linear transformation of $\boldsymbol{x}$ with $\boldsymbol{a} = (0, 0, 1)^T$. Hence $I^{-1}(\beta) \equiv \boldsymbol{a}^T I^{-1}(\xi_k) \boldsymbol{a}$ corresponds to the third row and third column of the inverted $I(\xi_k)$. Using the same strategy as part (a) because of a similar situation, we can partition the 3 by 3 inverted matrix and focus on the bottom right 2 by 2 matrix to extract the bottom right element. We'll see that

$$\sqrt{n}\left(\widehat{\beta}_k - \beta\right) \to_d \mathcal{N}\left(0, \frac{1}{\sigma_k^2 e^{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2}}\right)$$

Since the MLE is consistent and by continuous mapping theorem, the asymptotic variance $V_k(\xi_k) = \frac{1}{\sigma_k^2 e^{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2}}$ can be estimated with the MLE $\widehat{\xi}_k$. Hence $\widehat{V}_k(\widehat{\xi}_k) = \frac{1}{\widehat{\sigma}_k^2 e^{\widehat{\alpha}_k + \frac{1}{2}\widehat{\sigma}_k^2\widehat{\beta}_k^2}}$.

(ii) To obtain a single estimator of β, one may consider $g(\widehat{\beta}_1, \widehat{\beta}_2)$, where $g(x,y)$ is a known, continuously differentiable, scalar function and $g(x,x) = x$ for any x. Find the asymptotic distribution of $\sqrt{2n}\left(g(\widehat{\beta}_1, \widehat{\beta}_2) - \beta\right)$ in terms of β, V_1, and V_2.

Solution

Before proceeding, using part (b)(i), notice that

$$\sqrt{n}\left(\begin{bmatrix}\widehat{\beta}_1\\ \widehat{\beta}_2\end{bmatrix} - \begin{bmatrix}\beta\\ \beta\end{bmatrix}\right) \to_d \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\right) \text{ where } V_k = \frac{1}{\sigma_k^2 e^{\alpha_k + \frac{1}{2}\sigma_k^2\beta^2}} \text{ for } k = 1, 2.$$

Using this, we can find the asymptotic distribution of $\sqrt{2n}\left(g(\widehat{\beta}_1, \widehat{\beta}_2) - \beta\right)$. By Delta Method and given that $g(x,x) = x$, we see that

$$\begin{aligned}\sqrt{2n}\left(g(\widehat{\beta}_1, \widehat{\beta}_2) - \beta\right) &= \sqrt{2n}\left(g(\widehat{\beta}_1, \widehat{\beta}_2) - g(\beta,\beta)\right)\\ &\to_d \sqrt{2}\cdot\nabla g(\beta,\beta)\cdot\mathcal{N}\left(\mathbf{0}, \begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\right)\\ &= \mathcal{N}\left(\mathbf{0}, 2\cdot\nabla g(\beta,\beta)\cdot\begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\cdot(\nabla g(\beta,\beta))^T\right)\end{aligned}$$

Note that $\nabla g(x,y) = \begin{bmatrix}\dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y}\end{bmatrix} \equiv \begin{bmatrix}\dot{g}_1(x,y) & \dot{g}_2(x,y)\end{bmatrix}$. Therefore $\nabla g(\beta,\beta) = \begin{bmatrix}\dot{g}_1(\beta) & \dot{g}_2(\beta)\end{bmatrix}$ So the asymptotic variance is

$$\begin{aligned}2\cdot\nabla g(\beta,\beta)\cdot\begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\cdot(\nabla g(\beta,\beta))^T &= 2\cdot\begin{bmatrix}\dot{g}_1(\beta) & \dot{g}_2(\beta)\end{bmatrix}\begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\begin{bmatrix}\dot{g}_1(\beta) & \dot{g}_2(\beta)\end{bmatrix}\\ &= 2\left((\dot{g}_1(\beta))^2 V_1 + (\dot{g}_2(\beta))^2 V_2\right)\end{aligned}$$

(iii) Next, the goal is to find a function $g_{opt}(x,y)$ which satisfies the conditions in (ii) and which minimizes the asymptotic variance in (ii). Note that g_{opt} may not be unique. Write down the constraints which implicitly define $g_{opt}(x,y)$ and show that one such function is $g^*_{opt}(x,y) = (V_2 x + V_1 y)/(V_1 + V_2)$.

Solution

The conditions in (ii) require $g(x,y)$ to be

(1) continuously differentiable: A function f is cont. diff. if $f'(x)$ exists and is itself continuous.

(2) a scalar function

(3) $g(x,x) = x$ for any x

For simplicity, let us suppose $g(x,y)$ is a linear transformation of the vector $\begin{bmatrix}x\\ y\end{bmatrix}$ characterized as

$$g(x,y) = \begin{bmatrix}a & b\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = ax + by \text{ for constants } a \text{ and } b.$$

This is continuously differentiable and a scalar function. To satisfy the third condition, notice that

$$\begin{aligned}g(x,x) &= ax + bx = x\\ \Leftrightarrow a + b &= 1\\ \Rightarrow b &= 1 - a\end{aligned}$$

So for a linear transformation satisfying all three conditions, $g(x,y) = ax + (1-a)y$ is one such function. Now to minimize the asymptotic variance and obtain g^*_{opt} using part (ii), we have

$$\begin{aligned}\nabla g(\beta,\beta)\cdot\begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\cdot(\nabla g(\beta,\beta))^T &= \begin{bmatrix}a & 1-a\end{bmatrix}\begin{bmatrix}V_1 & 0\\ 0 & V_2\end{bmatrix}\begin{bmatrix}a\\ 1-a\end{bmatrix}\\ &= a^2 V_1 + (1-a)^2 V_2\\ &\equiv \Sigma(a)\end{aligned}$$

Minimize the variance by differentiating with respect to a.

$$\begin{aligned}\frac{\partial\Sigma(a)}{\partial a} &= 2aV_1 - 2(1-a)V_2 = 0\\ \Leftrightarrow a(V_1 + V_2) &= V_2\\ \Rightarrow a &= \frac{V_2}{V_1 + V_2}\end{aligned}$$

So one possible choice for g is $g^*_{opt}(x,y) = \dfrac{V_2 x + V_1 y}{V_1 + V_2}$.

(iv) Replacing V_1 and V_2 in $g^*_{opt}(x,y)$ in (iii) by $\widehat{V}_1$ and $\widehat{V}_2$ yields $\widehat{g}^*_{opt}(x,y)$. Derive the asymptotic distribution of $\sqrt{2n}\left[\widehat{g}^*_{opt}(\widehat{\beta}_1,\widehat{\beta}_2)-\beta\right]$. What is the asymptotic relative efficiency of $\widehat{g}^*_{opt}(\widehat{\beta}_1,\widehat{\beta}_2)$ with respect to $\widehat{\beta}$ given in (a)?

Solution

Using $g^*_{opt}(x,y)$ from part (iii), the asymptotic variance of $\sqrt{2n}\left(\widehat{g}^*_{opt}(\widehat{\beta}_1,\widehat{\beta}_2)-\beta\right)$ is

$$
\begin{aligned}
2\left((\dot{g}_1(\beta))^2 V_1 + (\dot{g}_2(\beta))^2 V_2\right) &= 2\left(\left(\frac{V_2}{V_1+V_2}\right)^2 V_1 + \left(\frac{V_1}{V_1+V_2}\right)^2 V_2\right) \\
&= \frac{2V_1V_2}{V_1+V_2}
\end{aligned}
$$

where $V_k = \frac{1}{\sigma_k^2 e^{\alpha_k+\frac{1}{2}\sigma_k^2\beta^2}}$ for $k=1,2$.

From part (a), the asymptotic variance of $\sqrt{2n}\left(\widehat{\beta}-\beta\right)$ was

$$
\begin{aligned}
\frac{2}{\sigma_1^2 e^{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2}+\sigma_2^2 e^{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2}} &= 2\left[\sigma_1^2 e^{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2}+\sigma_2^2 e^{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2}\right]^{-1} \\
&= 2\left[V_1^{-1}+V_2^{-1}\right]^{-1} \\
&= 2\left[\frac{V_2+V_1}{V_1V_2}\right]^{-1} \\
&= \frac{2V_1V_2}{V_1+V_2}
\end{aligned}
$$

Hence the asymptotic relative efficiency of $\widehat{g}^*_{opt}(\widehat{\beta}_1,\widehat{\beta}_2)$ equals that of $\widehat{\beta}$.

(c) Now, suppose that we assume $\alpha_1 = \alpha_2 = \alpha$. Let $\widehat{\beta}_r$ be the maximum likelihood estimator using the combined data from both centers. That is, we conduction maximum likelihood estimation using data from both centers, as in (a), under this additional restriction. What is the asymptotic distribution of $\sqrt{2n}(\widehat{\beta}_r - \beta)$? Give a sufficient and necessary condition that $\widehat{\beta}_r$ and $\widehat{\beta}$ (in (a)) have the same asymptotic variances.

Solution

Let us denote $\xi = (\sigma_1^2, \sigma_2^2, \alpha, \beta)$. The joint likelihood and log-likelihood are

$$
\begin{aligned}
P(\boldsymbol{Y}, \boldsymbol{X}|\xi) &= \prod_k \prod_i P(Y_{ik}, X_{ik}|\xi) \\
&= \prod_k \prod_i P\left(X_{ik}|\sigma_k^2\right) P(Y_{ik}|X_{ik}; \alpha, \beta) \\
\Rightarrow l_{2n}(\xi) \equiv \log(P(\boldsymbol{Y}, \boldsymbol{X}|\xi)) &= \sum_k \sum_i \log\left[P\left(X_{ik}|\sigma_k^2\right) P(Y_{ik}|X_{ik}; \alpha, \beta)\right] \\
&= \sum_k \sum_i \log\left(P\left(X_{ik}|\sigma_k^2\right)\right) + \sum_k \sum_i \log(P(Y_{ik}|X_{ik}; \alpha, \beta))
\end{aligned}
$$

For σ_k^2, we have the same results as part (a).

$$
\begin{aligned}
E\left[-\frac{\partial^2 l_{2n}(\xi_k)}{\partial(\sigma_k^2)^2}\right] &= \frac{n}{2\sigma_k^4} \\
E\left[-\frac{\partial^2 l_{2n}(\xi_k)}{\partial\sigma_k^2 \partial\sigma_{k'}^2}\right] &= E\left[-\frac{\partial^2 l_{2n}(\xi_k)}{\partial\sigma_k^2 \partial\alpha}\right] = E\left[-\frac{\partial^2 l_{2n}(\xi_k)}{\partial\sigma_k^2 \partial\beta}\right] = 0
\end{aligned}
$$

For α, we have

$$
\begin{aligned}
\frac{\partial l_{2n}(\xi)}{\partial\alpha} &= \sum_k \sum_i -\exp(\alpha + \beta x_{ik}) + y_{ik} \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial\alpha^2} &= -\sum_k \sum_i \exp(\alpha + \beta x_{ik}) \\
&= -e^{\alpha} \sum_k \sum_i e^{\beta x_{ik}} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\alpha^2}\right] &= e^{\alpha} \sum_k \sum_i E\left[e^{\beta x_{ik}}\right] = e^{\alpha} \sum_k n \exp\left\{\frac{1}{2}\sigma_k^2\beta^2\right\} \\
&= ne^{\alpha} \sum_k e^{\frac{1}{2}\sigma_k^2\beta^2} \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial\alpha\partial\beta} &= \sum_k \sum_i -x_{ik}\exp(\alpha + \beta x_{ik}) \\
&= -e^{\alpha} \sum_k \sum_i x_{ik} e^{\beta x_{ik}} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\alpha\partial\beta}\right] &= e^{\alpha} \sum_k \sum_i E\left[x_{ik} e^{\beta x_{ik}}\right] = e^{\alpha} \sum_k \sum_i \sigma_k^2 \beta e^{\frac{1}{2}\sigma_k^2\beta^2} \\
&= n\beta e^{\alpha} \sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}
\end{aligned}
$$

For β, we have

$$
\begin{aligned}
\frac{\partial l_{2n}(\xi)}{\partial\beta} &= \sum_k \sum_i \{-x_{ik}\exp(\alpha + \beta x_{ik}) + y_{ik}x_{ik}\} \\
\Rightarrow \frac{\partial^2 l_{2n}(\xi)}{\partial\beta^2} &= -\sum_k \sum_i x_{ik}^2 \exp(\alpha + \beta x_{ik}) \\
&= -e^{\alpha} \sum_k \sum_i x_{ik}^2 e^{\beta x_{ik}} \\
\Rightarrow E\left[-\frac{\partial^2 l_{2n}(\xi)}{\partial\beta^2}\right] &= e^{\alpha} \sum_k \sum_i E\left[x_{ik}^2 e^{\beta x_{ik}}\right] \\
&= ne^{\alpha} \sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2} \left(\beta^2\sigma_k^2 + 1\right)
\end{aligned}
$$

The Fisher Information matrix is

$$
\begin{aligned}
I_{2n}(\xi) &= \begin{bmatrix} \frac{n}{2\sigma_1^4} & 0 & 0 & 0 \\ 0 & \frac{n}{2\sigma_2^4} & 0 & 0 \\ 0 & 0 & ne^{\alpha}\sum_k e^{\frac{1}{2}\sigma_k^2\beta^2} & n\beta e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2} \\ 0 & 0 & n\beta e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2} & ne^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\left(\beta^2\sigma_k^2+1\right) \end{bmatrix} \\
&= 2n \begin{bmatrix} \frac{1}{4\sigma_1^4} & 0 & 0 & 0 \\ 0 & \frac{1}{4\sigma_2^4} & 0 & 0 \\ 0 & 0 & \frac{1}{2}e^{\alpha}\sum_k e^{\frac{1}{2}\sigma_k^2\beta^2} & \frac{1}{2}\beta e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2} \\ 0 & 0 & \frac{1}{2}\beta e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2} & \frac{1}{2}e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\left(\beta^2\sigma_k^2+1\right) \end{bmatrix} \\
&= 2n \cdot I(\xi)
\end{aligned}
$$

By the Weak Law of Large Numbers, $\dfrac{1}{2n}I_{2n}(\xi) \to_p I(\xi)$. From the MLE theory with the regularity conditions satisfied, we have

$$\sqrt{2n}\left(\widehat{\xi}-\xi\right) \to_d \mathcal{N}\left(\mathbf{0}, I^{-1}(\xi)\right)$$

and by Delta Method or continuous mapping theorem,

$$
\begin{aligned}
\sqrt{2n}\left(\widehat{\beta}_r-\beta\right) &= \sqrt{2n}\left(g(\widehat{\xi})-g(\xi)\right) \\
&\to_d \nabla g(\xi)\cdot\mathcal{N}\left(\mathbf{0}, I^{-1}(\xi)\right) \\
&= \mathcal{N}\left(0, \nabla g(\xi)\cdot I^{-1}(\xi)\cdot(\nabla g(\xi))^T\right) \\
&= \mathcal{N}\left(0, I^{-1}(\beta)\right)
\end{aligned}
$$

Just like in part (a) and (b)(i), we want to extract the bottom right element of the inverted $I(\xi)$ matrix. After partitioning $I(\xi)$, we just need to re-apply the hint on the bottom right 2 by 2 matrix to get the bottom right element. Let $\begin{bmatrix} A_* & B_* \\ B_*^T & C_* \end{bmatrix}$ where $A_* = \frac{1}{2}e^{\alpha}\sum_k e^{\frac{1}{2}\sigma_k^2\beta^2}$, $B_* = \frac{1}{2}\beta e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}$, and $C_* = \frac{1}{2}e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\left(\beta^2\sigma_k^2+1\right)$. The asymptotic variance is therefore

$$
\begin{aligned}
\left(C_* - B_*^T A_*^{-1} B_*\right)^{-1} &= \left(C_* - B_*^2 A_*^{-1}\right)^{-1} \\
&= \left(\frac{1}{2}e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\left(\beta^2\sigma_k^2+1\right) - \left[\frac{1}{2}\beta e^{\alpha}\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\right]^2 \frac{1}{\frac{1}{2}e^{\alpha}\sum_k e^{\frac{1}{2}\sigma_k^2\beta^2}}\right)^{-1} \\
&= 2e^{-\alpha}\left(\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\left(\beta^2\sigma_k^2+1\right) - \left[\sum_k \sigma_k^2 e^{\frac{1}{2}\sigma_k^2\beta^2}\right]^2 \frac{\beta^2}{\sum_k e^{\frac{1}{2}\sigma_k^2\beta^2}}\right)^{-1} \\
&\quad \text{Note: Suppose that } \sigma_1^2 = \sigma_2^2 \equiv \sigma^2 \text{ and } \alpha_1 = \alpha_2 \equiv \alpha \\
&= 2e^{-\alpha}\left(2\sigma^2 e^{\frac{1}{2}\sigma^2\beta^2}\left(\beta^2\sigma^2+1\right) - \left[2\sigma^2 e^{\frac{1}{2}\sigma^2\beta^2}\right]^2 \frac{\beta^2}{2e^{\frac{1}{2}\sigma^2\beta^2}}\right)^{-1} \\
&= 2e^{-\alpha}\left(2\sigma^2 e^{\frac{1}{2}\sigma^2\beta^2}\left(\beta^2\sigma^2+1\right) - 2\beta^2\sigma^4 e^{\frac{1}{2}\sigma^2\beta^2}\right)^{-1} \\
&= 2e^{-\alpha}\left(2\sigma^2 e^{\frac{1}{2}\sigma^2\beta^2}\right)^{-1} \\
&= \frac{1}{\sigma^2\exp\left\{\alpha+\frac{1}{2}\sigma^2\beta^2\right\}}
\end{aligned}
$$

With what we supposed, the asymptotic variance in part (a) would become

$$\frac{2}{\sigma_1^2 e^{\alpha_1+\frac{1}{2}\sigma_1^2\beta^2} + \sigma_2^2 e^{\alpha_2+\frac{1}{2}\sigma_2^2\beta^2}} = \frac{2}{\sigma^2 e^{\alpha+\frac{1}{2}\sigma^2\beta^2} + \sigma^2 e^{\alpha+\frac{1}{2}\sigma^2\beta^2}} = \frac{1}{\sigma^2\exp\left\{\alpha+\frac{1}{2}\sigma^2\beta^2\right\}}$$

Therefore, as long as $\alpha = \alpha_1 = \alpha_2$ and $\sigma^2 = \sigma_1^2 = \sigma_2^2$, then both asymptotic variances will be equal.

2 Theory 2010

2 Part 1

2.1.1 Question 1

1. Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d., where

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N}_2(\mu, \Sigma),$$

$i = 1, \ldots, n$, where $\mu = (\mu_1, \mu_2)^T$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

(a) Suppose that $\sigma_{12} = 0$ and all other parameters are unknown. Consider the hypothesis $H_0 : \frac{\sigma_2^2}{\sigma_1^2} = \Delta_0$ versus $H_1 : \frac{\sigma_2^2}{\sigma_1^2} \neq \Delta_0$, where $\Delta_0 > 0$ is a specified constant. Derive the UMPU size α test for this hypothesis and find the simplest possible form of the test statistic and critical value for the test.

Solution

The hypotheses can be restated as

$$H_0 : \frac{\sigma_2^2}{\sigma_1^2} = \Delta_0 \Leftrightarrow H_0 : \frac{1}{2\Delta_0\sigma_1^2} - \frac{1}{2\sigma_2^2} = 0 \Leftrightarrow H_0 : \theta = 0 \text{ vs. } H_1 : \theta \neq 0.$$

To derive the UMPU test, we need to find the sufficient statistic U for θ and sufficient statistic T for the nuisance parameter ξ by writing the likelihood in a multi-parameter exponential family form

$$p_{\theta,\xi}(\boldsymbol{x}, \boldsymbol{y}) = c(\theta, \xi) \exp\left[\theta u(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{k} \xi_j T_j(\boldsymbol{x}, \boldsymbol{y})\right].$$

The joint likelihood is

$$\begin{aligned}
P(\boldsymbol{Y}, \boldsymbol{X} | \mu, \Sigma) &= \prod_{i=1}^{n} P(Y_i, X_i | \mu, \Sigma) \\
&= \prod_{i=1}^{n} (2\pi \det(\Sigma))^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{(x_i - \mu_x)^2}{\sigma_1^2} + \frac{(y_i - \mu_y)^2}{\sigma_2^2}\right)\right\} \\
&\propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\left(\frac{x_i^2 - 2x_i\mu_x + \mu_x^2}{\sigma_1^2} + \frac{y_i^2 - 2y_i\mu_y + \mu_y^2}{\sigma_2^2}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{\sum_i x_i^2 - 2\mu_x \sum_i x_i + n\mu_x^2}{\sigma_1^2} + \frac{\sum_i y_i^2 - 2\mu_y \sum_i y_i + n\mu_y^2}{\sigma_2^2}\right)\right\} \\
&\propto \exp\left\{-\frac{\sum_i x_i^2 - 2\mu_x \sum_i x_i}{2\sigma_1^2} - \frac{\sum_i y_i^2 - 2\mu_y \sum_i y_i}{2\sigma_2^2}\right\} \\
&\propto \exp\left\{-\frac{\sum_i x_i^2}{2\sigma_1^2} + \frac{\mu_x \sum_i x_i}{\sigma_1^2} - \frac{\sum_i y_i^2}{2\sigma_2^2} + \frac{\mu_y \sum_i y_i}{\sigma_2^2}\right\}
\end{aligned}$$

Notice that

$$\begin{aligned}
-\frac{1}{2\sigma_1^2}\sum_i x_i^2 - \frac{1}{2\sigma_2^2}\sum_i y_i^2 &= -\frac{1}{2\sigma_1^2}\sum_i x_i^2 - \frac{1}{2\sigma_2^2}\sum_i y_i^2 + \left[\frac{1}{2\sigma_1^2\Delta_0}\sum_i y_i^2 - \frac{1}{2\sigma_1^2\Delta_0}\sum_i y_i^2\right] \\
&= \left(\frac{1}{2\sigma_1^2\Delta_0} - \frac{1}{2\sigma_2^2}\right)\sum_i y_i^2 - \frac{1}{2\sigma_1^2}\left(\frac{1}{\Delta_0}\sum_i y_i^2 + \sum_i x_i^2\right) \\
&\equiv \theta \sum_i y_i^2 - \xi_1 \left(\frac{1}{\Delta_0}\sum_i y_i^2 + \sum_i x_i^2\right)
\end{aligned}$$

Plugging this result back into the joint likelihood,

$$\begin{aligned}
P(\boldsymbol{Y}, \boldsymbol{X} | \mu, \Sigma) &\propto \exp\left\{\theta \sum_i y_i^2 - \xi_1\left(\frac{1}{\Delta_0}\sum_i y_i^2 + \sum_i x_i^2\right) + \frac{n\mu_x \bar{x}}{\sigma_1^2} + \frac{n\mu_y \bar{y}}{\sigma_2^2}\right\} \\
&= \exp\left\{\theta \sum_i y_i^2 - \xi_1\left(\frac{1}{\Delta_0}\sum_i y_i^2 + \sum_i x_i^2\right) + \xi_2 \bar{x} + \xi_3 \bar{y}\right\}
\end{aligned}$$

And so, we see that $U(\boldsymbol{x},\boldsymbol{y}) = \sum_i y_i^2, T_1(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\Delta_0}\sum_i y_i^2 + \sum_i x_i^2, T_2(\boldsymbol{x},\boldsymbol{y}) = \bar{x}, T_3(\boldsymbol{x},\boldsymbol{y}) = \bar{y}$. Now we need to find a function of U and $T \equiv (T_1, T_2, T_3)$ to construct the hypothesis test.

Hint: Since the X and Y are Gaussian distributed and U and T are functions of $\bar{x}$, $\bar{y}$, $\sum_i x_i^2$, and $\sum_i y_i^2$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, perhaps the statistic is at the very least a "function" of the F distribution. In addition, the null hypothesis is testing whether or not the ratio of two variances is constant.

Based on what is given, we know that

$$\begin{aligned}
\frac{\sum_i (y_i-\bar{y})^2/\sigma_2^2}{\sum_i (x_i-\bar{x})^2/\sigma_1^2} &= \frac{(n-1)S_y^2/\sigma_2^2}{(n-1)S_x^2/\sigma_1^2} \\
&= \frac{\chi^2_{n-1}}{\chi^2_{n-1}} \\
&= \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{n-1}/(n-1)} \sim F(n-1,n-1)
\end{aligned}$$

And under $H_0 : \frac{\sigma_2^2}{\sigma_1^2} = \Delta_0$,

$$\frac{\sum_i (y_i-\bar{y})^2/\sigma_2^2}{\sum_i (x_i-\bar{x})^2/\sigma_1^2} = \frac{\sum_i (y_i-\bar{y})^2/(\Delta_0\sigma_1^2)}{\sum_i (x_i-\bar{x})^2/\sigma_1^2} = \frac{\frac{1}{\Delta_0}\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2} \sim^{H_0} F(n-1,n-1)$$

Unfortunately, for this potential statistic, we would need to know the conditional distribution of $U|T$. Using Remark 2.6 from BIOS 761 regarding the two sided test, if we can find $V \equiv h(U,T) = a(t)U + b(t)$ with $a(t) > 0$, the 2nd constraint used in finding the constants of ϕ becomes

$$E_{\theta_0}\left[V\phi(\boldsymbol{x},\boldsymbol{y})|T=t\right] = \alpha E_{\theta_0}\left[V|T=t\right]$$

and if V is independent of T on the boundary, then the test is unconditional. So if we can transform our potential statistic above so that it fits this description, we'll be finished.

What would the ratio of their statistics look like under the null? Specifically, if the null were true, then the total variance of $\boldsymbol{X}$ and $\frac{1}{\Delta_0}\boldsymbol{Y}$ would equal the variance of $\frac{1}{\Delta_0}\boldsymbol{Y}$.

$$\begin{aligned}
\frac{\frac{1}{\Delta_0}\widehat{\sigma}_2^2}{\widehat{\sigma}_1^2 + \frac{1}{\Delta_0}\widehat{\sigma}_2^2} &= \frac{\frac{1}{\Delta_0(n-1)}\sum_i (y_i-\bar{y})^2}{\frac{1}{(n-1)}\sum_i (x_i-\bar{x})^2 + \frac{1}{\Delta_0(n-1)}\sum_i (y_i-\bar{y})^2} \\
&= \frac{\frac{(n-1)}{\Delta_0}\frac{\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2}}{(n-1) + \frac{(n-1)}{\Delta_0}\frac{\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2}} \\
&= \frac{(n-1)F}{(n-1)+(n-1)F} \\
&= \frac{\frac{1}{\Delta_0}(U - nT_3^2)}{T_1 - nT_2^2 - \frac{1}{\Delta_0}nT_3^2} \\
&\equiv V(U,T)
\end{aligned}$$

where F corresponds to a random variable following the F distribution. There is something to note. If $X \sim F(d_1,d_2) \Rightarrow \frac{d_1 X}{d_2 + d_1 X} \sim Beta\left(\frac{d_1}{2},\frac{d_2}{2}\right)$. In this case, our statistic follows $Beta\left(\frac{n-1}{2},\frac{n-1}{2}\right)$. Notice that the statistic is linear in U, $a(t) > 0$, and $V \perp T$ because V's distribution doesn't depend on T so we have our test statistic V. Our test is therefore

$$\phi(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} 1 & \text{if } V < c_1 \text{ or } V > c_2 \\ 0 & \text{otherwise} \end{cases}$$

Using the two constraints to find c_1 and c_2, $E\left[\phi(\boldsymbol{x},\boldsymbol{y})\right] = \alpha$ and $E\left[V\phi(\boldsymbol{x},\boldsymbol{y})\right] = \alpha E\left[V\right]$. Under H_0, notice that V (bounded between 0 and 1) is symmetric ($V \stackrel{d}{=} 1-V$), so $c_1 = 1 - c_2$ which simplifies the calculation. So

$$\begin{aligned}
\alpha = E\left[\phi(\boldsymbol{x},\boldsymbol{y})\right] &= P\left(V < c_1 \text{ or } V > c_2\right) \\
&= P\left(V < c_1\right) + P\left(V > c_2\right) = P\left(V < c_1\right) + P\left(1 - V < 1 - c_2\right) \\
&= 2P\left(V < c_1\right) \\
\Leftrightarrow c_1 &= F_V^{-1}\left(\frac{\alpha}{2}\right)
\end{aligned}$$

(b) Derive the simplest possible form of the size α likelihood ratio test corresponding to part (a), and compare it to the UMPU test.

Solution

Using part (a), the joint likelihood and log likelihood are

$$\begin{aligned}
P\left(\boldsymbol{X}, \boldsymbol{Y} | \mu, \Sigma\right) &= (2\pi\sigma_1^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_1^2}\sum_i (x_i-\mu_1)^2\right\} \times (2\pi\sigma_2^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_2^2}\sum_i (y_i-\mu_2)^2\right\} \\
&\propto (\sigma_1^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_1^2}\sum_i (x_i-\mu_1)^2\right\} \times (\sigma_2^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_2^2}\sum_i (y_i-\mu_2)^2\right\} \\
&= \exp\left\{-\frac{n}{2}\log(\sigma_1^2) - \frac{1}{2\sigma_1^2}\sum_i (x_i-\mu_1)^2 - \frac{n}{2}\log(\sigma_2^2) - \frac{1}{2\sigma_2^2}\sum_i (y_i-\mu_2)^2\right\} \\
\Rightarrow l_n(\mu, \Sigma) &= -\frac{n}{2}\log(\sigma_1^2) - \frac{1}{2\sigma_1^2}\sum_i (x_i-\mu_1)^2 - \frac{n}{2}\log(\sigma_2^2) - \frac{1}{2\sigma_2^2}\sum_i (y_i-\mu_2)^2
\end{aligned}$$

Under H_0, the MLEs are $\tilde{\mu}_1 = \bar{x}$, $\tilde{\mu}_2 = \bar{y}$, and $\tilde{\sigma}^2 = \frac{1}{2n}\left(\sum_i (x_i-\bar{x})^2 + \frac{1}{\Delta_0}\sum_i (y_i-\bar{y})^2\right)$. Under $H_0 \cup H_1$, the MLEs are $\widehat{\mu}_1 = \bar{x}$, $\widehat{\mu}_2 = \bar{y}$, $\widehat{\sigma}_1^2 = \frac{\sum_i (x_i-\bar{x})^2}{n}$, and $\widehat{\sigma}_2^2 = \frac{\sum_i (y_i-\bar{y})^2}{n}$.

The LRT has the form

$$\begin{aligned}
\Lambda &\equiv \frac{\sup_{\theta\in\Theta_0} L(\mu,\Sigma)}{\sup_{\theta\in\Theta_0\cup\Theta_1} L(\mu,\Sigma)} < k \\
\Leftrightarrow \log(\Lambda) &< k' \\
\Leftrightarrow l_n(\tilde{\mu}, \tilde{\Sigma}) - l_n(\widehat{\mu}, \widehat{\Sigma}) &< k'
\end{aligned}$$

and we find that $l_n(\tilde{\mu}, \tilde{\Sigma}) = -n\log(\tilde{\sigma}^2) - \frac{n}{2}\log(\Delta_0) - \frac{n}{2}$ as well as $l_n(\widehat{\mu}, \widehat{\Sigma}) = -\frac{n}{2}\log(\widehat{\sigma}_1^2) - \frac{n}{2}\log(\widehat{\sigma}_2^2) - \frac{n}{2}$. So their difference is

$$\begin{aligned}
l_n(\tilde{\mu}, \tilde{\Sigma}) - l_n(\widehat{\mu}, \widehat{\Sigma}) &= \frac{n}{2}\log\left(\frac{\widehat{\sigma}_1^2\widehat{\sigma}_2^2}{\tilde{\sigma}^4\Delta_0}\right) \\
&= \frac{n}{2}\log\left(\frac{\sum_i (x_i-\bar{x})^2 \cdot \sum_i (y_i-\bar{y})^2}{\left(\frac{1}{2}\sum_i (x_i-\bar{x})^2 + \frac{1}{2\Delta_0}\sum_i (y_i-\bar{y})^2\right)^2\Delta_0}\right) \\
&= \frac{n}{2}\log\left(\frac{\frac{\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2}}{\left(\frac{1}{2} + \frac{1}{2\Delta_0}\frac{\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2}\right)^2\Delta_0}\right) < k \\
&\Leftrightarrow \frac{\frac{1}{\Delta_0}\frac{\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2}}{\left(1 + \frac{1}{\Delta_0}\frac{\sum_i (y_i-\bar{y})^2}{\sum_i (x_i-\bar{x})^2}\right)^2} < k' \\
&\Leftrightarrow V(1-V) < k' \\
&\Leftrightarrow V < c_1 \text{ or } V > c_2
\end{aligned}$$

Hence the LRT is equivalent to the UMPU test in part (a). The only difference is that the UMPU test requires unbiasedness, that is, (c_1, c_2) must satisfy $P(V < c_1) + P(V > c_2) = \alpha$.

(c) Now suppose that σ_{12} is *unknown* and all other parameters are also *unknown*. Let $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}}$ denote the population correlation coefficient. Suppose we wish to test $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$.

(i) Show that the size α likelihood ratio test rejects H_0 when $|R| > c$, where R denotes the sample correlation coefficient and c is chosen to make the test size α.

Solution

Note: The answer below uses matrix properties and matrix calculus, refer to the matrix cookbook for guidance.

Let's temporarily denote $z_i = (x_i, y_i)^T$. The joint likelihood and log likelihood are

$$\begin{aligned}
P\left(\boldsymbol{X},\boldsymbol{Y}|\mu,\Sigma\right) &= \prod_{i=1}^{n} P\left(x_i, y_i|\mu,\Sigma\right) \\
&= \prod_{i=1}^{n} \left(2\pi\det(\Sigma)\right)^{-1/2} \exp\left\{-\frac{(z_i-\mu)^T\Sigma^{-1}(z_i-\mu)}{2}\right\} \\
&= \left(2\pi\det(\Sigma)\right)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(z_i-\mu)^T\Sigma^{-1}(z_i-\mu)\right\} \\
\Rightarrow l_n(\mu,\Sigma) &= -\frac{n}{2}\log\left(\det(\Sigma)\right) - \frac{1}{2}\sum_{i=1}^{n}(z_i-\mu)^T\Sigma^{-1}(z_i-\mu)
\end{aligned}$$

Under H_0,

$$\begin{aligned}
l_n(\mu,\Sigma_0) &= -\frac{n}{2}\log(\sigma_1^2\sigma_2^2) - \frac{1}{2}\left[\frac{\sum_i(x_i-\mu_1)^2}{\sigma_1^2} + \frac{\sum_i(y_i-\mu_2)^2}{\sigma_2^2}\right] \\
&= -\frac{n}{2}\left[\log(\sigma_1^2) + \log(\sigma_2^2)\right] - \frac{1}{2}\left[\frac{\sum_i(x_i-\mu_1)^2}{\sigma_1^2} + \frac{\sum_i(y_i-\mu_2)^2}{\sigma_2^2}\right] \\
&\Rightarrow \tilde{\mu}_1 = \bar{x}, \tilde{\mu}_2 = \bar{y} \\
&\Rightarrow \tilde{\sigma}_1^2 = \frac{\sum_i(x_i-\bar{x})^2}{n}, \tilde{\sigma}_2^2 = \frac{\sum_i(y_i-\bar{y})^2}{n}
\end{aligned}$$

Under the full parameter space,

$$\begin{aligned}
l_n(\mu,\Sigma) &= -\frac{n}{2}\log(\det(\Sigma)) - \frac{1}{2}\sum_{i=1}^{n}(z_i-\mu)^T\Sigma^{-1}(z_i-\mu) \\
\Rightarrow \frac{\partial l_n(\mu,\Sigma)}{\partial\mu} &= -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial\mu}\left[(z_i-\mu)^T\Sigma^{-1}(z_i-\mu)\right] = 0
\end{aligned}$$

Note: From the matrix cookbook, $\frac{\partial \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\boldsymbol{B}+\boldsymbol{B}^T)\boldsymbol{x}$

$$\begin{aligned}
&\Leftrightarrow \sum_{i=1}^{n} 2\Sigma^{-1}(z_i-\mu) = 0 \\
&\Leftrightarrow \sum_{i=1}^{n}(z_i-\mu) = 0 \\
&\Rightarrow \widehat{\mu} = \bar{z} = \begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix}
\end{aligned}$$

We can plug this result back into the log likelihood.

$$l_n(\widehat{\mu},\Sigma) = -\frac{n}{2}\log(\det(\Sigma)) - \frac{1}{2}\sum_{i=1}^{n}(z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z})$$

Looking at the second term, since the function is scalar, we have

$$\begin{aligned}
-\frac{1}{2}\sum_{i=1}^{n}(z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z}) &= -\frac{1}{2}\text{trace}\left[\sum_{i=1}^{n}(z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z})\right] = -\frac{1}{2}\sum_{i=1}^{n}\text{trace}\left[(z_i-\bar{z})^T\Sigma^{-1}(z_i-\bar{z})\right] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\text{trace}\left[\Sigma^{-1}(z_i-\bar{z})(z_i-\bar{z})^T\right] \\
&= -\frac{1}{2}\text{trace}\left[\Sigma^{-1}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})^T\right] \\
&\quad \text{Note: Let's define } B \equiv \sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})^T \\
&= -\frac{1}{2}\text{trace}\left[\Sigma^{-1}B\right]
\end{aligned}$$

So our log likelihood to be maximized is now

$$\begin{aligned}
l_n(\widehat{\mu},\Sigma) &= -\frac{n}{2}\log(\det(\Sigma)) - \frac{1}{2}\text{trace}\left[\Sigma^{-1}B\right] \\
\Rightarrow \frac{\partial l_n(\widehat{\mu},\Sigma)}{\partial\Sigma} &= -\frac{n}{2}\frac{\partial}{\partial\Sigma}\left[\log(\det(\Sigma))\right] - \frac{1}{2}\frac{\partial}{\partial\Sigma}\left[\text{trace}\left[\Sigma^{-1}B\right]\right] = 0 \\
&= -\frac{n}{2}\frac{\frac{\partial}{\partial\Sigma}\det(\Sigma)}{\det(\Sigma)} - \frac{1}{2}\frac{\partial}{\partial\Sigma}\left[\text{trace}\left[\Sigma^{-1}B\right]\right] = 0 \\
&\quad \text{Note: } \frac{\partial\det(\boldsymbol{X})}{\partial\boldsymbol{X}} = \det(\boldsymbol{X})\left(\boldsymbol{X}^{-1}\right)^T \\
&\quad \text{Note: } \frac{\partial}{\partial\boldsymbol{X}}\text{trace}\left[\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\right] = -\left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{X}^{-1}\right)^T \\
&\Leftrightarrow\ 0 = -\frac{n}{2}\left(\Sigma^{-1}\right)^T - \frac{1}{2}\left(-\Sigma^{-1}B\Sigma^{-1}\right)^T \\
&\Leftrightarrow\ 0 = \left[n\boldsymbol{I} - \Sigma^{-1}B\right]\left(\Sigma^{-1}\right)^T \\
&\Leftrightarrow\ n\boldsymbol{I} = \Sigma^{-1}B \\
&\Rightarrow\ \widehat{\Sigma} = \frac{1}{n}B = \frac{1}{n}\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})^T
\end{aligned}$$

The LRT is of the form $\phi(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} 1 & \text{if } \Lambda < k \\ 0 & \text{if } \Lambda > k \end{cases}$. In this case,

$$\begin{aligned}
\Lambda &< k \\
\Leftrightarrow \log(\Lambda) &< \log(k) \\
\Leftrightarrow l_n(\tilde{\mu},\tilde{\Sigma}) - l_n(\widehat{\mu},\widehat{\Sigma}) &< \log(k) \\
&\quad \text{Note: } l_n(\tilde{\mu},\tilde{\Sigma}) = -\frac{n}{2}\log(\det(\tilde{\Sigma})) - n \\
&\quad \text{Note: } l_n(\widehat{\mu},\widehat{\Sigma}) = -\frac{n}{2}\log(\det(\widehat{\Sigma})) - \frac{n}{2} \\
\Leftrightarrow \frac{n}{2}\log(\det(\widehat{\Sigma})) - \frac{n}{2}\log(\det(\tilde{\Sigma})) - \frac{n}{2} &< \log(k) \\
\Leftrightarrow \log\left(\frac{\det(\widehat{\Sigma})}{\det(\tilde{\Sigma})}\right) &< k' \\
&\quad \text{Note: } \det(\widehat{\Sigma}) = \widehat{\sigma}_1^2\widehat{\sigma}_2^2(1-\widehat{\rho}^2) \\
&\quad \text{Note: } \det(\tilde{\Sigma}) = \widehat{\sigma}_1^2\widehat{\sigma}_2^2 \\
\Leftrightarrow \log(1-\widehat{\rho}^2) &< k' \\
\Rightarrow |\widehat{\rho}| &> k'' \equiv k
\end{aligned}$$

We have that $\rho = \dfrac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}}$ and so $\widehat{\rho} = \dfrac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2 \sum_i (y_i-\bar{y})^2}} \equiv R$, the sample correlation coefficient.

(ii) Derive the exact distribution of R under the null hypothesis and hence find an explicit expression of c for (i) above.

Solution

Assume $\rho = 0$. Let's re-express R or $\widehat{\rho}$ from part (i) in matrix/vector form where $\boldsymbol{I}$ denotes the $n \times n$ identity matrix and $\boldsymbol{M}_J = \boldsymbol{J}_n(\boldsymbol{J}_n^T\boldsymbol{J}_n)^{-1}\boldsymbol{J}_n^T$ where $\boldsymbol{J}_n$ denotes a column of n ones. I'm borrowing on some linear algebra concepts.

$$\begin{aligned}
R &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \\
&= \frac{((\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x})^T \times (\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y}}{\sqrt{\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \times \boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y}}} \\
&= \frac{((\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y})^T \times (\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y} \times \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}} \\
&= \left(\frac{(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y}}{\sqrt{n-1}\sqrt{\frac{1}{n-1}\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y}}}\right)^T \times \left(\frac{(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{n-1}\sqrt{\frac{1}{n-1}\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}}\right) \\
&\quad \text{Note: Let } S_Y = \sqrt{\frac{1}{n-1}\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y}} \text{ and } A_Y = \frac{(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{y}}{\sqrt{n-1}S_Y} \\
&= \frac{A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{n-1}S_X} = \frac{A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}/\sigma_1}{\sqrt{n-1}S_X/\sigma_1}
\end{aligned}$$

Notice that in this form, without loss of generality, set $\sigma_1^2 = \sigma_2^2 = 1$ and that

$$\begin{aligned}
(n-1)S_X^2 - (A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x})^2 &= (n-1)S_X^2 - \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)A_Y A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \\
&= \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} - \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)A_Y A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \\
&= \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)(\boldsymbol{I} - A_Y A_Y^T)(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \\
&\equiv \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}
\end{aligned}$$

where B_Y is an orthogonal projection matrix. We see that for fixed $\boldsymbol{y}$,

$$\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \perp A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \text{ b/c } B_Y A_Y^T = 0$$

Also one can see that $A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \sim \mathcal{N}(0,1)$ and $\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \sim \chi^2_{n-2}$. This is because

$$\begin{aligned}
\boldsymbol{x} &\sim \mathcal{N}(\mu_1, \boldsymbol{I}) \\
\Rightarrow (\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} &\sim \mathcal{N}(0, \boldsymbol{I} - \boldsymbol{M}_J) \\
\Rightarrow A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} &\sim \mathcal{N}(0, A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)A_Y) \\
&= \mathcal{N}(0,1) \\
\Rightarrow \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)(\boldsymbol{I} - A_Y A_Y^T)(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} &= \boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} \\
&\quad \text{Note: trace}\left[(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\right] = \text{trace}\left[(\boldsymbol{I} - \boldsymbol{M}_J)B_Y\right] = n-2 \\
&\sim \chi^2_{n-2}
\end{aligned}$$

Therefore $\dfrac{\sqrt{n-2}A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}} \sim t_{n-2}$ (when conditioning on $\boldsymbol{y}$). So R can be re-expressed once more as

$$R = \frac{A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{n-1}S_X} = \frac{A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{\boldsymbol{x}^T(\boldsymbol{I} - \boldsymbol{M}_J)B_Y(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x} + (A_Y^T(\boldsymbol{I} - \boldsymbol{M}_J)\boldsymbol{x})^2}}$$

Moving on, the cumulative density for R is

$$
\begin{aligned}
P\left(R \leq r\right) &= E\left[E\left[\mathbf{1}\left\{R \leq r\right\}|\boldsymbol{y}\right]\right] \\
&= E\left[P\left(R \leq r|\boldsymbol{y}\right)\right] \\
&= E\left[P\left(\left.\frac{A_Y^T(\boldsymbol{I}-\boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{\boldsymbol{x}^T(\boldsymbol{I}-\boldsymbol{M}_J)B_Y(\boldsymbol{I}-\boldsymbol{M}_J)\boldsymbol{x}+\left(A_Y^T(\boldsymbol{I}-\boldsymbol{M}_J)\boldsymbol{x}\right)^2}} \leq r \right| \boldsymbol{y}\right)\right] \\
&= E\left[P\left(\left.\frac{A_Y^T(\boldsymbol{I}-\boldsymbol{M}_J)\boldsymbol{x}}{\sqrt{\boldsymbol{x}^T(\boldsymbol{I}-\boldsymbol{M}_J)B_Y(\boldsymbol{I}-\boldsymbol{M}_J)\boldsymbol{x}}} \leq \frac{r}{\sqrt{1-r^2}} \right| \boldsymbol{y}\right)\right] \\
&= E\left[P\left(\left. t_{n-2} \leq \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| \boldsymbol{y}\right)\right] \\
&\quad \text{Note: If } T \sim t_\nu \Rightarrow f_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\
\Rightarrow P\left(R \leq r\right) &= \frac{\Gamma(\frac{n-1}{2})}{\sqrt{(n-1)\pi}\Gamma(\frac{n-2}{2})}\int_0^{\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}}\left(1+\frac{u^2}{n-2}\right)^{-\frac{n-1}{2}} du
\end{aligned}
$$

With this, $f(r) = \frac{\partial}{\partial r}\left[P\left(R \leq r\right)\right]$ and we can solve $\int_{-c}^{c} f(r)dr = 1-\alpha$ to obtain c.
The distribution of R is supposedly $Beta(1, n-2)$.

(iii) Derive the (appropriately normalized) asymptotic distribution of R assuming $\rho = 0$.

Solution

Looking at $R = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$, we can apply the **Central Limit Theorem** (CLT), **Weak Law of Large Numbers** (WLLN), **Continuous Mapping Theorem** (CMT), and **Slutsky's Theorem** after normalizing R.

For the numerator,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_i (x_i - \bar{x})(y_i - \bar{y}) &= \frac{1}{\sqrt{n}} \left(\sum_i (x_i - \mu_1 + \mu_1 - \bar{x})(y_i - \mu_2 + \mu_2 - \bar{y}) \right) \\
&= \frac{1}{\sqrt{n}} \left(\sum_i (x_i - \mu_1)(y_i - \mu_2) + (\mu_2 - \bar{y}) \sum_i (x_i - \mu_1) + (\mu_1 - \bar{x}) \sum_i (y_i - \mu_2) \right) + \\
&\quad \frac{1}{\sqrt{n}} \cdot n(\mu_1 - \bar{x})(\mu_2 - \bar{y}) \\
&\quad \text{Note: } \frac{1}{\sqrt{n}} (\mu_2 - \bar{y}) \sum_i (x_i - \mu_1) = O_p(1) o_p(1) = o_p(1) \\
&\quad \text{Note: } \frac{1}{\sqrt{n}} (\mu_1 - \bar{x}) \sum_i (y_i - \mu_2) = O_p(1) o_p(1) = o_p(1) \\
&\quad \text{Note: } \frac{1}{\sqrt{n}} \cdot n(\mu_1 - \bar{x})(\mu_2 - \bar{y}) = O_p(1) o_p(1) = o_p(1) \\
&\quad \text{Note: } o_p(1) + o_p(1) = o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_i (x_i - \mu_1)(y_i - \mu_2) + o_p(1) \\
&= \sigma_1 \sigma_2 \frac{1}{\sqrt{n}} \sum_i \frac{x_i - \mu_1}{\sigma_1} \cdot \frac{y_i - \mu_2}{\sigma_2} + o_p(1) \\
&\quad \text{Note: } A_i \equiv \frac{x_i - \mu_1}{\sigma_1}, B_i \equiv \frac{y_i - \mu_2}{\sigma_2} \text{ where } A_i \perp B_i \text{ b/c } \rho = 0 \\
&= \sigma_1 \sigma_2 \frac{1}{\sqrt{n}} \sum_i A_i B_i + o_p(1)
\end{aligned}$$

Notice that $E[A_i B_i] = E[A_i] E[B_i] = 0$ and

$$\begin{aligned}
V[A_i B_i] &= E[V[A_i B_i | B_i]] + V[E[A_i B_i | B_i]] \\
&= E\left[B_i^2 V[A_i]\right] + V[B_i E[A_i]] \\
&= E\left[B_i^2\right] = 1
\end{aligned}$$

By the CLT, $\sqrt{n} \left(\frac{1}{n} \sum_i A_i B_i - 0 \right) \to_d \mathcal{N}(0,1)$. By CMT,

$$\begin{aligned}
\frac{1}{n} \sum_i (x_i - \bar{x})^2 &\to_p \sigma_1^2 \Rightarrow \frac{1}{\sqrt{\frac{1}{n} \sum_i (x_i - \bar{x})^2}} \to_p \frac{1}{\sigma_1} \\
\frac{1}{n} \sum_i (y_i - \bar{y})^2 &\to_p \sigma_2^2 \Rightarrow \frac{1}{\sqrt{\frac{1}{n} \sum_i (y_i - \bar{y})^2}} \to_p \frac{1}{\sigma_2} \\
\Rightarrow \frac{1}{\sqrt{\frac{1}{n^2} \sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} &\to_p \frac{1}{\sigma_1 \sigma_2}
\end{aligned}$$

So by Slutsky's Theorem,

$$\begin{aligned}
\sqrt{n} R &= \frac{\frac{1}{\sqrt{n}} \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{\sqrt{n^2}} \sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \\
&= \frac{\frac{1}{\sqrt{n}} \sum_i (x_i - \mu_1)(y_i - \mu_2) + o_p(1)}{\sqrt{\frac{1}{n^2} \sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \\
&\to_d \mathcal{N}(0,1)
\end{aligned}$$

2.1.2 Question 2

2. Suppose that $X_1, \ldots, X_n$ are i.i.d from the uniform distribution $U(\theta, \theta+1)$, where θ is an unknown, finite, real-valued, scalar parameter.

   (a) Derive the maximum likelihood estimator (MLE) of θ.

   Solution

   The joint likelihood is

$$\begin{aligned}
P(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} P(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{(\theta+1)-\theta} \mathbf{1}\{\theta \leq x_i \leq \theta+1\} \\
&= \prod_{i=1}^{n} \mathbf{1}\{\theta \leq x_i \leq \theta+1\} = \prod_{i=1}^{n} \mathbf{1}\{\theta \leq x_i\} \mathbf{1}\{x_i \leq \theta+1\} \\
&= \prod_{i=1}^{n} \mathbf{1}\{\theta \leq x_i\} \cdot \prod_{i=1}^{n} \mathbf{1}\{x_i \leq \theta+1\} \\
&= \mathbf{1}\left\{\theta \leq x_{(1)}\right\} \cdot \mathbf{1}\left\{x_{(n)} \leq \theta+1\right\} \\
&= \mathbf{1}\left\{\theta \leq x_{(1)}\right\} \cdot \mathbf{1}\left\{x_{(n)} - 1 \leq \theta\right\} \\
&= \mathbf{1}\left\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\right\}
\end{aligned}$$

   And so we see that $X_{(n)} - 1 \leq \widehat{\theta}_{MLE} \leq X_{(1)}$ is the MLE of θ.

   (b) Consider estimating θ under absolute error loss, that is, assume the loss function is given by $L(\theta, a) = |\theta - a|$. Suppose that the prior for θ is given by $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$, where (μ_0, σ_0^2) are specified hyperparameters. Derive the Bayes estimator for θ.

   Solution

   To find the Bayes estimator, first find the posterior distribution. Let $\phi(x)$ and $\Phi(x)$ denote the probability density and cumulative density of $X \sim \mathcal{N}(0,1)$.

$$\begin{aligned}
P(\theta|\boldsymbol{x}) &= \frac{P(\boldsymbol{x}|\theta)P(\theta)}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|\theta)P(\theta)}{\int_\theta P(\boldsymbol{x}|\theta)P(\theta)\,d\theta} \\
&= \frac{\mathbf{1}\left\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\right\} \cdot \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right)}{\int_\theta \mathbf{1}\left\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\right\} \cdot \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right) d\theta} = \frac{\mathbf{1}\left\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\right\} \cdot \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right)}{\int_{x_{(n)}-1}^{x_{(1)}} \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right) d\theta} \\
&= \frac{\mathbf{1}\left\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\right\} \cdot \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right)}{\Phi\left(\frac{x_{(1)}-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right)}
\end{aligned}$$

   Recall that under absolute error loss, the Bayes estimator is the posterior median. Letting $d_\Lambda(\boldsymbol{x})$ denote the Bayes estimator, we have

$$\begin{aligned}
\frac{1}{2} = P(\theta \leq d_\Lambda(\boldsymbol{x})|\boldsymbol{x}) = \int_{-\infty}^{d_\Lambda(\boldsymbol{x})} P(\theta|\boldsymbol{x})\,d\theta &= \int_{-\infty}^{d_\Lambda(\boldsymbol{x})} \frac{\mathbf{1}\left\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\right\} \cdot \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right)}{\Phi\left(\frac{x_{(1)}-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right)} d\theta \\
&= \frac{\int_{x_{(n)}-1}^{d_\Lambda(\boldsymbol{x})} \phi\left(\frac{\theta-\mu_0}{\sigma_0}\right) d\theta}{\Phi\left(\frac{x_{(1)}-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right)} \\
\Leftrightarrow \frac{1}{2}\left[\Phi\left(\frac{x_{(1)}-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right)\right] &= \Phi\left(\frac{d_\Lambda(\boldsymbol{x})-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right) \\
\Leftrightarrow \frac{1}{2}\left[\Phi\left(\frac{x_{(1)}-\mu_0}{\sigma_0}\right) + \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right)\right] &= \Phi\left(\frac{d_\Lambda(\boldsymbol{x})-\mu_0}{\sigma_0}\right) \\
\Rightarrow d_\Lambda(\boldsymbol{x}) &= \mu_0 + \sigma_0 \cdot \Phi^{-1}\left(\frac{1}{2}\left[\Phi\left(\frac{x_{(1)}-\mu_0}{\sigma_0}\right) + \Phi\left(\frac{x_{(n)}-1-\mu_0}{\sigma_0}\right)\right]\right)
\end{aligned}$$

   We have our Bayes estimator.

(c) Under squared error loss, consider the class of estimators given by $d(X) = aX_{(1)} + bX_{(n)} + c$, where (a, b, c) are constants, $X = (X_1, \ldots, X_n)$, and $X_{(j)}$ is the jth order statistic. Within this class of estimators, derive an admissible estimator of θ.

Solution

To show that an estimator is admissible for θ within the specified class, we can use BIOS 761's **Theorem 1.14** stating that any unique Bayes estimator (with finite risk) is admissible. In other words, find a, b, c such that the frequentist risk is free of θ.

The frequentist risk is

$$\begin{aligned}
R(\theta, d(\boldsymbol{x})) &= E\left[L(\theta, d(\boldsymbol{x})\right] = E\left[(\theta - d(\boldsymbol{x}))^2\right] \\
&= V\left[\theta - d(\boldsymbol{x})\right] + \{E\left[\theta - d(\boldsymbol{x})\right]\}^2 \\
&= V\left[d(\boldsymbol{x})\right] + \{E\left[d(\boldsymbol{x})\right] - \theta\}^2 \\
&= V\left[ax_{(1)} + bx_{(n)} + c\right] + \left\{E\left[ax_{(1)} + bx_{(n)} + c\right] - \theta\right\}^2 \\
&= V\left[ax_{(1)} + bx_{(n)}\right] + \left\{E\left[ax_{(1)} + bx_{(n)}\right] + c - \theta\right\}^2 \\
&= V\left[a\left(x_{(1)} - \theta + \theta\right) + b\left(x_{(n)} - \theta + \theta\right)\right] + \left\{E\left[a\left(x_{(1)} - \theta + \theta\right) + b\left(x_{(n)} - \theta + \theta\right)\right] + c - \theta\right\}^2 \\
&= V\left[a\left(x_{(1)} - \theta\right) + b\left(x_{(n)} - \theta\right)\right] + \left\{E\left[a\left(x_{(1)} - \theta\right) + b\left(x_{(n)} - \theta\right)\right] + (a + b - 1)\theta + c\right\}^2 \\
& \text{Note: Transform } Z = X - \theta \Rightarrow Z \sim U(0,1) \\
&= V\left[az_{(1)} + bz_{(n)}\right] + \left\{aE\left[z_{(1)}\right] + bE\left[z_{(n)}\right] + (a + b - 1)\theta + c\right\}^2
\end{aligned}$$

(**) At this point, we don't actually need to calculate the expectations and variances because they're free of θ but I'll do so anyway. Notice that

$$\begin{aligned}
V\left[az_{(1)} + bz_{(n)}\right] &= V\left[\begin{bmatrix} a & b \end{bmatrix}\begin{bmatrix} z_{(1)} \\ z_{(n)} \end{bmatrix}\right] = \begin{bmatrix} a & b \end{bmatrix} V\begin{bmatrix} z_{(1)} \\ z_{(n)} \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} \\
&= \begin{bmatrix} a & b \end{bmatrix}\begin{bmatrix} V\left[z_{(1)}\right] & \text{Cov}(z_{(1)}, z_{(n)}) \\ \text{Cov}(z_{(1)}, z_{(n)}) & V\left[z_{(n)}\right] \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} \\
&= a^2 V\left[z_{(1)}\right] + b^2 V\left[z_{(n)}\right] + 2ab\text{Cov}(z_{(1)}, z_{(n)})
\end{aligned}$$

Since $Z \sim U(0,1)$, $f_{Z_{(1)}}(z) = n(1-z)^{n-1}$ and $f_{Z_{(n)}}(z) = nz^{n-1}$. Therefore

$$\begin{aligned}
E\left[z_{(1)}\right] &= \int_0^1 z f_{Z_{(1)}}(z)dz = \int_0^1 zn(1-z)^{n-1}dz = n\int_0^1 z^{2-1}(1-z)^{n-1}dz = n \cdot Beta(2, n) = n \cdot \frac{(2-1)!(n-1)!}{(2+n-1)!} \\
&= \frac{1}{n+1} \\
E\left[z_{(1)}^2\right] &= \int_0^1 z^2 f_{Z_{(1)}}(z)dz = \int_0^1 z^2 n(1-z)^{n-1}dz = n\int_0^1 z^{3-1}(1-z)^{n-1}dz = n \cdot Beta(3, n) = n \cdot \frac{(3-1)!(n-1)!}{(3+n-1)!} \\
&= \frac{2}{(n+2)(n+1)} \\
V\left[z_{(1)}\right] &= E\left[z_{(1)}^2\right] - \left\{E\left[z_{(1)}\right]\right\}^2 = \frac{2}{(n+2)(n+1)} - \frac{1}{(n+1)^2} = \frac{1}{n+1}\left(\frac{2}{n+2} - \frac{1}{n+1}\right) = \frac{1}{n+1} \cdot \frac{n}{(n+1)(n+2)} \\
&= \frac{n}{(n+1)^2(n+2)} \\
E\left[z_{(n)}\right] &= \int_0^1 z f_{Z_{(n)}}(z)dz = \int_0^1 znz^{n-1}dz = n\int_0^1 z^{n+1-1}(1-z)^{1-1}dz = n \cdot Beta(n+1, 1) = n \cdot \frac{(n+1-1)!(1-1)!}{(n+1+1-1)!} \\
&= \frac{n}{n+1} \\
E\left[z_{(n)}^2\right] &= \int_0^1 z^2 f_{Z_{(n)}}(z)dz = \int_0^1 z^2 nz^{n-1}dz = n\int_0^1 z^{n+2-1}(1-z)^{1-1}dz = n \cdot Beta(n+2, 1) \\
&= n \cdot \frac{(n+2-1)!(1-1)!}{(n+2+1-1)!} = \frac{n}{n+2} \\
V\left[z_{(n)}\right] &= E\left[z_{(n)}^2\right] - \left\{E\left[z_{(n)}\right]\right\}^2 = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} \\
&= \frac{n}{(n+1)^2(n+2)} = V\left[z_{(1)}\right]
\end{aligned}$$

For the covariance, we need the joint distribution or $f_{Z_{(1)},Z_{(n)}}(x,y) = \dfrac{n!}{(n-2)!}(y-x)^{n-2} \cdot \mathbf{1}\{0 \le x \le y \le 1\}$.

$$
\begin{aligned}
E\left[z_{(1)}z_{(n)}\right] &= \int_0^1 \int_0^y xy f_{Z_{(1)},Z_{(n)}}(x,y)dxdy = \frac{n!}{(n-2)!}\int_0^1\int_0^y xy(y-x)^{n-2}dxdy, \text{ let } r = \frac{x}{y} \\
&= \frac{n!}{(n-2)!}\int_0^1\int_0^1 ry^3(y-yr)^{n-2}drdy = \frac{n!}{(n-2)!}\int_0^1 y^{n+1}dy \cdot \int_0^1 r^{2-1}(1-r)^{n-1-1}dr \\
&= \frac{n!}{(n-2)!}\int_0^1 y^{n+1}dy \cdot Beta(2,n-1) = \frac{n!}{(n-2)!}\int_0^1 y^{n+1}dy \cdot \frac{(2-1)!(n-1-1)!}{(2+n-1-1)!} = \int_0^1 y^{n+1}dy \\
&= \frac{1}{n+2} \\
\text{Cov}(z_{(1)},z_{(n)}) &= E\left[z_{(1)}z_{(n)}\right] - E\left[z_{(1)}\right]E\left[z_{(n)}\right] = \frac{1}{n+2} - \frac{n}{(n+1)^2} \\
&= \frac{1}{(n+1)^2(n+2)}
\end{aligned}
$$

Hence the frequentist risk is

$$
\begin{aligned}
R(\theta, d(\boldsymbol{x})) &= a^2V\left[z_{(1)}\right] + b^2V\left[z_{(n)}\right] + 2ab\text{Cov}(z_{(1)},z_{(n)}) + \left\{aE\left[z_{(1)}\right] + bE\left[z_{(n)}\right] + (a+b-1)\theta + c\right\}^2 \\
&= \frac{n(a^2+b^2)}{(n+1)^2(n+2)} + \frac{2ab}{(n+1)^2(n+2)} + \left\{\frac{a}{n+1} + \frac{bn}{n+1} + (a+b-1)\theta + c\right\}^2 \\
&= \frac{n(a^2+b^2)+2ab}{(n+1)^2(n+2)} + \left\{\frac{a+bn}{n+1} + (a+b-1)\theta + c\right\}^2
\end{aligned}
$$

So we need for $a+b=1$ for the frequentist risk to be free of θ. We now have

$$
\begin{aligned}
R(\theta, d(\boldsymbol{x})) &= \frac{n(a^2+(1-a)^2)+2a(1-a)}{(n+1)^2(n+2)} + \left\{\frac{a+(1-a)n}{n+1} + c\right\}^2 \\
&= \frac{n}{(n+1)^2(n+2)}(a^2+(1-a)^2) + \frac{2}{(n+1)^2(n+2)}a(1-a) + \left\{\frac{1-n}{n+1}a + \frac{n}{n+1} + c\right\}^2 \\
&\equiv X(a^2+(1-a)^2) + Ya(1-a) + \left\{Za + \frac{n}{n+1} + c\right\}^2 \\
\frac{\partial R(\theta, d(\boldsymbol{x}))}{\partial a} &= X(2a-2(1-a)) + Y(1-2a) + 2Z\left\{Za + \frac{n}{n+1} + c\right\}^2 = 0 \\
&= 2X(2a-1) - Y(2a-1) + 2Z^2a + 2Z\frac{n}{n+1} + 2Zc = 0 \\
&\Rightarrow c = -\frac{(2X-Y)(2a-1)}{2Z} - Za - \frac{n}{n+1} \\
\frac{\partial R(\theta, d(\boldsymbol{x}))}{\partial c} &= 2\left\{Za + \frac{n}{n+1} + c\right\} = 0 \\
&\Rightarrow c = -Za - \frac{n}{n+1} \\
\Leftrightarrow -Za - \frac{n}{n+1} &= -\frac{(2X-Y)(2a-1)}{2Z} - Za - \frac{n}{n+1} \\
&\Leftrightarrow 0 = -\frac{(2X-Y)(2a-1)}{2Z} \\
&\Rightarrow a = \frac{1}{2} \text{ and } c = -\frac{1}{2}
\end{aligned}
$$

So the decision rule $d^*(\boldsymbol{x}) = \frac{1}{2}\left(X_{(1)} + X_{(n)} - 1\right)$ is admissible because under squared error loss (unique Bayes) the risk is finite for all θ.

(d) Under squared error loss, obtain a minimax estimator for θ.

Solution

The estimator from part (c) is minimax because it has finite risk and is minimized at $a = b = \frac{1}{2}$ and $c = -\frac{1}{2}$.

Might need to double check this.

(e) Derive the (appropriately normalized) asymptotic distribution of $R_n = X_{(n)} - X_{(1)}$.

Solution

Notice that $R_n = X_{(n)} - X_{(1)} = Z_{(n)} - Z_{(1)}$ where $Z \sim U(0,1)$. Using the joint distribution, we can transform. Let $R_n = Z_{(n)} - Z_{(1)}$ and $V = Z_{(1)}$.

$$
\begin{aligned}
f_{R_n,V}(r,v) &= f_{Z_{(1)},Z_{(n)}}(g_1^{-1}(r,v), g_2^{-1}(r,v)) \cdot \left|J\left(g_1^{-1}(r,v), g_2^{-1}(r,v)\right)\right| \\
&= n(n-1)r^{n-2}\mathbf{1}\left\{0 \le v \le 1-r \le 1\right\}\begin{vmatrix}1 & 0\\ 1 & 1\end{vmatrix} \\
&= n(n-1)r^{n-2}\mathbf{1}\left\{0 \le v \le 1-r \le 1\right\} \\
f_{R_n}(r) &= \int_0^1 f_{R_n,V}(r,v)dv \\
&= \int_0^1 n(n-1)r^{n-2}\mathbf{1}\left\{0 \le v \le 1-r \le 1\right\}dv \\
&= n(n-1)r^{n-2}\int_0^{1-r} dv \\
&= n(n-1)r^{n-2}(1-r)\mathbf{1}\left\{0 \le r \le 1\right\} \\
&= \frac{1}{Beta(n-1,2)} r^{n-1-1}(1-r)^{2-1}\mathbf{1}\left\{0 \le r \le 1\right\} \\
\Rightarrow R_n &\sim Beta(n-1,2) \\
F_{R_n}(r) = P\left(R_n \le r\right) &= \int_0^r f_{R_n}(u)du = \int_0^r n(n-1)u^{n-2}(1-u)du \\
&= n(n-1)\int_0^r u^{n-2} - u^{n-1}du = n(n-1)\left\{\left(\frac{u^{n-1}}{n-1} - \frac{u^n}{n}\right)\Bigg|_0^r\right\} \\
&= n(n-1)\left(\frac{r^{n-1}}{n-1} - \frac{r^n}{n}\right) = nr^{n-1} - (n-1)r^n \\
&= r^{n-1}\left(n(1-r) + r\right)
\end{aligned}
$$

Take the normalization $X_n \equiv n(1-R_n) \to_d X$ and find it's asymptotic distribution.

$$
\begin{aligned}
P\left(n(1-R_n) \le t\right) &= 1 - P\left(R_n \le 1 - \frac{t}{n}\right) \\
&= 1 - \left(1 - \frac{t}{n}\right)^{n-1}\left(t + 1 - \frac{t}{n}\right) \\
&\to 1 - e^{-t}(1+t) = F_X(t) = P\left(X \le t\right) \\
f_X(t) &= \frac{\partial}{\partial t}F_X(t) \\
&= -e^{-t} + (1+t)e^{-t} = te^{-t} \\
\Rightarrow n(1-R_n) &\to_d Gamma(2,1)
\end{aligned}
$$

2.1.3 Question 3

3. Let $X_1, \ldots, X_n$ be an i.i.d sample of real random variables with $E[X_1] = 0$ and $0 < \text{var}(X_1) = \sigma^2 < \infty$. Define $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Do the following:

(a) Show that for x close to zero, $e^x - 1 - x = x^2/2 + o(x^2)$.

Solution

Using a Maclaurin Series expansion a.k.a.

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(x)\big|_{x=0}\, x^j}{j!} = f(0) + f'(0)x + \frac{f''(0)x^2}{2!} + \cdots$$

of e^x centered around 0, we have

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots \\ & \quad \text{Note: Using big-O and little-o notation, } \frac{x^3}{3!} = o(x^2) \text{ where } x \to 0 \\ &= 1 + x + \frac{x^2}{2} + o(x^2) \\ \Rightarrow e^x - 1 - x &= \frac{x^2}{2} + o(x^2) \end{aligned}$$

(b) Show that $e^{\bar{X}_n} - 1 - \bar{X}_n \to 0$,

$$\frac{e^{\bar{X}_n} - \bar{X}_n - 1}{\bar{X}_n} \to 0 \text{ and } \frac{e^{\bar{X}_n} - \bar{X}_n - 1}{\bar{X}_n^2} \to \frac{1}{2}$$

in probability.

Solution

By WLLN, $\bar{X}_n \to_p 0$ hence $\bar{X}_n = o_p(1)$. Using part (a), we have

$$\begin{aligned} \frac{e^{\bar{X}_n} - \bar{X}_n - 1}{\bar{X}_n} &= \frac{\frac{\bar{X}_n^2}{2} + o_p(\bar{X}_n^2)}{\bar{X}_n} = \frac{\bar{X}_n}{2} + o_p(\bar{X}_n) \\ & \quad \text{Note: } o_p(\bar{X}_n) = o_p(1)o_p(1) = o_p(1) \\ &= o_p(1) + o_p(1) \\ &= o_p(1) \\ &\to_p 0 \\ \frac{e^{\bar{X}_n} - \bar{X}_n - 1}{\bar{X}_n^2} &= \frac{\frac{\bar{X}_n^2}{2} + o_p(\bar{X}_n^2)}{\bar{X}_n^2} = \frac{1}{2} + \frac{o_p(\bar{X}_n^2)}{\bar{X}_n^2} \\ &= \frac{1}{2} + o_p(1) \\ &\to_p \frac{1}{2} \end{aligned}$$

(c) Show that

$$\frac{2n}{S_n^2}\left(e^{\bar{X}_n} - 1 - \bar{X}_n\right)$$

converges in distribution to a χ^2 random variable with 1 degree of freedom.

Solution

Using the previous parts, Slutsky's Theorem, WLLN, and CMT,

$$\begin{aligned} \frac{2n}{S_n^2}\left(e^{\bar{X}_n} - 1 - \bar{X}_n\right) &= \frac{2n}{S_n^2}\left(\frac{\bar{X}_n^2}{2} + o_p(\bar{X}_n^2)\right) \\ &= \frac{n}{S_n^2}\left(\bar{X}_n^2 + o_p(1)\right) \\ &= \frac{1}{S_n^2}\left\{\sqrt{n}\left(\bar{X}_n - 0\right)\right\}^2 + o_p(1) \\ &\to_d \frac{1}{\sigma^2} \cdot \left\{\mathcal{N}(0, \sigma^2)\right\}^2 \\ &= \frac{1}{\sigma^2}\left\{\sigma \cdot \mathcal{N}(0,1)\right\}^2 \\ &= \chi_1^2 \end{aligned}$$

(d) Show that
$$\frac{2\sqrt{n}}{S_n}\left(\frac{e^{\bar{X}_n}-1-\bar{X}_n}{\bar{X}_n}\right)$$
converges in distribution to a $\mathcal{N}(0,1)$ random variable.

Solution

Using the previous parts, Slutsky's Theorem, WLLN, CMT,
$$\begin{aligned}\frac{2\sqrt{n}}{S_n}\left(\frac{e^{\bar{X}_n}-1-\bar{X}_n}{\bar{X}_n}\right) &= \frac{2\sqrt{n}}{S_n}\bar{X}_n\left(\frac{e^{\bar{X}_n}-1-\bar{X}_n}{\bar{X}_n^2}\right)\\ &= \frac{2\sqrt{n}}{S_n}\bar{X}_n\left(\frac{1}{2}+o_p(1)\right)\\ &= \frac{\sqrt{n}}{S_n}\bar{X}_n\left(1+o_p(1)\right)\\ &\to_d \mathcal{N}(0,1)\end{aligned}$$

(e) Show that
$$\frac{2n}{S_n^2}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\tan\bar{X}_n \to 0$$
in probability, where tan denotes the tangent function.

Solution

From part (c), we know that
$$\frac{2n}{S_n^2}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\to_d \chi_1^2$$
and by WLLN along with CMT,
$$\tan\bar{X}_n \to_p 0.$$
So by Slutsky's Theorem,
$$\frac{2n}{S_n^2}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\tan\bar{X}_n \to_d 0$$
and hence, by a theorem from BIOS 760,
$$\frac{2n}{S_n^2}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\tan\bar{X}_n \to_p 0$$

(f) Show that
$$\frac{2\sqrt{n}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\tan\bar{X}_n}{S_n\bar{X}_n^2}$$
converges in distribution to a $\mathcal{N}(0,1)$ random variable.

Solution

Using previous parts,
$$\begin{aligned}\frac{2\sqrt{n}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\tan\bar{X}_n}{S_n\bar{X}_n^2} &= \frac{2\sqrt{n}\tan\bar{X}_n}{S_n}\left(\frac{1}{2}+o_p(1)\right)\\ &= \frac{\sqrt{n}\tan\bar{X}_n}{S_n}\left(1+o_p(1)\right)\end{aligned}$$

We know that
$$\frac{\sqrt{n}(\bar{X}_n-0)}{S_n}\to_d \mathcal{N}(0,1)$$
Using the Delta Method, let $g(x)=\tan(x)$. Therefore $\nabla g(x)=\sec^2(x)$. So $\nabla g(0)=1$. And so we see that
$$\frac{2\sqrt{n}\left(e^{\bar{X}_n}-1-\bar{X}_n\right)\tan\bar{X}_n}{S_n\bar{X}_n^2}\to_d \mathcal{N}(0,1)$$

2 Part 2

2.2.1 Question 1

1. Consider independent observations $(X_1, Y_1), \cdots, (X_n, Y_n)$, where Y_i takes values 0 and 1. Suppose that $X_i|Y_i = m \sim \mathcal{N}(\mu_m, \sigma^2)$ and $P(Y_i = m) = \pi_m$ for $m = 0, 1$, where $\pi_0 + \pi_1 = 1$ and $\pi_0 \in (0, 1)$.

   (a) Show that $P(Y_i = m|X_i), m = 0, 1$ satisfies a logistic model, that is

   $$\text{logit}\left(P\left(Y_i = 1|X_i, \alpha\right)\right) = \alpha_0 + \alpha_1 X_i,$$

   where $\text{logit}(u) = \log(u/(1-u)), \alpha = (\alpha_0, \alpha_1)$, and α_0 and α_1 are unknown parameters. Derive the explicit form of $\alpha = g(\theta)$ as a function of $\theta = (\pi_1, \mu_0, \mu_1, \sigma^2)$.

   Solution

   Starting with the logistic model and substituting terms, we have

   $$\begin{aligned}
   \text{logit}\left(P\left(Y_i = 1|X_i, \alpha\right)\right) &= \log\left(\frac{P\left(Y_i = 1|X_i, \alpha\right)}{P\left(Y_i = 0|X_i, \alpha\right)}\right) = \log\left(\frac{P\left(X_i|Y_i = 1, \alpha\right) P\left(Y_i = 1\right) / P\left(X_i\right)}{P\left(X_i|Y_i = 0, \alpha\right) P\left(Y_i = 0\right) / P\left(X_i\right)}\right) \\
   &= \log\left(\frac{P\left(X_i|Y_i = 1, \alpha\right) P\left(Y_i = 1\right)}{P\left(X_i|Y_i = 0, \alpha\right) P\left(Y_i = 0\right)}\right) = \log\left(\frac{\pi_1}{\pi_0}\right) + \log\left(\frac{\phi\left(\frac{X_i - \mu_1}{\sigma}\right)}{\phi\left(\frac{X_i - \mu_0}{\sigma}\right)}\right) \\
   &= \log\left(\frac{\pi_1}{\pi_0}\right) + \log\left(\frac{\exp\left\{-\frac{(X_i - \mu_1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(X_i - \mu_0)^2}{2\sigma^2}\right\}}\right) \\
   &= \log\left(\frac{\pi_1}{\pi_0}\right) + \frac{(X_i - \mu_0)^2}{2\sigma^2} - \frac{(X_i - \mu_1)^2}{2\sigma^2} \\
   &= \log\left(\frac{\pi_1}{\pi_0}\right) + \frac{(X_i - \mu_o - (X_i - \mu_1))(X_i - \mu_0 + X_i - \mu_1)}{2\sigma^2} \\
   &= \log\left(\frac{\pi_1}{\pi_0}\right) + \frac{-(\mu_o - \mu_1)(2X_i - \mu_0 - \mu_1)}{2\sigma^2} \\
   &= \log\left(\frac{\pi_1}{\pi_0}\right) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \frac{(\mu_1 - \mu_0)}{\sigma^2} X_i \\
   &= \alpha_0 + \alpha_1 X_i
   \end{aligned}$$

   And so,

   $$\alpha = (\alpha_0, \alpha_1) = g(\theta) = \left(\log(\pi_1) - \log(\pi_0) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \frac{\mu_1 - \mu_0}{\sigma^2}\right)$$

(b) Based on the logistic model in (a), please give the explicit form of the Newton-Raphson algorithm for calculating the maximum likelihood estimate of α, denoted by $\widehat{\alpha} = (\widehat{\alpha}_0, \widehat{\alpha}_1)$, and derive the asymptotic covariance matrix of $\widehat{\alpha}$.

Solution

The joint likelihood and log likelihood are as follows.

$$\begin{aligned} P\left(\boldsymbol{y}|\boldsymbol{x};\alpha\right) &= \prod_{i=1}^{n} P\left(y_i|x_i;\alpha\right) = \prod_{i=1}^{n} \frac{\exp\left\{y_i(\alpha_0+\alpha_1 x_i)\right\}}{1+\exp\left\{\alpha_0+\alpha_1 x_i\right\}} \\ l_n(\alpha) \equiv \log\left\{P\left(\boldsymbol{y}|\boldsymbol{x};\alpha\right)\right\} &= \sum_{i=1}^{n} y_i(\alpha_0+\alpha_1 x_i) - \log\left(1+\exp\left\{\alpha_0+\alpha_1 x_i\right\}\right) \end{aligned}$$

The Newton-Raphson algorithm involves taking the first and second derivative of the log likelihood. From a Taylor series expansion of the score equation, it has the form

$$\alpha_{(t+1)} \approx \alpha_{(t)} - \left[\ddot{l}_n(\alpha_{(t)})\right]^{-1}\left[\dot{l}_n(\alpha_{(t)})\right]$$

But first, let's denote $\alpha = \begin{bmatrix}\alpha_0 \\ \alpha_1\end{bmatrix}$ and $z_i = \begin{bmatrix}1 \\ x_i\end{bmatrix}$.

$$\begin{aligned} l_n(\alpha) &= \sum_{i=1}^{n} y_i(\alpha^T z_i) - \log\left(1+\exp\left\{\alpha^T z_i\right\}\right) \\ \dot{l}_n(\alpha) \equiv \frac{\partial l_n(\alpha)}{\partial \alpha} &= \sum_{i=1}^{n} y_i z_i - \frac{\exp\left\{\alpha^T z_i\right\}}{1+\exp\left\{\alpha^T z_i\right\}} z_i \\ &= \sum_{i=1}^{n} \left(y_i - \frac{\exp\left\{\alpha^T z_i\right\}}{1+\exp\left\{\alpha^T z_i\right\}}\right) z_i \\ \ddot{l}_n(\alpha) \equiv \frac{\partial^2 l_n(\alpha)}{\partial \alpha \partial \alpha^T} &= -\sum_{i=1}^{n} \frac{\left(1+\exp\left\{\alpha^T z_i\right\}\right)\exp\left\{\alpha^T z_i\right\} - \exp\left\{\alpha^T z_i\right\}\exp\left\{\alpha^T z_i\right\}}{\left(1+\exp\left\{\alpha^T z_i\right\}\right)^2} z_i z_i^T \\ &= -\sum_{i=1}^{n} \frac{\exp\left\{\alpha^T z_i\right\}}{\left(1+\exp\left\{\alpha^T z_i\right\}\right)^2} z_i z_i^T \end{aligned}$$

Therefore the Newton-Raphson algorithm is

$$\alpha_{(t+1)} \approx \alpha_{(t)} + \left[\sum_{i=1}^{n} \frac{\exp\left\{\alpha_{(t)}^T z_i\right\}}{\left(1+\exp\left\{\alpha_{(t)}^T z_i\right\}\right)^2} z_i z_i^T\right]^{-1} \left[\sum_{i=1}^{n}\left(y_i - \frac{\exp\left\{\alpha_{(t)}^T z_i\right\}}{1+\exp\left\{\alpha_{(t)}^T z_i\right\}}\right) z_i\right]$$

Assuming the regularity conditions hold, by MLE theory, we know that

$$\sqrt{n}\left(\widehat{\alpha}-\alpha\right) \to_d \mathcal{N}\left(\mathbf{0}, I^{-1}(\alpha)\right)$$

where $I^{-1}(\alpha)$ is the asymptotic covariance of $\widehat{\alpha}$, $\dfrac{1}{n} I_n(\alpha) \to_p I(\alpha)$, and $I_n(\alpha) = E\left[-\ddot{l}_n(\alpha)\right] = \sum_{i=1}^{n} \dfrac{\exp\left\{\alpha^T z_i\right\}}{\left(1+\exp\left\{\alpha^T z_i\right\}\right)^2} z_i z_i^T$.

(c) Please write down the joint distribution of $\{(X_i, Y_i) : i = 1, \cdots, n\}$ and calculate the maximum likelihood estimate of θ, denoted by $\widehat{\theta}_F$, and its asymptotic covariance matrix.

Solution

The joint likelihood and log likelihood are as follows.

$$\begin{aligned}
P(\boldsymbol{x}, \boldsymbol{y}|\theta) &= \prod_{i=1}^n P(x_i, y_i|\theta) = \prod_{i=1}^n [P(x_i, y_i = 1|\theta)]^{y_i} [P(x_i, y_i = 0|\theta)]^{1-y_i} \\
&= \prod_{i=1}^n [P(x_i|y_i = 1; \theta) P(y_i = 1|\theta)]^{y_i} [P(x_i|y_i = 0; \theta) P(y_i = 0|\theta)]^{1-y_i} \\
&= \prod_{i=1}^n \left[\phi\left(\frac{x_i - \mu_1}{\sigma}\right)\pi_1\right]^{y_i} \left[\phi\left(\frac{x_i - \mu_0}{\sigma}\right)(1 - \pi_1)\right]^{1-y_i} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n y_i(x_i - \mu_1)^2}{2\sigma^2} - \frac{\sum_{i=1}^n (1 - y_i)(x_i - \mu_0)^2}{2\sigma^2}\right\} \prod_{i=1}^n \pi_1^{y_i}(1 - \pi_1)^{1-y_i} \\
l_n(\theta) \equiv \log(P(\boldsymbol{x}, \boldsymbol{y}|\theta)) &\propto -\frac{n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n y_i(x_i - \mu_1)^2}{2\sigma^2} - \frac{\sum_{i=1}^n (1 - y_i)(x_i - \mu_0)^2}{2\sigma^2} + \\
&\quad \sum_{i=1}^n y_i \log(\pi_1) + (1 - y_i)\log(1 - \pi_1)
\end{aligned}$$

Maximize the log likelihood to obtain $\widehat{\theta}_F$.

$$\begin{aligned}
\frac{\partial l_n(\theta)}{\partial \pi_1} &= \sum_{i=1}^n \frac{y_i}{\pi_1} - \frac{1 - y_i}{1 - \pi_1} = 0 \Rightarrow \widehat{\pi}_1 = \frac{\sum_i y_i}{n} = \bar{y} \\
\frac{\partial l_n(\theta)}{\partial \mu_0} &= \frac{1}{\sigma^2}\sum_i (1 - y_i)(x_i - \mu_0) = 0 \Rightarrow \widehat{\mu}_0 = \frac{\sum_i (1 - y_i)x_i}{\sum_i (1 - y_i)} \\
\frac{\partial l_n(\theta)}{\partial \mu_1} &= \frac{1}{\sigma^2}\sum_i y_i(x_i - \mu_1) = 0 \Rightarrow \widehat{\mu}_1 = \frac{\sum_i y_i x_i}{\sum_i y_i} \\
\frac{\partial l_n(\theta)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\sum_i y_i(x_i - \mu_1)^2}{2\sigma^4} + \frac{\sum_i (1 - y_i)(x_i - \mu_0)^2}{2\sigma^4} = 0 \\
&\Rightarrow \widehat{\sigma}^2 = \frac{1}{n}\sum_i \left\{y_i(x_i - \widehat{\mu}_1)^2 + (1 - y_i)(x_i - \widehat{\mu}_0)^2\right\}
\end{aligned}$$

For the asymptotic covariance matrix,

$$\begin{aligned}
\frac{\partial^2 l_n(\theta)}{\partial \pi^2} &= \sum_i -\frac{y_i}{\pi_1^2} - \frac{1 - y_i}{(1 - \pi_1)^2} \Rightarrow E\left[-\frac{\partial^2 l_n(\theta)}{\partial \pi^2}\right] = \sum_i \frac{1}{\pi_1} - \frac{1}{1 - \pi_1} = \frac{n}{\pi_1(1 - \pi_1)} \\
E\left[-\frac{\partial^2 l_n(\theta)}{\partial \pi \partial \mu_1}\right] &= E\left[-\frac{\partial^2 l_n(\theta)}{\partial \pi \partial \mu_0}\right] = E\left[-\frac{\partial^2 l_n(\theta)}{\partial \pi \partial \sigma^2}\right] = 0 \\
\frac{\partial^2 l_n(\theta)}{\partial \mu_0^2} &= -\frac{\sum_i (1 - y_i)}{\sigma^2} \Rightarrow E\left[-\frac{\partial^2 l_n(\theta)}{\partial \mu_0^2}\right] = \frac{n(1 - \pi_1)}{\sigma^2} \\
E\left[-\frac{\partial^2 l_n(\theta)}{\partial \mu_0 \partial \mu_1}\right] &= 0 \\
\frac{\partial^2 l_n(\theta)}{\partial \mu_0 \partial \sigma^2} &= -\frac{\sum_i (1 - y_i)(x_i - \mu_0)}{\sigma^4} \Rightarrow E\left[-\frac{\partial^2 l_n(\theta)}{\partial \mu_0 \partial \sigma^2}\right] = 0 \\
\frac{\partial^2 l_n(\theta)}{\partial \mu_1^2} &= -\frac{\sum_i y_i}{\sigma^2} \Rightarrow E\left[-\frac{\partial^2 l_n(\theta)}{\partial \mu_1^2}\right] = \frac{n\pi_1}{\sigma^2} \\
\frac{\partial^2 l_n(\theta)}{\partial \mu_1 \partial \sigma^2} &= -\frac{\sum_i y_i(x_i - \mu_1)}{\sigma^4} \Rightarrow E\left[-\frac{\partial^2 l_n(\theta)}{\partial \mu_1 \partial \sigma^2}\right] = 0 \\
\frac{\partial^2 l_n(\theta)}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{\sum_i y_i(x_i - \mu_1)^2}{\sigma^6} - \frac{\sum_i (1 - y_i)(x_i - \mu_0)^2}{\sigma^6} \Rightarrow E\left[-\frac{\partial^2 l_n(\theta)}{\partial (\sigma^2)^2}\right] = \frac{n}{2\sigma^4}
\end{aligned}$$

Note: The expectations involving y_i and x_i jointly utilized conditional expectation.

So we have

$$I_n(\theta) = n \cdot \begin{bmatrix} \frac{1}{\pi_1(1-\pi_1)} & 0 & 0 & 0 \\ 0 & \frac{1-\pi_1}{\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{\pi_1}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma^4} \end{bmatrix} \text{ and } \frac{1}{n}I_n(\theta) \to_p I(\theta) = \begin{bmatrix} \frac{1}{\pi_1(1-\pi_1)} & 0 & 0 & 0 \\ 0 & \frac{1-\pi_1}{\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{\pi_1}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

Using MLE theory, we have that $\sqrt{n}\left(\widehat{\theta}_F - \theta\right) \to_d \mathcal{N}\left(0, I^{-1}(\theta)\right)$ where

$$I^{-1}(\theta) = \begin{bmatrix} \pi_1(1-\pi_1) & 0 & 0 & 0 \\ 0 & \dfrac{\sigma^2}{1-\pi_1} & 0 & 0 \\ 0 & 0 & \dfrac{\sigma^2}{\pi_1} & 0 \\ 0 & 0 & 0 & 2\sigma^4 \end{bmatrix}$$

(d) Calculate the asymptotic covariance matrix of $g(\widehat{\theta}_F)$.

Solution

From part(a), $g(\theta) = \left(\log(\pi_1) - \log(1-\pi_1) + \dfrac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \dfrac{\mu_1 - \mu_0}{\sigma^2}\right)$ and so

$$\nabla g(\theta) = \begin{bmatrix} \dfrac{1}{\pi_1} + \dfrac{1}{1-\pi_1} & \dfrac{\mu_0}{\sigma^2} & -\dfrac{\mu_1}{\sigma^2} & -\dfrac{\mu_0^2 - \mu_1^2}{2\sigma^4} \\ 0 & -\dfrac{1}{\sigma^2} & \dfrac{1}{\sigma^2} & -\dfrac{\mu_1 - \mu_0}{\sigma^4} \end{bmatrix}$$

By the Delta Method, $\sqrt{n}\left(g(\widehat{\theta}_F) - g(\theta)\right) \to_d \nabla g(\theta) \cdot \mathcal{N}\left(0, I^{-1}(\theta)\right)$. The asymptotic covariance is

$$\begin{aligned}
\nabla g(\theta) \cdot I^{-1}(\theta) \cdot \nabla g(\theta)^T &= \begin{bmatrix} \frac{1}{\pi_1} + \frac{1}{1-\pi_1} & \frac{\mu_0}{\sigma^2} & -\frac{\mu_1}{\sigma^2} & -\frac{\mu_0^2-\mu_1^2}{2\sigma^4} \\ 0 & -\frac{1}{\sigma^2} & \frac{1}{\sigma^2} & -\frac{\mu_1-\mu_0}{\sigma^4} \end{bmatrix} \cdot \begin{bmatrix} \pi_1(1-\pi_1) & 0 & 0 & 0 \\ 0 & \frac{\sigma^2}{1-\pi_1} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{\pi_1} & 0 \\ 0 & 0 & 0 & 2\sigma^4 \end{bmatrix} \cdot \nabla g(\theta)^T \\
&= \begin{bmatrix} 1 & \frac{\mu_0}{1-\pi_1} & -\frac{\mu_1}{\pi_1} & \mu_1^2 - \mu_0^2 \\ 0 & -\frac{1}{1-\pi_1} & \frac{1}{\pi_1} & 2(\mu_0 - \mu_1) \end{bmatrix} \begin{bmatrix} \frac{1}{\pi_1(1-\pi_1)} & 0 \\ \frac{\mu_0}{\sigma^2} & -\frac{1}{\sigma^2} \\ -\frac{\mu_1}{\sigma^2} & \frac{1}{\sigma^2} \\ -\frac{\mu_0^2-\mu_1^2}{2\sigma^4} & -\frac{\mu_1-\mu_0}{\sigma^4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\pi_1(1-\pi_1)} + \frac{\mu_0^2}{\sigma^2(1-\pi_1)} + \frac{\mu_1^2}{\sigma^2\pi_1} + \frac{(\mu_1^2-\mu_0^2)^2}{2\sigma^4} & -\frac{\mu_0}{\sigma^2(1-\pi_1)} - \frac{\mu_1}{\sigma^2\pi_1} - \frac{(\mu_1-\mu_0)^2(\mu_1+\mu_0)}{\sigma^4} \\ -\frac{\mu_0}{\sigma^2(1-\pi_1)} - \frac{\mu_1}{\sigma^2\pi_1} - \frac{(\mu_0-\mu_1)^2(\mu_0+\mu_1)}{\sigma^4} & \frac{1}{\sigma^2(1-\pi_1)} + \frac{1}{\sigma^2\pi_1} + \frac{2(\mu_0-\mu_1)^2}{\sigma^4} \end{bmatrix}
\end{aligned}$$

(e) In this part, suppose that $\mu_0 = \mu_1$. Show that $\mathrm{Cov}(\widehat{\alpha})^{-1}\mathrm{Cov}(g(\widehat{\theta}_F))$ converges to a matrix, which does not depend on θ. Please interpret the results.

Solution

From part(a), $\alpha = g(\theta) = \left(\log(\pi_1) - \log(1-\pi_1) + \dfrac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \dfrac{\mu_1 - \mu_0}{\sigma^2}\right)$. Now with $\mu_0 = \mu_1$, we see that

$$\alpha = g(\theta) = (\log(\pi_1) - \log(1-\pi_1), 0) \Rightarrow \alpha_0 = \log\left(\frac{\pi_1}{1-\pi_1}\right), \alpha_1 = 0$$

Let's find the asymptotic covariance of $\widehat{\alpha}_0$ and let $\mu \equiv \mu_0 = \mu_1$. Combined with part (a), we have that

$$\begin{aligned}
I_n(\alpha_0) &= \sum_i \frac{\exp(\alpha_0)}{(1+\exp(\alpha_0))^2} z_i z_i^T = \pi_1(1-\pi_1)\sum_i z_i z_i^T \\
&= \pi_1(1-\pi_1)\begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \\
\frac{1}{n} I_n(\alpha_0) &= \pi_1(1-\pi_1)\begin{bmatrix} 1 & \frac{\sum_i x_i}{n} \\ \frac{\sum_i x_i}{n} & \frac{\sum_i x_i^2}{n} \end{bmatrix} \to_p \pi_1(1-\pi_1)\begin{bmatrix} 1 & \mu \\ \mu & \sigma^2 \end{bmatrix} \\
&= I(\alpha_0) = \mathrm{Cov}(\widehat{\alpha}_0)^{-1}
\end{aligned}$$

Using part (d), we have

$$\begin{aligned}
\mathrm{Cov}(g(\widehat{\theta}_F)) &= I^{-1}(g(\theta)) = \nabla g(\theta) \cdot I^{-1}(\theta) \cdot \nabla g(\theta)^T = \begin{bmatrix} \frac{\sigma^2+\mu^2}{\sigma^2\pi_1(1-\pi_1)} & -\frac{\mu}{\sigma^2\pi_1(1-\pi_1)} \\ -\frac{\mu}{\sigma^2\pi_1(1-\pi_1)} & \frac{1}{\sigma^2\pi_1(1-\pi_1)} \end{bmatrix} \\
&= \frac{1}{\sigma^2\pi_1(1-\pi_1)}\begin{bmatrix} \sigma^2+\mu^2 & -\mu \\ -\mu & 1 \end{bmatrix}
\end{aligned}$$

With these two pieces, we'll see that $\mathrm{Cov}(\widehat{\alpha}_0)^{-1}\mathrm{Cov}(g(\widehat{\theta}_F)) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. So one sees that the covariances are asymptotically and equivalently efficient. Assuming $\mu_0 = \mu_1$ means that $X \perp Y$. So conditioning on X will give you the same resulting inferences as if you did not.

(f) Now, suppose that π_1 is known. Will the results in (b)-(e) be changed? Please explain. If so, then please derive the corresponding results and compare with those obtained above.

Solution

From part (a), we see that even if we know π_1, we still won't know α_0 or α_1. So the answer to part (b) will remain unchanged. For part (c), the score equation with respect to π_1 didn't depend on μ_0, μ_1, or σ^2. The only difference here is that $I_n(\theta)$ and $I(\theta)$ are 3×3 matrices. For part (d), the Delta Method involves $g(\cdot)$, a function of μ_0, μ_1, and σ^2. So the asymptotic covariance becomes

$$\nabla g(\theta) \cdot I^{-1}(\theta) \cdot \nabla g(\theta)^T = \begin{bmatrix} \frac{\mu_0^2}{\sigma^2(1-\pi_1)} + \frac{\mu_1^2}{\sigma^2\pi_1} + \frac{(\mu_1^2-\mu_0^2)^2}{2\sigma^4} & -\frac{\mu_0}{\sigma^2(1-\pi_1)} - \frac{\mu_1}{\sigma^2\pi_1} - \frac{(\mu_1-\mu_0)^2(\mu_1+\mu_0)}{\sigma^4} \\ -\frac{\mu_0}{\sigma^2(1-\pi_1)} - \frac{\mu_1}{\sigma^2\pi_1} - \frac{(\mu_0-\mu_1)^2(\mu_0+\mu_1)}{\sigma^4} & \frac{1}{\sigma^2(1-\pi_1)} + \frac{1}{\sigma^2\pi_1} + \frac{2(\mu_0-\mu_1)^2}{\sigma^4} \end{bmatrix}$$

where the first row first column is strictly larger than previously because we have more information about Y.

Regarding part (e), the term $\text{Cov}(\widehat{\alpha}_0)^{-1}$ remains unchanged but for the second term we have

$$\begin{aligned} \text{Cov}(g(\widehat{\theta}_F)) &= \begin{bmatrix} \frac{\mu^2}{\sigma^2\pi_1(1-\pi_1)} & -\frac{\mu}{\sigma^2\pi_1(1-\pi_1)} \\ -\frac{\mu}{\sigma^2\pi_1(1-\pi_1)} & \frac{1}{\sigma^2\pi_1(1-\pi_1)} \end{bmatrix} \\ &= \frac{1}{\sigma^2\pi_1(1-\pi_1)} \begin{bmatrix} \mu^2 & -\mu \\ -\mu & 1 \end{bmatrix} \end{aligned}$$

Now taking the product of the two matrices,

$$\begin{aligned} \text{Cov}(\widehat{\alpha}_0)^{-1} \cdot \text{Cov}(g(\widehat{\theta}_F)) &= \pi_1(1-\pi_1) \begin{bmatrix} 1 & \mu \\ \mu & \sigma^2 \end{bmatrix} \cdot \frac{1}{\sigma^2\pi_1(1-\pi_1)} \begin{bmatrix} \mu^2 & -\mu \\ -\mu & 1 \end{bmatrix} \\ &= \frac{\mu^2-\sigma^2}{\sigma^2} \begin{bmatrix} 0 & 0 \\ \mu & -1 \end{bmatrix} \end{aligned}$$

This result shows that estimating the intercept is infinitely more efficient with π_1 known while the efficiency in the slope is unchanged.

2.2.2 Question 2

2. Consider the following model:

$$Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots + X_{ip}\beta_p + U + \epsilon_i,$$

$i = 1, \ldots, n$, where $\beta_1, \ldots, \beta_p$ are unknown parameters, $Y = (Y_1, \ldots, Y_n)$ is the vector of responses, and $X_{ij}, i = 1, \ldots, n, j = 1, \ldots, p$, are fixed covariates. Assume that $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, $U \sim \mathcal{N}(\alpha, k\sigma^2)$, where α and $\sigma^2 > 0$ are unknown, $k > 0$ is known, and ϵ_i are independent of each other and of U. Assume further that the $(n \times p)$ matrix with entries $X_{ij} - \bar{X}_{\cdot j}$ has rank p, where $\bar{X}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$.

(a) Find the distribution of Y, and show that the variance-covariance matrix for Y is positive-definite. (Hint: For a constant c the inverse of a matrix $I + cJ$ is in the form of $I + dJ$ for certain constant d, where I is the $(n \times n)$ identity matrix and J is the $(n \times n)$ matrix with all entries equal to 1.)

Solution

Let's begin by defining a shorthand notation. $Y_i = X_i^T\beta + U + \epsilon_i$.

$$\begin{aligned}
\epsilon_i &\sim \mathcal{N}\left(0, \sigma^2\right) \text{ and } U \sim \mathcal{N}\left(\alpha, k\sigma^2\right) \\
U + \epsilon_i &\sim \mathcal{N}\left(\alpha, \sigma^2(1+k)\right) \\
Y_i|X_i &\sim \mathcal{N}\left(X_i^T\beta + \alpha, \sigma^2(1+k)\right) \equiv \mathcal{N}\left(Z_i^T\theta, \sigma^2(1+k)\right) \\
V\left[Y_i|X_i\right] &= \sigma^2(1+k) \\
\text{Cov}\left[Y_i|X_i, Y_j|X_j\right] &= \text{Cov}\left[X_i^T\beta + U + \epsilon_i, X_j^T\beta + U + \epsilon_j\right] \\
&= \text{Cov}\left[U, U\right] = V\left[U\right] \\
&= k\sigma^2 \\
\Rightarrow V\begin{bmatrix} Y_i|X_i \\ Y_j|X_j \end{bmatrix} &= \begin{bmatrix} \sigma^2(1+k) & \sigma^2 k \\ \sigma^2 k & \sigma^2(1+k) \end{bmatrix} = \sigma^2\left(I_2 + kJ_2J_2^T\right) \\
\Rightarrow V\left[Y|X\right] &= \sigma^2\left(I_n + kJ_nJ_n^T\right) \\
\Rightarrow E\left[Y|X\right] &= Z\theta
\end{aligned}$$

where I_n is an $n \times n$ identity matrix and J_n is an $n \times 1$ vector of ones.

Without using the hint, we can easily show that the variance-covariance matrix for Y is positive-definite. Recall that a square matrix $\boldsymbol{A}$ is positive-definite if for any nonzero vector $\boldsymbol{x}$, $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$. In this case, we know that $\sigma^2 > 0$ and $k > 0$ so

$$\begin{aligned}
x^T\left(I_n + kJ_n\right)x &= x^Tx + kx^TJ_nJ_n^Tx \\
&= \sum_i x_i^2 + k\left(\sum_i x_i\right)^2 \\
&> 0
\end{aligned}$$

We can conclude that $V\left[Y|X\right]$ is positive-definite (hence invertible).

(b) Show that $\beta_1, \ldots, \beta_p$ and α are estimable.

Solution

Recall that if some matrix X is full rank, any linear combination of β, hence $\Lambda^T\beta$, is estimable. Therefore we just need to show $\begin{bmatrix} J_n & X \end{bmatrix}$ has rank $p+1$. Define

$$X_* = \begin{bmatrix} J_n & X - \bar{X} \end{bmatrix} = \begin{bmatrix} J_n & (I_n - M_J)X \end{bmatrix} \text{ and } X_{**} = \begin{bmatrix} J_n & X \end{bmatrix}$$

where $X - \bar{X}$ represents the $n \times p$ matrix of centered covariates and $M_J = \frac{1}{n} J_n J_n^T$. Realize that X_* and X_{**} have the same rank because elementary row (column) operations do not change the row (column) rank of a matrix. From BIOS 762, we know that $\text{rank}(X) = \text{rank}(X^TX)$. So notice that

$$\begin{aligned}
\text{rank}(X_{**}) &= \text{rank}\left(X_{**}^T X_{**}\right) = \text{rank}\left(X_*^T X_*\right) \\
&= \text{rank}\left(\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & (X-\bar{X})^T(X-\bar{X}) \end{bmatrix}\right) \\
&= \text{rank}\left(\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & X^T(I_n - M_J)X \end{bmatrix}\right) \\
&= 1 + \text{rank}\left((X-\bar{X})^T(X-\bar{X})\right) \\
&= 1 + \text{rank}\left(X - \bar{X}\right) = 1 + p
\end{aligned}$$

The matrix $X - \bar{X}$ has rank p from the given information.

(c) Let $\theta = (\alpha, \beta_1, \ldots, \beta_p)^T$. Derive the maximum likelihood estimator for θ, denoted by $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T$. What is the distribution of $\widehat{\theta}$?

Solution

Define the model as $Y = Z\theta + e$ where $Z = \begin{bmatrix} J_n & X \end{bmatrix}$, $\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, and $e \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 V\right)$ where $V = I_n + kJ_nJ_n^T$ and is known (since k is known). Recall V is positive-definite hence invertible. The joint likelihood and log likelihood are

$$\begin{aligned}
P\left(Y \middle| Z; \theta, \sigma^2, k\right) &= (2\pi)^{-n/2} \left|\sigma^2 V\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(Y - Z\theta\right)^T \left(\sigma^2 V\right)^{-1} \left(Y - Z\theta\right)\right\} \\
l_n(\theta, \sigma^2) \equiv \log\left(P\left(Y \middle| Z; \theta, \sigma^2, k\right)\right) &\propto -\frac{1}{2}\log(\left|\sigma^2 V\right|) - \frac{1}{2\sigma^2}\left(Y - Z\theta\right)^T V^{-1} \left(Y - Z\theta\right)
\end{aligned}$$

Using the matrix cookbook, we can maximize the log likelihood with respect to θ and conclude that the MLE of θ is

$$\widehat{\theta} = \left(Z^T V^{-1} Z\right)^{-1} Z^T V^{-1} Y$$

To find the distribution of $\widehat{\theta}$, we have $Y \sim \mathcal{N}\left(Z\theta, \sigma^2 V\right) \Rightarrow \widehat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2 \left(Z^T V^{-1} Z\right)^{-1}\right)$.

(d) Let $\tilde{\theta} = \left(\tilde{\alpha}, \tilde{\beta}_1, \ldots, \tilde{\beta}_p\right)$ be the value of the vector θ minimizing the sum of squares

$$\sum_{i=1}^{n} \left[Y_i - (\alpha + X_{i1}\beta_1 + \cdots + X_{ip}\beta_p)\right]^2$$

Is it true that $\text{Var}(\widehat{\alpha}) < \text{Var}(\tilde{\alpha})$? Carefully justify your answer.

Solution

The least squares estimate of θ is $\tilde{\theta} = (Z^T Z)^{-1} Z^T Y$. So its variance is

$$\begin{aligned}
V\left[\tilde{\theta}\right] &= V\left[(Z^T Z)^{-1} Z^T Y\right] = (Z^T Z)^{-1} Z^T \cdot V[Y] \cdot Z(Z^T Z)^{-1} \\
&= (Z^T Z)^{-1} Z^T \cdot \sigma^2 V \cdot Z(Z^T Z)^{-1} \\
&= \sigma^2 (Z^T Z)^{-1} Z^T V Z (Z^T Z)^{-1} \\
\Rightarrow V[\tilde{\alpha}] &= V\left[C_1^T \tilde{\theta}\right] = C_1^T V\left[\tilde{\theta}\right] C_1 \\
&= \sigma^2 C_1^T (Z^T Z)^{-1} Z^T V Z (Z^T Z)^{-1} C_1
\end{aligned}$$

where C_1^T is a row vector of length $p+1$ with all elements equaling 0 except for the first equaling 1.

From part (c), we know the variance of $\widehat{\theta}$.

$$V\left[\widehat{\theta}\right] = \sigma^2 \left(Z^T V^{-1} Z\right)^{-1} \Rightarrow V[\widehat{\alpha}] = \sigma^2 C_1^T \left(Z^T V^{-1} Z\right)^{-1} C_1$$

- One approach:
  From BIOS 762, a corollary under the weighted least squares section states, suppose X is full rank p. Then the weighted least squares estimate of β is given by $\tilde{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$, and the ordinary least squares estimate of β is given by $\widehat{\beta} = (X^T X)^{-1} X^T Y$. Then $\tilde{\beta} = \widehat{\beta}$ if and only if $C(VX) = C(X)$.
  Seeing that this corollary applies to this problem, we'll see that $V\left[\tilde{\theta}\right] = V\left[\widehat{\theta}\right]$ proving that the claim is false. The variances for the two estimates of α are actually **equal**!
- Second approach:
  This approach utilizes a lemma of inverting a sum of matrices that applies to this problem. The lemma states that if matrices A and $A+B$ are invertible, and B has rank 1, then

$$(A+B)^{-1} = A^{-1} - \frac{1}{1 + \text{trace}\left[BA^{-1}\right]} A^{-1} B A^{-1}$$

  I'll denote $V = I_n + kJ_nJ_n^T$. Let's look at the difference in variances.

$$V[\tilde{\alpha}] - V[\widehat{\alpha}] = \sigma^2 C_1^T \left[(Z^T Z)^{-1} Z^T V Z (Z^T Z)^{-1} - \left(Z^T V^{-1} Z\right)^{-1}\right] C_1$$

  I'll expand the first inner term.

$$\begin{aligned}
(Z^T Z)^{-1} Z^T V Z (Z^T Z)^{-1} &= (Z^T Z)^{-1} Z^T \left(I_n + kJ_nJ_n^T\right) Z \\
&= (Z^T Z)^{-1} Z^T Z (Z^T Z)^{-1} + k(Z^T Z)^{-1} Z^T J_n J_n^T Z (Z^T Z)^{-1} \\
&= (Z^T Z)^{-1} + k(Z^T Z)^{-1} Z^T J_n J_n^T Z (Z^T Z)^{-1} \\
&\equiv W
\end{aligned}$$

  Looking at the second inner term, I'll first expand V^{-1} using the lemma.

$$\begin{aligned}
V^{-1} &= \left(I_n + kJ_nJ_n^T\right)^{-1} = I_n - \frac{1}{1 + \text{trace}\left[kJ_nJ_n^T I_n\right]} I_n \left(kJ_nJ_n^T\right) I_n \\
&= I_n - \frac{k}{1+kn} J_n J_n^T \\
&\equiv I_n - dJ_nJ_n^T
\end{aligned}$$

  Now to expanding the second inner term.

$$\begin{aligned}
\left(Z^T V^{-1} Z\right)^{-1} &= \left(Z^T \left(I_n - dJ_nJ_n^T\right) Z\right)^{-1} = \left(Z^T Z - dZ^T J_n J_n^T Z\right)^{-1} \\
&\quad \text{Note: Re-apply the lemma b/c the conditions are satisfied.} \\
&= (Z^T Z)^{-1} - \frac{1}{1 + \text{trace}\left[-dZ^T J_n J_n^T Z (Z^T Z)^{-1}\right]} (Z^T Z)^{-1} \left(-dZ^T J_n J_n^T Z\right) (Z^T Z)^{-1} \\
&= (Z^T Z)^{-1} + \frac{d}{1 - d \cdot \text{trace}\left[Z^T J_n J_n^T Z (Z^T Z)^{-1}\right]} (Z^T Z)^{-1} \left(Z^T J_n J_n^T Z\right) (Z^T Z)^{-1}
\end{aligned}$$

  Regarding the trace term,

$$\text{trace}\left[Z^T J_n J_n^T Z (Z^T Z)^{-1}\right] = \text{trace}\left[J_n^T Z (Z^T Z)^{-1} Z^T J_n\right] = \text{trace}\left[J_n^T J_n\right] = n$$

  Notice that $\dfrac{d}{1-dn} = k$. And so $\left(Z^T V^{-1} Z\right)^{-1} \equiv W$. And so we've just shown that $V\left[\widehat{\theta}\right] = V\left[\tilde{\theta}\right] \Rightarrow V[\widehat{\alpha}] = V[\tilde{\alpha}]$.

2.2.3 Question 3

3. To evaluate the diagnostic performance using two continuous biomarkers, we randomly select n diseased subjects and m non-diseased subjects. Let $X_1 = (X_{11}, X_{12})^T, \ldots, X_n = (X_{n1}, X_{n2})^T$ be these two measured biomarkers for the diseased subjects and $Y_1 = (Y_{11}, Y_{12})^T, \ldots, Y_m = (Y_{m1}, Y_{m2})^T$ be the same two measured biomarkers for the non-diseased subjects. We aim to find an optimal linear combination of these biomarkers to maximize some measure of the diagnostic perfomance. In particular, we need to find $\beta = (\beta_1, \beta_2)^T$ such that the area under the receiver operating characteristics curve, defined by $AUC(\beta) \equiv P\left(\beta^T \boldsymbol{X}_1 \geq \beta^T \boldsymbol{Y}_1\right)$, is maximized.

   Assume $X_1, \ldots, X_n$ are i.i.d from $MN(\mu_1, \Sigma)$ and $Y_1, \ldots, Y_m$ are i.i.d from $MN(\mu_2, \Sigma)$, where $\mu_1 = (\mu_{11}, \mu_{21})^T, \mu_2 = (\mu_{12}, \mu_{22})^T$, $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ are unknown parameters and Σ is a positive definite matrix. Moreover, assume $m = \tau n$ for a fixed constant $\tau > 0$.

   (a) Show $AUC(\beta) = \Phi\left(\beta^T(\mu_1 - \mu_2)/\sqrt{2\beta^T\Sigma\beta}\right)$, where $\Phi(x)$ is the cumulative distribution function of $\mathcal{N}(0,1)$.

   Solution

   Using what's given, we know that

$$\begin{aligned}
\boldsymbol{X}_1 &\sim \mathcal{N}(\mu_1, \Sigma) \Rightarrow \beta^T \boldsymbol{X}_1 \sim \mathcal{N}(\beta^T\mu_1, \beta^T\Sigma\beta) \\
\boldsymbol{Y}_1 &\sim \mathcal{N}(\mu_2, \Sigma) \Rightarrow \beta^T \boldsymbol{Y}_1 \sim \mathcal{N}(\beta^T\mu_2, \beta^T\Sigma\beta) \\
\Rightarrow \beta^T \boldsymbol{X}_1 - \beta^T \boldsymbol{Y}_1 &\sim \mathcal{N}(\beta^T(\mu_1 - \mu_2), 2\beta^T\Sigma\beta)
\end{aligned}$$

   And so

$$\begin{aligned}
AUC(\beta) &= P\left(\beta^T \boldsymbol{X}_1 \geq \beta^T \boldsymbol{Y}_1\right) \\
&= P\left(\beta^T \boldsymbol{X}_1 - \beta^T \boldsymbol{Y}_1 \geq 0\right) \\
&= P\left(\frac{\beta^T \boldsymbol{X}_1 - \beta^T \boldsymbol{Y}_1 - \beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}} \geq \frac{-\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) \\
&= P\left(\mathcal{N}(0,1) \geq \frac{-\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) = P\left(\mathcal{N}(0,1) \leq \frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right) \\
&= \Phi\left(\frac{\beta^T(\mu_1 - \mu_2)}{\sqrt{2\beta^T\Sigma\beta}}\right)
\end{aligned}$$

(b) Show that the maximum of $AUC(\beta)$, denote as $A^{optimal}$, is

$$\Phi\left(\left[(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)/2\right]^{1/2}\right).$$

Hint: The β maximizing $AUC(\beta)$ is unique up to some multiplicative scale.

Solution

Since $AUC(\beta)$ is the cumulative distribution function of the form $F_X(x)$, which is monotone increasing in x, we simply need to maximize the input $x(\beta) = \dfrac{\beta^T(\mu_1-\mu_2)}{\sqrt{2\beta^T\Sigma\beta}}$.

$$\begin{aligned}
\frac{\partial x(\beta)}{\partial \beta} &= \frac{\sqrt{2\beta^T\Sigma\beta}(\mu_1-\mu_2) - \beta^T(\mu_1-\mu_2)\cdot\frac{\partial\sqrt{2\beta^T\Sigma\beta}}{\partial\beta}}{2\beta^T\Sigma\beta} = 0 \\
&\Leftrightarrow \sqrt{\beta^T\Sigma\beta}(\mu_1-\mu_2) = \beta^T(\mu_1-\mu_2)\cdot\frac{\partial\sqrt{\beta^T\Sigma\beta}}{\partial\beta}
\end{aligned}$$

For the inner partial derivative,

$$\begin{aligned}
\frac{\partial\sqrt{\beta^T\Sigma\beta}}{\partial\beta} &= \frac{1}{2}\left(\beta^T\Sigma\beta\right)^{-1/2}\cdot\frac{\partial}{\partial\beta}\left[\beta^T\Sigma\beta\right] \\
&\quad \text{Note: From the matrix cookbook, } \frac{\partial \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}}{\partial\boldsymbol{x}} = (\boldsymbol{B}+\boldsymbol{B}^T)\boldsymbol{x} \\
&= \frac{1}{2}\left(\beta^T\Sigma\beta\right)^{-1/2}\cdot(2\Sigma\beta) \\
&= \frac{\Sigma\beta}{\sqrt{\beta^T\Sigma\beta}}
\end{aligned}$$

Plugging this back into the main expression,

$$\begin{aligned}
\sqrt{\beta^T\Sigma\beta}(\mu_1-\mu_2) &= \beta^T(\mu_1-\mu_2)\cdot\frac{\Sigma\beta}{\sqrt{\beta^T\Sigma\beta}} \\
&\Leftrightarrow \beta^T\Sigma\beta(\mu_1-\mu_2) = \beta^T(\mu_1-\mu_2)\Sigma\beta
\end{aligned}$$

Notice that for the dimensions in the equation directly above to match up, we would need

$$\beta^T\Sigma\beta = \beta^T(\mu_1-\mu_2) \text{ and } \mu_1-\mu_2 = \Sigma\beta$$

These equations point to $\widehat{\beta} = \Sigma^{-1}(\mu_1-\mu_2)$ which does satisfy both equations. We're told that Σ is positive definite and hence its inverse exists. Plugging this solution back into $x(\beta)$ gives us

$$\begin{aligned}
x(\widehat{\beta}) &= \frac{\widehat{\beta}^T(\mu_1-\mu_2)}{\sqrt{2\widehat{\beta}^T\Sigma\widehat{\beta}}} \\
&= \frac{(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)}{\sqrt{2(\mu_1-\mu_2)^T\Sigma^{-1}\Sigma\Sigma^{-1}(\mu_1-\mu_2)}} \\
&= \frac{(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)}{\sqrt{2(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)}} \\
&= \left[\frac{(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)}{2}\right]^{1/2}
\end{aligned}$$

And so

$$A^{optimal} = AUC(\widehat{\beta}) = \Phi\left(\left[\frac{(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)}{2}\right]^{1/2}\right)$$

(c) Calculate the maximum likelihood estimator for $A^{optimal}$ and denote it by $\widehat{A}$.

Solution

Denote $x_i = (x_{i1}, x_{i2})^T$ for $i = 1, \ldots, n$ and $y_j = (y_{j1}, y_{j2})^T$ for $j = 1, \ldots, m$. The joint likelihood and log likelihood of the observed data is

$$\begin{aligned}
P(X_1, Y_1|\mu_1, \mu_2, \Sigma) &= \prod_{i=1}^{n} P(x_i|\mu_1, \Sigma) \times \prod_{j=1}^{m} P(y_j|\mu_2, \Sigma) \\
&= \prod_{i=1}^{n} (2\pi)^{-1/2} |\Sigma|^{-1/2} \exp\left\{-\frac{(x_i-\mu_1)^T\Sigma^{-1}(x_i-\mu_1)}{2}\right\} \times \\
&\quad \prod_{j=1}^{m} (2\pi)^{-1/2} |\Sigma|^{-1/2} \exp\left\{-\frac{(y_j-\mu_2)^T\Sigma^{-1}(y_j-\mu_2)}{2}\right\} \\
&= (2\pi)^{-\frac{n+m}{2}} |\Sigma|^{-\frac{n+m}{2}} \exp\left\{-\frac{\sum_i (x_i-\mu_1)^T\Sigma^{-1}(x_i-\mu_1)}{2}\right\} \exp\left\{-\frac{\sum_j (y_j-\mu_2)^T\Sigma^{-1}(y_j-\mu_2)}{2}\right\} \\
&\propto |\Sigma|^{-\frac{n+m}{2}} \exp\left\{-\frac{\sum_i (x_i-\mu_1)^T\Sigma^{-1}(x_i-\mu_1)}{2}\right\} \exp\left\{-\frac{\sum_j (y_j-\mu_2)^T\Sigma^{-1}(y_j-\mu_2)}{2}\right\} \\
l_{n+m}(\mu_1, \mu_2, \Sigma) &\propto -\frac{n+m}{2}\log(|\Sigma|) - \frac{\sum_i (x_i-\mu_1)^T\Sigma^{-1}(x_i-\mu_1)}{2} - \frac{\sum_j (y_j-\mu_2)^T\Sigma^{-1}(y_j-\mu_2)}{2}
\end{aligned}$$

To maximize the log likelihood with respect to μ_1 and μ_2, we will see that $\widehat{\mu}_1 = \bar{x}$ and $\widehat{\mu}_2 = \bar{y}$. To maximize with respect to Σ, refer to the matrix cookbook.

$$\begin{aligned}
l_{n+m}(\widehat{\mu}_1, \widehat{\mu}_2, \Sigma) &\propto -\frac{n+m}{2}\log(|\Sigma|) - \frac{\sum_i (x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x})}{2} - \frac{\sum_j (y_j-\bar{y})^T\Sigma^{-1}(y_j-\bar{y})}{2} \\
&= -\frac{n+m}{2}\log(|\Sigma|) - \text{trace}\left[\frac{\sum_i (x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x})}{2} - \frac{\sum_j (y_j-\bar{y})^T\Sigma^{-1}(y_j-\bar{y})}{2}\right] \\
&= -\frac{n+m}{2}\log(|\Sigma|) - \frac{1}{2}\text{trace}\left[\Sigma^{-1}\sum_i (x_i-\bar{x})(x_i-\bar{x})^T + \Sigma^{-1}\sum_j (y_j-\bar{y})(y_j-\bar{y})^T\right] \\
&\equiv -\frac{n+m}{2}\log(|\Sigma|) - \frac{1}{2}\text{trace}\left[\Sigma^{-1}B + \Sigma^{-1}C\right] \\
&\equiv -\frac{n+m}{2}\log(|\Sigma|) - \frac{1}{2}\text{trace}\left[\Sigma^{-1}D\right] \\
\frac{\partial l_{n+m}(\widehat{\mu}_1, \widehat{\mu}_2, \Sigma)}{\partial \Sigma} &= -\frac{n+m}{2}\frac{\partial \log(|\Sigma|)}{\partial \Sigma} - \frac{1}{2}\frac{\partial \text{trace}\left[\Sigma^{-1}D\right]}{\partial \Sigma} = 0 \\
&\quad \text{Note: } \frac{\partial \log(|\boldsymbol{X}|)}{\partial \boldsymbol{X}} = \left(\boldsymbol{X}^T\right)^{-1} \\
&\quad \text{Note: } \frac{\partial \text{trace}\left[\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\right]}{\partial \boldsymbol{X}} = -\left(-\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{X}^{-1}\right)^T \\
0 &= -\frac{n+m}{2}\left(\Sigma^T\right)^{-1} - \frac{1}{2}\left[-\Sigma^{-1}D\Sigma^{-1}\right]^T \\
0 &= -(n+m)\Sigma^{-1} - \Sigma^{-1}D\Sigma^{-1} \\
&\Leftrightarrow (n+m)\Sigma^{-1} = \Sigma^{-1}D\Sigma^{-1} \\
&\Leftrightarrow \left((n+m)\boldsymbol{I} - \Sigma^{-1}D\right)\Sigma^{-1} = 0 \\
&\Rightarrow \widehat{\Sigma} = \frac{1}{n+m}D = \frac{1}{n+m}\left(\sum_i (x_i-\bar{x})(x_i-\bar{x})^T + \sum_j (y_j-\bar{y})(y_j-\bar{y})^T\right)
\end{aligned}$$

Since the MLE for $A^{optimal}$ is invariant, we have

$$\widehat{A} = \widehat{AUC(\widehat{\beta})} = \Phi\left(\left[\frac{(\widehat{\mu}_1-\widehat{\mu}_2)^T\widehat{\Sigma}^{-1}(\widehat{\mu}_1-\widehat{\mu}_2)}{2}\right]^{1/2}\right)$$

(d) Describe how you will obtain the asymptotic distribution of $\sqrt{n+m}\left(\widehat{A}-A^{optimal}\right)$ (I think the exam had a typo). You do not need to give the explicit expression of the asymptotic variance.

Solution

Let us denote the set of parameters by $\theta=(\mu_1,\mu_2,\Sigma)$. With the regularity conditions satisfied and by MLE theory,

$$\sqrt{n+m}\left(\widehat{\theta}-\theta\right)\to_d \mathcal{N}(\mathbf{0},I^{-1}(\theta))$$

where the Fisher Information is $I_{n+m}(\theta)=E\left[-\frac{\partial^2 l_{n+m}(\theta)}{\partial\theta\partial\theta^T}\right]$ and by WLLN we have $\frac{1}{n+m}I_{n+m}(\theta)\to_p I(\theta)$. The Fisher Information can be obtained from the log likelihood in part (c).

By the Delta Method, we see that

$$\begin{aligned}\sqrt{n+m}\left(\widehat{A}-A^{optimal}\right) &= \sqrt{n+m}\left(\Phi(h(\widehat{\theta}))-\Phi(h(\theta))\right)\\ &\to_d \nabla\Phi(h(\theta))\cdot\mathcal{N}(\mathbf{0},I^{-1}(\theta))\\ &= \mathcal{N}\left(0,\nabla\Phi(h(\theta))\cdot I^{-1}(\theta)\cdot(\nabla\Phi(h(\theta)))^T\right)\end{aligned}$$

With regards to $\nabla\Phi(h(\theta))$, denote $\Phi\left(h(\theta)\right)=\Phi\left(h(\mu_1,\mu_2,\Sigma)\right)$ where $h(\mu_1,\mu_2,\Sigma)=\left[\frac{(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)}{2}\right]^{1/2}$. We'll see that

$$\begin{aligned}\nabla\Phi(h(\theta)) &= \begin{bmatrix}\frac{\partial\Phi(h(\theta))}{\partial\mu_1} & \frac{\partial\Phi(h(\theta))}{\partial\mu_2} & \frac{\partial\Phi(h(\theta))}{\partial\Sigma}\end{bmatrix}\\ &= \begin{bmatrix}\phi(h(\theta))\frac{\partial h(\theta)}{\partial\mu_1} & \phi(h(\theta))\frac{\partial h(\theta)}{\partial\mu_2} & \phi(h(\theta))\frac{\partial h(\theta)}{\partial\Sigma}\end{bmatrix}\\ &= \phi(h(\theta))\begin{bmatrix}\frac{\partial h(\theta)}{\partial\mu_1} & \frac{\partial h(\theta)}{\partial\mu_2} & \frac{\partial h(\theta)}{\partial\Sigma}\end{bmatrix}\end{aligned}$$

where $\phi(z)$ is the pdf of Z if $Z\sim\mathcal{N}(0,1)$.

(e) To test whether the combination of the two biomarkers is useful for diagnosis, we formulate the hypothesis $H_0: A^{optimal}=1/2$ vs $H_1: A^{optimal}>1/2$ and reject H_0 when $\widehat{A}>c_{n+m}$ for some threshold value c_{n+m} (depending on $n+m$). (Again, I think the exam had a small typo.)

i. Determine c_{n+m} such that the type I error converges to a given level α, where c_{n+m} is a constant depending *only* on $n+m$ and α, that is, $\lim_{n+m\to\infty}P\left(\widehat{A}>c_{n+m}\middle|H_0\right)=\alpha$.

Solution

Let's denote the asymptotic variance (a scalar) in part (d) by S such that $\sqrt{n+m}\left(\widehat{A}-A^{optimal}\right)\to_d\mathcal{N}(0,S)$. We'll see that

$$\begin{aligned}P\left(\widehat{A}>c_{n+m}\middle|H_0\right) &= P\left(\widehat{A}-\frac{1}{2}>c_{n+m}-\frac{1}{2}\middle|H_0\right)=P\left(\frac{\sqrt{n+m}\left(\widehat{A}-\frac{1}{2}\right)}{\sqrt{S}}>\frac{\sqrt{n+m}\left(c_{n+m}-\frac{1}{2}\right)}{\sqrt{S}}\middle|H_0\right)\\ &\approx P\left(Z>\frac{\sqrt{n+m}\left(c_{n+m}-\frac{1}{2}\right)}{\sqrt{S}}\middle|H_0\right)=1-P\left(Z\le\frac{\sqrt{n+m}\left(c_{n+m}-\frac{1}{2}\right)}{\sqrt{S}}\middle|H_0\right)\\ \alpha &= 1-\Phi\left(\frac{\sqrt{n+m}\left(c_{n+m}-\frac{1}{2}\right)}{\sqrt{S}}\right)\\ &\Rightarrow\ c_{n+m}=\sqrt{\frac{S}{n+m}}\cdot\Phi^{-1}(1-\alpha)+\frac{1}{2}=z_{1-\alpha}\sqrt{\frac{S}{n+m}}+\frac{1}{2}\end{aligned}$$

One thing to notice is that under the null, if $A^{optimal}=1/2$, then $\Phi(h(\theta))=1/2$ means that $h(\theta)=0$ implying that H_0 can be restated as $H_0:\mu_1=\mu_2$.

ii. Calculate the asymptotic power of this test at a local alternative $H_1 : A^{optimal} = 1/2 + \delta/\sqrt{n+m}$ where δ is a fixed positive constant. (Again, I think there's a small typo, n should be $n+m$.)

Solution

The simple hypotheses are

$$H_0 : A^{optimal} = 1/2 \text{ vs. } H_1 : A^{optimal} = 1/2 + \delta/\sqrt{n+m}$$

Power is calculated as

$$\begin{aligned}
\text{Power} &= P\left(\widehat{A} > c_{n+m} \middle| H_1 \text{ is true}\right) \\
&= P\left(\widehat{A} - \frac{1}{2} - \frac{\delta}{\sqrt{n+m}} > c_{n+m} - \frac{1}{2} - \frac{\delta}{\sqrt{n+m}} \middle| H_1 \text{ is true}\right) \\
&= P\left(\frac{\sqrt{n+m}\left(\widehat{A} - \frac{1}{2} - \frac{\delta}{\sqrt{n+m}}\right)}{\sqrt{S}} > \frac{\sqrt{n+m}\left(c_{n+m} - \frac{1}{2} - \frac{\delta}{\sqrt{n+m}}\right)}{\sqrt{S}} \middle| H_1 \text{ is true}\right) \\
&\approx P\left(Z > \frac{\sqrt{n+m}\left(c_{n+m} - \frac{1}{2} - \frac{\delta}{\sqrt{n+m}}\right)}{\sqrt{S}}\right) \\
& \quad \text{Note: From part (e)(i), } c_{n+m} = z_{1-\alpha}\sqrt{\frac{S}{n+m}} + \frac{1}{2} \\
&= P\left(Z > \frac{z_{1-\alpha}\sqrt{S} - \delta}{\sqrt{S}}\right) \\
&= P\left(Z > z_{1-\alpha} - \frac{\delta}{\sqrt{S}}\right) \\
\Rightarrow \text{Power} &= 1 - \Phi^{-1}\left(z_{1-\alpha} - \frac{\delta}{\sqrt{S}}\right)
\end{aligned}$$

3 Theory 2011

3 Part 1

3.1.1 Question 1

1. Let $X_1, X_2, \ldots,$ be a sequence of i.i.d. real random variables with $E[X_1] = 0$. Let N be a Poisson random variable with parameter $\lambda \geq 0$ and independent of $X_1, X_2, \ldots,$. For each integer $m \geq 0$, let $\bar{X}_m = m^{-1}\sum_{i=1}^{m} X_i$, where we define $\bar{X}_0 = 0$.

   (a) Assume $\sigma^2 = E\left[X_1^2\right] < \infty$ and do the following:

   (i) Show that $\operatorname{var}\left(\bar{X}_N\right) \leq \sigma^2 \left[P\left(N < \lambda^{1/3}\right) + \dfrac{P\left(N \geq \lambda^{1/3}\right)}{\lambda^{1/3}}\right]$. (I think there was a typo, it should be $\lambda^{1/3}$ in the denominator.)

   Solution
   Starting with the variance, we have

$$
\begin{aligned}
V\left[\bar{X}_N\right] &= E\left[V\left[\bar{X}_N|N\right]\right] + V\left[E\left[\bar{X}_N|N\right]\right] \\
&= E\left[V\left[\frac{1}{N}\sum_i X_i \middle| N\right]\right] + V\left[E\left[\frac{1}{N}\sum_i X_i \middle| N\right]\right] \\
&= E\left[\frac{1}{N^2}V\left[\sum_i X_i \middle| N\right]\right] + V\left[\frac{1}{N}E\left[\sum_i X_i \middle| N\right]\right] \\
&= E\left[\frac{1}{N^2}NV\left[X_i\right]\right] + V\left[\frac{1}{N}NE\left[X_i\right]\right] \\
&= \sigma^2 E\left[\frac{1}{N}\right] \\
&= \sigma^2\left[E\left[\frac{\mathbf{1}\left\{N < \lambda^{1/3}\right\}}{N}\right] + E\left[\frac{\mathbf{1}\left\{N \geq \lambda^{1/3}\right\}}{N}\right]\right]
\end{aligned}
$$

   For the first term,

$$
E\left[\frac{\mathbf{1}\left\{N < \lambda^{1/3}\right\}}{N}\right] \leq E\left[\mathbf{1}\left\{N < \lambda^{1/3}\right\}\right] = P\left(\lambda < \lambda^{1/3}\right).
$$

   For the second term,

$$
E\left[\frac{\mathbf{1}\left\{N \geq \lambda^{1/3}\right\}}{N}\right] \leq E\left[\frac{\mathbf{1}\left\{N \geq \lambda^{1/3}\right\}}{\lambda^{1/3}}\right] = \frac{1}{\lambda^{1/3}}P\left(N \geq \lambda^{1/3}\right)
$$

   And so we've shown that

$$
\operatorname{var}\left(\bar{X}_N\right) \leq \sigma^2\left[P\left(N < \lambda^{1/3}\right) + \frac{P\left(N \geq \lambda^{1/3}\right)}{\lambda^{1/3}}\right]
$$

   (ii) Show that $P\left(N < \lambda^{1/3}\right) \to 0$, as $\lambda \to \infty$. Hint: Use Chebyshev's inequality.
   Solution
   We start with

$$
\begin{aligned}
P\left(N < \lambda^{1/3}\right) &= P\left(N - \lambda < \lambda^{1/3} - \lambda\right) = P\left(N - \lambda < \lambda^{1/3} - \lambda\right) \\
&= P\left(-(N-\lambda) > \lambda - \lambda^{1/3}\right) = P\left(-\frac{N-\lambda}{\lambda - \lambda^{1/3}} > 1\right) \\
&= E\left[\mathbf{1}\left\{-\frac{N-\lambda}{\lambda-\lambda^{1/3}} > 1\right\}\right] \\
&\quad \text{Note: } -\frac{N-\lambda}{\lambda-\lambda^{1/3}} \leq \left|\frac{N-\lambda}{\lambda-\lambda^{1/3}}\right| \\
&\leq E\left[\mathbf{1}\left\{\left|\frac{N-\lambda}{\lambda-\lambda^{1/3}}\right| > 1\right\}\right] = P\left(|N-\lambda| > \left|\lambda - \lambda^{1/3}\right|\right)
\end{aligned}
$$

   By Chebyshev's inequality, we know that $P\left(|X-\mu| > \epsilon\right) \leq \dfrac{V[X]}{\epsilon^2}$.
   Hence

$$
\begin{aligned}
P\left(N < \lambda^{1/3}\right) &\leq P\left(|N-\lambda| > \left|\lambda - \lambda^{1/3}\right|\right) \\
&\leq \frac{V[X]}{\left(\lambda-\lambda^{1/3}\right)^2} = \frac{\lambda}{\left(\lambda-\lambda^{1/3}\right)^2} \\
&\to 0 \text{ as } \lambda \to \infty
\end{aligned}
$$

(iii) Show that $\lim_{\lambda\to\infty} P\left(\left|\bar{X}_N\right| \geq \epsilon\right) = 0$ for every $\epsilon > 0$.

Solution

Notice that $E\left[\bar{X}_N\right] = E\left[E\left[\bar{X}_N|N\right]\right] = E\left[\frac{1}{N}NE\left[X_i\right]\right] = 0$. So starting with the probability in question,

$$\begin{aligned}
P\left(\left|\bar{X}_N\right| \geq \epsilon\right) &= P\left(\left|\bar{X}_N - 0\right| \geq \epsilon\right) \\
&\leq \frac{V\left[\bar{X}_N\right]}{\epsilon^2} \text{ (by Chebyshev's inequality)} \\
&\leq \frac{\sigma^2}{\epsilon^2}\left[P\left(N < \lambda^{1/3}\right) + \frac{P\left(N \geq \lambda^{1/3}\right)}{\lambda^{1/3}}\right] \\
&\leq \frac{\sigma^2}{\epsilon^2}\left[P\left(N < \lambda^{1/3}\right) + \frac{1}{\lambda^{1/3}}\right] \\
&\to 0 \text{ as } \lambda \to \infty
\end{aligned}$$

(b) Let $\psi(t)$ be the characteristic function of a standard normal random variable, and define $Z_m = m^{1/2}\bar{X}_m/\sigma$. Continue to assume $\sigma^2 < \infty$. Do the following:

i. Show that for any real t,

$$\left|E\left(e^{itZ_N}\right) - \psi(t)\right| \leq 2P\left(N < \lambda^{1/3}\right) + \max_{m\geq\lambda^{1/3}}\left|E\left(e^{itZ_m}\right) - \psi(t)\right|.$$

Solution

Starting with the LHS

$$\begin{aligned}
\left|E\left[e^{itZ_N}\right] - \psi(t)\right| &= \left|E\left[e^{itZ_N} - \psi(t)\right]\right| \\
&= \left|E\left[E\left[e^{itZ_N} - \psi(t)|N\right]\right]\right| \\
&= \left|E\left[e^{itZ_N} - \psi(t)\middle|N < \lambda^{1/3}\right]P\left(N < \lambda^{1/3}\right) + E\left[e^{itZ_N} - \psi(t)\middle|N \geq \lambda^{1/3}\right]P\left(N \geq \lambda^{1/3}\right)\right| \\
&\leq \left|E\left[e^{itZ_N} - \psi(t)\middle|N < \lambda^{1/3}\right]P\left(N < \lambda^{1/3}\right) + E\left[e^{itZ_N} - \psi(t)\middle|N \geq \lambda^{1/3}\right]\right| \\
&\quad \text{Note: By the Triangle inequality, } |a+b| \leq |a| + |b| \\
&\leq \left|E\left[e^{itZ_N} - \psi(t)\middle|N < \lambda^{1/3}\right]\right|P\left(N < \lambda^{1/3}\right) + \left|E\left[e^{itZ_N} - \psi(t)\middle|N \geq \lambda^{1/3}\right]\right|
\end{aligned}$$

For the first expectation, notice that $\left|E\left[e^{itZ_N} - \psi(t)\middle|N < \lambda^{1/3}\right]\right| \leq \left|E\left[e^{itZ_N}\right]\right| + |\psi(t)| \leq 1 + 1 = 2$.
For the second expectation, we see that

$$\left|E\left[e^{itZ_N} - \psi(t)\middle|N \geq \lambda^{1/3}\right]\right| \leq \max_{m\geq\lambda^{1/3}}\left|E\left[e^{itZ_m}\right] - \psi(t)\right|$$

And so we've shown that

$$\left|E\left(e^{itZ_N}\right) - \psi(t)\right| \leq 2P\left(N < \lambda^{1/3}\right) + \max_{m\geq\lambda^{1/3}}\left|E\left(e^{itZ_m}\right) - \psi(t)\right|$$

ii. Show that for any real t, $\left|E\left(e^{itZ_N}\right) - \psi(t)\right| \to 0$, as $\lambda \to \infty$.
Solution
From part (b)(i), we know that

$$\left|E\left(e^{itZ_N}\right) - \psi(t)\right| \le 2P\left(N < \lambda^{1/3}\right) + \max_{m \ge \lambda^{1/3}} \left|E\left(e^{itZ_m}\right) - \psi(t)\right|$$

From part (a)(ii), we know that

$$P\left(N < \lambda^{1/3}\right) \to 0 \text{ as } \lambda \to \infty$$

and so this can be re-expressed as such. For all $\epsilon > 0$, $\exists \lambda_0$ such that $\forall \lambda > \lambda_0$,

$$\left|P\left(N < \lambda^{1/3}\right) - 0\right| = P\left(N < \lambda^{1/3}\right) < \frac{\epsilon}{4} \Rightarrow 2P\left(N < \lambda^{1/3}\right) < \frac{\epsilon}{2}$$

So we only need show that the second term converges to 0.
Portmanteau's Theorem says that if $X_n \to_d X$, then for a <u>bounded</u> and <u>continuous</u> function $g(\cdot)$,

$$E\left[g(X_n)\right] \to E\left[g(X)\right]$$

In this case, $Z_m = \dfrac{\sqrt{m}\bar{X}_m}{\sigma} \to_d \mathcal{N}(0,1)$ when $m \to \infty$, which happens when $\lambda \to \infty$. Notice that if we let $g(x) = e^{itx}$, then it's continuous and bounded for any real x. And so $E\left[e^{itZ_m}\right] \to E\left[e^{itZ}\right] = \psi(t)$ is equivalent to stating that $\forall \epsilon > 0$, $\exists M \in \mathbb{R}$ such that $\forall m \ge M$,

$$\left|E\left[e^{itZ_m}\right] - \psi(t)\right| < \frac{\epsilon}{2}$$

And so if we pick $\lambda_M = \max(\lambda_0, M)$, then $\forall \lambda > \lambda_M$,

$$\left|E\left[e^{itZ_N}\right] - \psi(t)\right| \le 2P\left(N < \lambda^{1/3}\right) + \max_{m \ge \lambda^{1/3}} \left|E\left[e^{itZ_m}\right] - \psi(t)\right| \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which is equivalent to saying that

$$\left|E\left[e^{itZ_N}\right] - \psi(t)\right| \to 0$$

(c) Now do not assume $\sigma^2 < \infty$. Do the following:

(i) Show that for each $\epsilon > 0$,

$$P\left(\left|\bar{X}_N\right| \ge \epsilon\right) \le P\left(N < \lambda^{1/3}\right) + P\left(\max_{m \ge \lambda^{1/3}} \left|\bar{X}_M\right| \ge \epsilon\right)$$

Solution
Starting with the LHS,

$$\begin{aligned} P\left(\left|\bar{X}_N\right| \ge \epsilon\right) &= P\left(\left|\bar{X}_N\right| \ge \epsilon, N < \lambda^{1/3}\right) + P\left(\left|\bar{X}_N\right| \ge \epsilon, N \ge \lambda^{1/3}\right) \\ &\le P\left(N < \lambda^{1/3}\right) + P\left(\max_{m \ge \lambda^{1/3}} \left|\bar{X}_m\right| \ge \epsilon\right) \end{aligned}$$

(ii) Show that $\bar{X}_N \to 0$, in probability, as $\lambda \to \infty$. Hint: Use the strong law of large numbers.
Solution
We know that $P\left(N < \lambda^{1/3}\right) \to 0$ as $\lambda \to \infty$. This can be restated as such. $\forall \delta > 0$, $\exists \lambda_0$ such that $\forall \lambda > \lambda_0$,

$$P\left(N < \lambda^{1/3}\right) < \frac{\delta}{2}$$

For the second term from part (c)(i), by the SLLN,

$$\bar{X}_m \to_{a.s.} E\left[X_i\right] = 0$$

This can be restated as such. $\forall \epsilon > 0$, there exists an $\delta > 0$ such that for $m > m_0$

$$P\left(\sup_{m \ge m_0} \left|\bar{X}_m - 0\right| \ge \epsilon\right) < \frac{\delta}{2}$$

So we see that if we define $\lambda^* = max\left\{\lambda_0, m_0^3\right\}$, then $\forall \lambda > \lambda^*$ we have

$$P\left(\left|\bar{X}_N\right| \ge \epsilon\right) \le P\left(N < \lambda^{1/3}\right) + P\left(\max_{m \ge \lambda^{1/3}} \left|\bar{X}_M\right| \ge \epsilon\right) \le \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

This implies that $P\left(\bar{X}_N > \epsilon\right) \to 0$ when $\lambda \to \infty$, in other words, $\bar{X}_N \to_p 0$.

3.1.2 Question 2

2. (a) Let X be a random variable and let ν be a parameter of interest in the distribution of X. Suppose that $T(X)$ is an unbiased estimator of ν. Show that any unbiased estimator of ν is of the form $T(X) - U(X)$, where $E[U(X)] = 0$.

Solution

To prove that the form is unique, let $W(X)$ be any other unbiased estimator of ν.

Since $T(X)$ and $W(X)$ are unbiased, their expected difference will equal 0. So let $U(X) = T(X) - W(X)$.

$$E[U(X)] = E[T(X) - W(X)] = \nu - \nu = 0$$

This shows us that any unbiased estimator of ν will have the form $W(X) = T(X) - U(X)$.

In the sequel, let X be a discrete random variable with $P(X = -1) = p, P(X = k) = (1-p)^2 p^k, k = 0, 1, 2, \dots,$ where $p \in (0,1)$ is unknown.

(b) Show that $E[U(X)] = 0$ if and only if $U(k) = ak$ for all $k = -1, 0, 1, 2, \dots$ and some a.

Solution

I'll try to prove $\Rightarrow$ direction first. Using what we're given and the LHS, we have

$$\begin{aligned}
0 = E[U(X)] &= \sum_{k=-1}^{\infty} U(k) P(X = k) \\
&= U(-1)P(X = -1) + \sum_{k=0}^{\infty} U(k)P(X = k) \\
&= U(-1)p + \sum_{k=0}^{\infty} U(k)(1-p)^2 p^k \\
0 &= U(-1)p + (1-p)^2 \sum_{k=0}^{\infty} U(k)p^k \\
\Leftrightarrow \frac{-U(-1)p}{(1-p)^2} &= \sum_{k=0}^{\infty} U(k)p^k \\
&\quad \text{Note: Let } -U(-1) = a \\
\Leftrightarrow \frac{ap}{(1-p)^2} &= \sum_{k=0}^{\infty} U(k)p^k
\end{aligned}$$

At this point, consider the infinite sum $\sum_{k=0}^{\infty} kp^k$ which equals $\frac{p}{(1-p)^2}$ with some algebra and calculus. Therefore we have

$$\begin{aligned}
\sum_{k=0}^{\infty} akp^k &= \sum_{k=0}^{\infty} U(k)p^k \\
&\Rightarrow U(k) = ak \text{ for some } a \text{ and } k = -1, 0, 1, 2, \dots
\end{aligned}$$

To prove $\Leftarrow$, assume $U(k) = ak$ for $k = -1, 0, 1, 2, \dots$ and some a. Then

$$\begin{aligned}
E[U(X)] &= \sum_{k=-1}^{\infty} U(k)P(X = k) \\
&= U(-1)P(X = -1) + \sum_{k=0}^{\infty} U(k)P(X = k) \\
&= -ap + \sum_{k=0}^{\infty} ak(1-p)^2 p^k \\
&= -ap + a(1-p)^2 \sum_{k=0}^{\infty} kp^k \\
&= -ap + a(1-p)^2 \cdot \frac{p}{(1-p)^2} \\
&= 0
\end{aligned}$$

(c) Using the results in (a) and (b), show that $\mathbf{1}\{X=0\}$ is the unique admissible estimator under squared error loss amongst all unbiased estimators of $(1-p)^2$, where $\mathbf{1}\{\cdot\}$ is the indicator function.

Solution

Define

- $\nu = (1-p)^2$,
- $T(X) = \mathbf{1}\{X=0\}$,
- $W(X) = T(X) - aX$, and
- $L(x,y) \equiv (x-y)^2$.

We have

$$\begin{aligned} E[T(X)] &= E[\mathbf{1}\{X=0\}] \\ &= P(X=0) = (1-p)^2 \end{aligned}$$

Hence $T(X)$ is unbiased. From part (a), we know that the class of unbiased estimators takes the form given by $W(X)$ due to part (b).

$$\begin{aligned} E[W(X)] &= E[T(X) - aX] \\ &= (1-p)^2 - aE[X] \\ &= (1-p)^2 \end{aligned}$$

Let's compare the variances of the two estimators.

$$\begin{aligned} V[W(X)] &= V[T(X) - aX] \\ &= V[T(X)] + V[aX] - 2\mathrm{Cov}(T(X), aX) \\ &\ \text{Note: } \mathrm{Cov}(T(X), aX) = E[T(X)aX] = E[\mathbf{1}\{X=0\}aX] = 0 \\ &= V[T(X)] + a^2V[X] \\ &\geq V[T(X)] \ \text{ with equality if } a = 0 \end{aligned}$$

Hence $T(X) = \mathbf{1}\{X=0\}$ is the unique admissible estimator under squared error loss amongst all unbiased estimators of $(1-p)^2$.

(d) Show that no unique admissible estimator exists for p under squared error loss amongst unbiased estimators for p.

Solution

Let $d(X) = \mathbf{1}\{X=-1\}$. We see that $d(X)$ is unbiased.

$$\begin{aligned} E[d(X)] &= E[\mathbf{1}\{X=-1\}] \\ &= P(X=-1) \\ &= p \end{aligned}$$

Let $d_2(X) = d(X) - aX$. Also we see that $d_2(X)$ is unbiased.

$$\begin{aligned} E[d_2(X)] &= E[d(X) - aX] \\ &= P(X=-1) - aE[X] \\ &= p \end{aligned}$$

Looking at the variances.

$$\begin{aligned} V[d_2(X)] &= V[d(X) - aX] \\ &= V[d(X)] + V[aX] - 2\mathrm{Cov}(d(X), aX) \\ &\ \text{Note: } \mathrm{Cov}(d(X), aX) = E[d(X)aX] = E[\mathbf{1}\{X=-1\}aX] = -aP(X=-1) = -ap \\ &= V[d(X)] + a^2V[X] + 2ap \end{aligned}$$

To minimize this variance with respect to a, we see that $a = \dfrac{-p}{V[X]}$. To calculate the variance, consider

$$\begin{aligned} E[(X+2)(X+1)] &= \sum_{k=-1}^{\infty}(k+2)(k+1)P(X=k) = \sum_{k=0}^{\infty}(k+2)(k+1)P(X=k) \\ &= \sum_{k=0}^{\infty}(k+2)(k+1)(1-p)^2p^k = (1-p)^2\sum_{k=0}^{\infty}(k+2)(k+1)p^k \\ &= (1-p)^2\sum_{k=0}^{\infty}\frac{\partial^2}{\partial p^2}\left[p^{k+2}\right] = (1-p)^2\frac{\partial^2}{\partial p^2}\left[\sum_{k=0}^{\infty}p^{k+2}\right] \\ &= (1-p)^2\frac{\partial^2}{\partial p^2}\left[\frac{p^2}{1-p}\right] = \frac{2}{1-p} \\ \Leftrightarrow E\left[X^2\right] + 2 &= \frac{2}{1-p} \\ \Leftrightarrow V[X] &= \frac{2p}{1-p} \end{aligned}$$

So $a = \dfrac{-p}{V[X]} = \dfrac{-p}{\left(\dfrac{2p}{1-p}\right)} = \dfrac{p-1}{2} = -\dfrac{1-p}{2}$ will minimize the variance. But this means our unbiased estimator

$$d_2(X) = \mathbf{1}\{X = -1\} + \frac{1-p}{2}X$$

While this estimator has smallest variance, it's a function of the unknown parameter p. So NO unique admissible estimator exists for p under squared error loss.

(e) Prove whether there exist unbiased estimators of p^{-1}. If so, then determine whether a unique admissible estimator exists under squared error loss amongst unbiased estimators for p^{-1}.

Solution

Let's assume $T(X)$ is an unbiased estimator of p^{-1}. Therefore

$$\begin{aligned}
\frac{1}{p} &= E[T(X)] = \sum_{k=-1}^{\infty} T(k)P(X=k) \\
&= T(-1)p + \sum_{k=0}^{\infty} T(k)(1-p)^2 p^k \\
\Leftrightarrow \frac{1}{p(1-p)^2} - \frac{T(-1)p}{(1-p)^2} &= \sum_{k=0}^{\infty} T(k)p^k \\
&\quad \text{Note: Recall that } \sum_{k=0}^{\infty} kp^k = \frac{p}{(1-p)^2} \\
\Leftrightarrow \frac{1}{p^2}\sum_{k=0}^{\infty} kp^k - T(-1)\sum_{k=0}^{\infty} kp^k &= \sum_{k=0}^{\infty} T(k)p^k \\
\Leftrightarrow \sum_{k=0}^{\infty}\left[\frac{1}{p^2} - T(-1)\right] kp^k &= \sum_{k=0}^{\infty} T(k)p^k \\
&\Rightarrow T(X) = \left[\frac{1}{p^2} - T(-1)\right] X
\end{aligned}$$

Since $T(X)$ is a function of p, there doesn't exist an unbiased estimator of p^{-1}. Proof by contradiction complete.

3.1.3 Question 3

3. Consider a sequence of numbers $x_1, x_2, \ldots$ and place vertical lines before x_1 and between x_j and x_{j+1} whenever $x_j > x_{j+1}$. We say that the runs are the segments between pairs of lines. Thus, each run is an increasing segment of the sequence $x_1, x_2, \ldots$.

   Suppose that $X_1, X_2, \ldots$ are independent and identically distributed uniform$(0,1)$ random variables and that we are interested in the lengths of the successive runs. Let L_j denote the length of the jth run.

   (a) Compute $P(L_1 \geq m), m = 1, 2, \ldots$

   Solution

   Notice that the expression $P(L_1 \geq m)$ simply asks what is the probability that the first run is at least of length m. In other words what is the probability of observing

$$X_1 \leq X_2 \leq \cdots \leq X_m$$

   a monotone increasing segment of x_i's. The probability is as follows. We'll try to see a pattern.

$$\begin{aligned}
P(X_1 \leq X_2) &= \int_0^1 P(x_1 \leq X_2 | X_1 = x_1) P(X_1 = x_1) \, dx_1 \\
&= \int_0^1 \int_{x_1}^1 P(X_2 = x_2) \, dx_2 P(X_1 = x_1) \, dx_1 \\
&= \int_0^1 \int_{x_1}^1 P(X_2 = x_2) \, dx_2 dx_1 = \int_0^1 (1 - x_1) dx_1 \\
&= \frac{1}{2} = \frac{1}{2!} \\
P(X_1 \leq X_2 \leq X_3) &= \int_0^1 \int_{x_1}^1 \int_{x_2}^1 P(X_3 = x_3) \, dx_3 \prod_{i=1}^{2} P(X_i = x_i) \, dx_2 dx_1 \\
&= \int_0^1 \int_{x_1}^1 \int_{x_2}^1 P(X_3 = x_3) \, dx_3 dx_2 dx_1 \\
&= \int_0^1 \int_{x_1}^1 (1 - x_2) dx_2 dx_1 \\
&= \int_0^1 \left(x_2 - \frac{x_2^2}{2} \right)\Bigg|_{x_1}^1 dx_1 = \int_0^1 \left(\frac{1}{2} - \left(x_1 - \frac{x_1^2}{2} \right) \right) dx_1 \\
&= \int_0^1 \frac{1}{2} - x_1 + \frac{x_1^2}{2} dx_1 \\
&= \frac{1}{2} - \left(\frac{x_1^2}{2} \right)\Bigg|_0^1 + \left(\frac{x_1^3}{6} \right)\Bigg|_0^1 \\
&= \frac{1}{6} = \frac{1}{3!} \\
\Rightarrow P(X_1 \leq X_2 \leq \cdots \leq X_m) &\stackrel{?}{=} \frac{1}{m!} \text{ (maybe)}
\end{aligned}$$

   <u>Proof by Induction</u>

   - Base Cases: $P(X_1 \leq X_2) = \dfrac{1}{2}$ and $P(X_1 \leq X_2 \leq X_3) = \dfrac{1}{3!}$
   - Induction Hypothesis: Assume $P(X_1 \leq \cdots \leq X_i) = \dfrac{1}{i!}$
   - Prove that $P(X_1 \leq \cdots \leq X_i \leq X_{i+1}) = \dfrac{1}{(i+1)!}$.

     Notice that

$$\begin{aligned}
P(X_1 \leq \cdots \leq X_i \leq X_{i+1}) &= P(X_1 \leq \cdots \leq X_i, \max\{X_1, \ldots, X_i\} \leq X_{i+1}) \\
&= P(X_1 \leq \cdots \leq X_i) P\left(X_{(i)} \leq X_{i+1}\right) \\
&= \frac{1}{i!} \cdot P\left(X_{(i)} \leq X_{i+1}\right) = \frac{1}{i!} \cdot \int_0^1 P\left(X_{(i)} \leq z, X_{i+1} = z\right) dz \\
&= \frac{1}{i!} \cdot \int_0^1 P\left(X_{(i)} \leq z\right) P(X_{i+1} = z) \, dz = \frac{1}{i!} \cdot \int_0^1 z^i dz \\
&= \frac{1}{i!} \cdot \frac{1}{i+1} \\
&= \frac{1}{(i+1)!}
\end{aligned}$$

And so, through a proof by induction, we conclude that

$$P\left(L_1 \geq m\right) = P\left(X_1 \leq X_2 \leq \cdots \leq X_m\right) = \frac{1}{m!}$$

Another way to think about this is "what is the probability that a set of m objects are ordered from smallest to largest?"

(b) Suppose we know that the jth run starts with the value x. Compute $P(L_j \geq m|x)$.

Solution

Notice here that the expression $P(L_j \geq m|x)$ is equivalent to observing

$$x \leq X_{j+1} \leq \cdots \leq X_{j+m-1}$$

and so we have

$$\begin{aligned}
P(L_j \geq m|x) &= P(x \leq X_{j+1} \leq \cdots \leq X_{j+m-1}) \\
&= P(X_{j+1} \leq \cdots \leq X_{j+m-1}|x \leq X_{j+1}, \ldots, x \leq X_{j+m-1}) \, P(x \leq X_{j+1}, \ldots, x \leq X_{j+m-1}) \\
&= P(X_{j+1} \leq \cdots \leq X_{j+m-1}) \cdot \prod_{i=1}^{m-1} P(x \leq X_{j+i}) \\
&= \frac{1}{(m-1)!} \cdot [P(x \leq X_{j+i})]^{m-1} = \frac{1}{(m-1)!} \cdot [1 - F_X(x)]^{m-1} \\
&= \frac{(1-x)^{m-1}}{(m-1)!}
\end{aligned}$$

(c) Let I_j denote the initial value of the jth run. Show that $p_n(y|x)$, the probability density that the $n+1$st run has $I_{n+1} = y$ given that the nth run has just begun with $I_n = x$, equals e^{1-x} if $y < x$ and $e^{1-x} - e^{y-x}$ if $y > x$.

Solution

Let $p_n(y|x) \equiv P(y|x)$. We will be applying what we've learned from part (b).

For $y < x$, notice that

$$\begin{aligned}
P(y|x) &= \sum_{j=1}^{\infty} P(x \leq X_1 \leq \cdots \leq X_{j-1}, X_j = y) = \sum_{j=1}^{\infty} P(x \leq X_1 \leq \cdots \leq X_{j-1}) \, P(X_j = y) \\
&= \sum_{j=1}^{\infty} P(x \leq X_1 \leq \cdots \leq X_{j-1}) = \sum_{j=1}^{\infty} P(L_k \geq j|x) \\
&= \sum_{j=1}^{\infty} \frac{(1-x)^{j-1}}{(j-1)!} = \sum_{j=0}^{\infty} \frac{(1-x)^j}{j!} \\
&= e^{1-x}
\end{aligned}$$

For $y > x$, notice that

$$\begin{aligned}
P(y|x) &= \sum_{j=2}^{\infty} P(x \leq X_1 \leq \cdots \leq X_{j-1}, X_{j-1} > y, X_j = y) = \sum_{j=1}^{\infty} P(x \leq X_1 \leq \cdots \leq X_{j-1}, X_{j-1} > y) \, P(X_j = y) \\
&= \sum_{j=1}^{\infty} P(x \leq X_1 \leq \cdots \leq X_{j-1}, X_{j-1} > y)
\end{aligned}$$

Note: So at some point, one of the X_k's was greater than y.

$$\begin{aligned}
&= \sum_{j=1}^{\infty} \sum_{k=1}^{j} P(x \leq X_1 \leq \cdots \leq X_{k-1} \leq y \leq X_k \leq \cdots \leq X_{j-1}) \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{j} P(x \leq X_1 \leq \cdots \leq X_{k-1} \leq y) \, P(y \leq X_k \leq \cdots \leq X_{j-1}) \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{j} \frac{(y-x)^{k-1}}{(k-1)!} \cdot \frac{(1-y)^{j-k+1}}{(j-k+1)!}
\end{aligned}$$

Note: Switch the order of sums. We have $1 \leq k \leq j \leq \infty$.

$$\begin{aligned}
&= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} \frac{(y-x)^{k-1}}{(k-1)!} \cdot \frac{(1-y)^{j-k+1}}{(j-k+1)!} = \sum_{k=1}^{\infty} \frac{(y-x)^{k-1}}{(k-1)!} \sum_{j=k}^{\infty} \frac{(1-y)^{j-k+1}}{(j-k+1)!} \\
&= \sum_{k=1}^{\infty} \frac{(y-x)^{k-1}}{(k-1)!} \sum_{z=1}^{\infty} \frac{(1-y)^z}{z!} = \sum_{k=1}^{\infty} \frac{(y-x)^{k-1}}{(k-1)!} \left(e^{1-y} - 1\right) \\
&= e^{y-x} \left(e^{1-y} - 1\right) \\
&= e^{1-x} - e^{y-x}
\end{aligned}$$

(d) Demonstrate that $\pi(y)$, the probability density function for I_n as $n \to \infty$, satisfies $\pi(y) = 2(1-y), 0 < y < 1$. You may do this by verifying the continuous state equilibrium equations for discrete time Markov chains: $\pi(y) = \int_0^1 \pi(x) p(y|x)\, dx$.

Solution

Define $\pi(x) = 2(1-x)\mathbf{1}\{0 < x < 1\}$. The goal is to show that $\pi(y) = \int_0^1 \pi(x) p(y|x)\, dx$. And so we have

$$\begin{aligned}
\int_0^1 \pi(x) p(y|x)\, dx &= \int_0^1 2(1-x)\left[p(y|x)\,\mathbf{1}\{y > x\} + p(y|x)\,\mathbf{1}\{y < x\}\right] dx \\
&= \int_0^1 2(1-x)\left[\left(e^{1-x} - e^{y-x}\right)\mathbf{1}\{y > x\} + e^{1-x}\mathbf{1}\{y < x\}\right] dx \\
&= \int_0^y 2(1-x)(e^{1-x} - e^{y-x})dx + \int_y^1 2(1-x)e^{1-x}dx \\
&= \int_0^1 2(1-x)e^{1-x}dx - \int_0^y 2(1-x)e^{y-x}dx \\
&\quad \text{Note: Let } z = 1-x, dz = -dx \\
&= \int_1^0 2ze^z(-dz) - \int_1^{1-y} 2ze^{y-1+z}(-dz) \\
&= 2\int_0^1 ze^z dz + 2e^{y-1}\int_1^{1-y} ze^z dz \\
&= 2\left(ze^z - e^z\right)\big|_0^1 + 2e^{y-1}\left(ze^z - e^z\right)\big|_1^{1-y} \\
&= 2 + 2e^{y-1}\left[(1-y)e^{1-y} - e^{1-y} - 0\right] \\
&= 2 - 2y \\
&= 2(1-y)
\end{aligned}$$

(e) Find $\lim_{n\to\infty} P(L_n \geq m)$.

Solution

We first need to derive $P(L_n \geq m)$. We already know $P(L_j \geq m|x)$ from part (b). From part (c), we know $p_n(y|x)$. In part (d), Notice that

$$\begin{aligned}
P(L_n \geq m) &= E\left[\mathbf{1}\{L_n \geq m\}\right] \\
&= E\left[E\left[\mathbf{1}\{L_n \geq m\}|I_n = X\right]\right] \\
&= E\left[P(L_n \geq m|I_n = X)\right] \\
&= \int_0^1 P(L_n \geq m|I_n = x)\,\pi_n(x)dx \\
\lim_{n\to\infty} P(L_n \geq m) &= \int_0^1 P(L_n \geq m|I_n = x)\,\pi(x)dx \\
&= \int_0^1 \frac{(1-x)^{m-1}}{(m-1)!}\cdot 2(1-x)dx \\
&= \frac{2}{(m-1)!}\int_0^1 (1-x)^m dx \\
&= \frac{2}{(m-1)!}\left(-\frac{(1-x)^{m+1}}{m+1}\right)\Bigg|_0^1 \\
&= \frac{2}{(m+1)(m-1)!}
\end{aligned}$$

(f) What is the average length of a run as $n \to \infty$, that is, $\lim_{n\to\infty} E[L_n]$?

Solution

First calculate the expectation.

$$
\begin{aligned}
E[L_n] &= \sum_{m=1}^{\infty} mP(L_n = m) \\
&= \sum_{m=1}^{\infty}\sum_{k=1}^{m} P(L_n = m) \\
&\quad \text{Note: Notice that } 1 \le k \le m \le \infty \\
&= \sum_{k=1}^{\infty}\sum_{m=k}^{\infty} P(L_n = m) \\
&= \sum_{k=1}^{\infty} P(L_n \ge k) \\
\Rightarrow \lim_{n\to\infty} E[L_n] &= \sum_{k=1}^{\infty} \lim_{n\to\infty} P(L_n \ge k) \\
&= \sum_{k=1}^{\infty} \frac{2}{(k+1)(k-1)!} \\
&= 2\sum_{k=1}^{\infty} \frac{k}{(k+1)!} \\
&= 2\sum_{k=1}^{\infty} \frac{k+1-1}{(k+1)!} \\
&= 2\left(\sum_{k=1}^{\infty} \frac{1}{k!} - \frac{1}{(k+1)!}\right) \\
&\quad \text{Note: } \sum_{k=1}^{\infty} \frac{1}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} - 1 = e^1 - 1 \\
&\quad \text{Note: } \sum_{k=1}^{\infty} \frac{1}{(k+1)!} = \sum_{z=2}^{\infty} \frac{1}{z!} = \sum_{z=0}^{\infty} \frac{1}{z!} - 2 = e^1 - 2 \\
&= 2\left(e^1 - 1 - \left(e^1 - 2\right)\right) \\
&= 2
\end{aligned}
$$

3 Part 2

3.2.1 Question 1, (f) incomplete

1. For a given $i = 1, \dots, n$ 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 $i = 1, \dots, n$. Note: in this problem, asymptotics refers to $n \to \infty$.

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

   Solution

   The joint likelihood and log likelihood are

$$
\begin{aligned}
L_1(\psi, \lambda_1, \cdots, \lambda_n) &\equiv P(\boldsymbol{x}, \boldsymbol{y} | \psi, \lambda_1, \dots, \lambda_n) = \prod_{i=1}^n P(x_i, y_i | \psi, \lambda_i) \\
&= \prod_{i=1}^n P(x_i | \psi, \lambda_i) P(y_i | \lambda_i) \\
&= \prod_{i=1}^n \psi\lambda_i e^{-\psi\lambda_i x_i} \cdot \lambda_i e^{-\lambda_i y_i} = \prod_{i=1}^n \psi\lambda_i^2 e^{-\lambda_i(\psi x_i + y_i)} \\
&= \psi^n \cdot \prod_{i=1}^n \lambda_i^2 \cdot e^{-\sum_i \lambda_i(\psi x_i + y_i)} \\
\Rightarrow l_n(\psi, \lambda_1, \dots, \lambda_n) &\equiv \log(L_1(\psi, \lambda_1, \dots, \lambda_n)) \\
&= n\log(\psi) - \sum_i \lambda_i(\psi x_i + y_i) + \sum_i 2\log(\lambda_i) \\
&= n\log(\psi) + \sum_i [2\log(\lambda_i) - \lambda_i(\psi x_i + y_i)]
\end{aligned}
$$

   Hence the score equations are

$$
\begin{aligned}
0 &= \frac{\partial}{\partial\psi} l_n(\psi, \lambda_1, \dots, \lambda_n) \\
&= \frac{n}{\psi} - \sum_i \lambda_i x_i = U_1(\psi) \\
0 &= \frac{\partial}{\partial\lambda_i} l_n(\psi, \lambda_1, \dots, \lambda_n) \\
&= \frac{2}{\lambda_i} - (\psi x_i + y_i)
\end{aligned}
$$

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

   Solution

   Looking at the log likelihood, the issue is that MLE theory relies on a finite number of parameters governing the likelihood as $n \to \infty$.

(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.

Solution

Define $\theta = (\psi, \lambda_1, \ldots, \lambda_n)$. Starting with the score equations, differentiate once more.

$$
\begin{aligned}
\frac{\partial^2}{\partial \psi^2} l_n(\theta) = -\frac{n}{\psi^2} \quad &\Rightarrow \quad E\left[-\frac{\partial^2}{\partial \psi^2} l_n(\theta)\right] = \frac{n}{\psi^2} \\
\frac{\partial^2}{\partial \psi \partial \lambda_i} l_n(\theta) = -x_i \quad &\Rightarrow \quad E\left[-\frac{\partial^2}{\partial \psi \partial \lambda_i} l_n(\theta)\right] = E\left[x_i\right] = \frac{1}{\psi \lambda_i} \\
\frac{\partial^2}{\partial \lambda_i^2} l_n(\theta) = -\frac{2}{\lambda_i^2} \quad &\Rightarrow \quad E\left[-\frac{\partial^2}{\partial \lambda_i^2} l_n(\theta)\right] = \frac{2}{\lambda_i^2} \\
\frac{\partial^2}{\partial \lambda_i \partial \lambda_j} l_n(\theta) = 0 \quad &\Rightarrow \quad E\left[-\frac{\partial^2}{\partial \lambda_i \partial \lambda_j} l_n(\theta)\right] = 0
\end{aligned}
$$

The Fisher Information for n observations and $I(\theta)$ are

$$
\begin{aligned}
I_n(\theta) &= E\left[-\frac{\partial^2}{\partial \theta^2} l_n(\theta)\right] = \begin{bmatrix} \frac{n}{\psi^2} & \frac{1}{\psi\lambda_1} & \cdots & \cdots & \frac{1}{\psi\lambda_n} \\ \frac{1}{\psi\lambda_1} & \frac{2}{\lambda_1^2} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \frac{1}{\psi\lambda_n} & 0 & \cdots & 0 & \frac{2}{\lambda_n^2} \end{bmatrix} \\
\frac{1}{n} I_n(\theta) &= \begin{bmatrix} \frac{1}{\psi^2} & \frac{1}{n\psi\lambda_1} & \cdots & \cdots & \frac{1}{n\psi\lambda_n} \\ \frac{1}{n\psi\lambda_1} & \frac{2}{n\lambda_1^2} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \frac{1}{n\psi\lambda_n} & 0 & \cdots & 0 & \frac{2}{n\lambda_n^2} \end{bmatrix} \to \frac{1}{\psi^2} = I(\theta) \\
I^{-1}(\theta) &= I^{-1}(\psi) = \psi^2
\end{aligned}
$$

From MLE theory, we know that

$$
\sqrt{n}\left(\widehat{\psi} - \psi\right) \to_d \mathcal{N}(0, I^{-1}(\psi)) = \mathcal{N}(0, \psi^2)
$$

(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.

Solution

Continuing with the score equations from part (a),

$$\begin{aligned}
\frac{2}{\lambda_i} &= \psi x_i + y_i \\
&\Leftrightarrow \lambda_i = \frac{2}{\psi x_i + y_i} \\
&\Leftrightarrow \lambda_i x_i = \frac{2x_i}{\psi x_i + y_i} \\
\frac{n}{\psi} &= \sum_i \lambda_i x_i \\
&\Leftrightarrow \psi^{-1} = \frac{1}{n}\sum_i \lambda_i x_i \\
&\Leftrightarrow \widehat{\psi}^{-1} = \frac{1}{n}\sum_i \frac{2x_i}{\widehat{\psi} x_i + y_i} = \frac{1}{n}\sum_i \frac{2}{\widehat{\psi} + \frac{y_i}{x_i}} \\
&\Leftrightarrow \widehat{\psi}^{-1} = \frac{1}{n}\sum_i \frac{2}{\widehat{\psi} + t_i^{-1}}
\end{aligned}$$

So we see that $U_1(\widehat{\psi})$ is indeed a function of $T_i = X_i/Y_i$.

Let's first find the distribution of $T = X/Y$. Define $W = Y$. And so

$$\begin{aligned}
f_{T,W}(t,w) &= f_{X,Y}\left(g_1^{-1}(t,w), g_2^{-1}(t,w)\right) \cdot |J(t,w)| \\
&= f_X(tw) f_Y(w) \begin{vmatrix} w & t \\ 0 & 1 \end{vmatrix} \\
&= \psi\lambda_i e^{-\psi\lambda_i tw} \lambda_i e^{-\lambda_i w} w \\
&= \psi\lambda_i^2 w e^{-\lambda_i(\psi t+1)w} \\
f_T(t) &= \int_w f_{T,W}(t,w)dw \\
&= \int_0^\infty \psi\lambda_i^2 w e^{-\lambda_i(\psi t+1)w} dw \\
&= \psi\lambda_i^2 \int_0^\infty w e^{-\lambda_i(\psi t+1)w} dw \\
&= \psi\lambda_i^2 \int_0^\infty \frac{\Gamma(2)\left[\lambda_i(\psi t+1)\right]^2}{\Gamma(2)\left[\lambda_i(\psi t+1)\right]^2} \cdot w^{2-1} e^{-\lambda_i(\psi t+1)w} dw \\
&= \psi\lambda_i^2 \frac{\Gamma(2)}{\left[\lambda_i(\psi t+1)\right]^2} \int_0^\infty \frac{\left[\lambda_i(\psi t+1)\right]^2}{\Gamma(2)} \cdot w^{2-1} e^{-\lambda_i(\psi t+1)w} dw \\
&= \frac{\psi}{(\psi t+1)^2} \mathbf{1}\{0 < t < \infty\}
\end{aligned}$$

We see that the PDF of T does not depend on λ_i.

(e) Use the density of $T_1, \cdots, T_n$ 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 maximum likelihood estimator based on L_2 using standard asymptotic calculations and compare with that in part (c). Discuss.

Solution

The joint likelihood and log likelihood are as follows.

$$\begin{aligned}
L_2(\psi) &= P(\boldsymbol{t}|\psi) = \prod_{i=1}^{n} P(t_i|\psi) \\
&= \prod_{i=1}^{n} \frac{\psi}{(\psi t_i + 1)^2} \\
l_n(\psi) &\equiv \log(P(\boldsymbol{t}|\psi)) \\
&= \sum_i \log(\psi) - 2\log(\psi t_i + 1) \\
&= n\log(\psi) - 2\sum_i \log(\psi t_i + 1)
\end{aligned}$$

Now to find the score equation and compare to $U_1(\psi)$.

$$\frac{\partial}{\partial \psi} l_n(\psi) = \frac{n}{\psi} - \sum_i \frac{2t_i}{\psi t_i + 1} = 0$$

This is identical to $U_1(\psi)$ where $t_i = x_i/y_i$. Since the score equations are the same, the MLEs for ψ will be the same. However, the asymptotic variances won't necessarily be the same.

$$\begin{aligned}
\frac{\partial^2}{\partial \psi^2} l_n(\psi) &= -\frac{n}{\psi^2} - \sum_i (2t_i)(-1)(\psi t_i + 1)^{-2}(t_i) \\
&= -\frac{n}{\psi^2} + 2\sum_i \frac{t_i^2}{(\psi t_i + 1)^2} \\
I_n(\psi) = E\left[-\frac{\partial^2}{\partial \psi^2} l_n(\psi)\right] &= \frac{n}{\psi^2} - 2\sum_i E\left[\frac{t_i^2}{(\psi t_i + 1)^2}\right] \\
&= \frac{n}{\psi^2} - 2n \cdot E\left[\frac{t_i^2}{(\psi t_i + 1)^2}\right]
\end{aligned}$$

For the inner expectation, one could integrate to find the answer. But if one can recall that the exponential distribution is a specific case of the gamma distribution and if $X \perp Y$ where $X \sim \Gamma(\alpha, \theta)$ and $Y \sim \Gamma(\beta, \theta)$, then $\dfrac{X}{X+Y} \sim Beta(\alpha, \beta)$. Notice that

$$\begin{aligned}
\frac{T}{\psi T + 1} &= \frac{X/Y}{\psi X/Y + 1} = \frac{\lambda_i X}{\psi \lambda_i X + \lambda_i Y} \\
&= \frac{1}{\psi} \cdot \frac{\psi \lambda_i X}{\psi \lambda_i X + \lambda_i Y} \\
& \text{Note: } (1) X \perp Y, (2) \psi\lambda_i X \sim Exp(1) \text{ and } (3) \lambda_i Y \sim Exp(1) \\
&\sim \frac{1}{\psi} \cdot Beta(1,1) \equiv \frac{1}{\psi} Z \\
& \text{Note: If } Z \sim Beta(\alpha, \beta) \Rightarrow E[Z] = \frac{\alpha}{\alpha+\beta}, V[Z] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
E\left[\frac{T}{\psi T + 1}\right] &= E\left[\frac{1}{\psi} \cdot Z\right] = \frac{1}{\psi} \cdot \frac{1}{2} = \frac{1}{2\psi} \\
V\left[\frac{T}{\psi T + 1}\right] &= V\left[\frac{1}{\psi} \cdot Z\right] = \frac{1}{\psi^2} \cdot \frac{1}{2^2(3)} = \frac{1}{12\psi^2} \\
\Rightarrow E\left[\left(\frac{T}{\psi T + 1}\right)^2\right] &= E\left[\left(\frac{1}{\psi} \cdot Z\right)^2\right] = V\left[\frac{1}{\psi} \cdot Z\right] + \left(E\left[\frac{1}{\psi} \cdot Z\right]\right)^2 = \frac{1}{12\psi^2} + \frac{1}{4\psi^2} = \frac{1}{3\psi^2}
\end{aligned}$$

Picking up where we left off,

$$\begin{aligned}
I_n(\psi) &= \frac{n}{\psi^2} - \frac{2n}{3\psi^2} = \frac{n}{3\psi^2} \\
\frac{1}{n} I_n(\psi) &\to I(\psi) = \frac{1}{3\psi^2}
\end{aligned}$$

And so, $\sqrt{n}\left(\widehat{\psi} - \psi\right) \to_d \mathcal{N}(0, I^{-1}(\psi)) = \mathcal{N}(0, 3\psi^2)$. The asymptotic variance is larger than in part (a) which makes sense since the variable transformation eliminated the estimation of λ_i.

(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?

Solution

With the estimating equation above, we have

$$\begin{aligned} 0 &= \sum_{i=1}^{n} w_i g_i(\psi) = \sum_{i=1}^{n} w_i \left(Y_i - \psi X_i\right) \\ \Leftrightarrow \quad \widehat{\psi} &= \frac{\sum_i w_i Y_i}{\sum_i w_i X_i} \end{aligned}$$

To determine the asymptotic variance, use a Taylor Series expansion of $S_n(\widehat{\psi}) \equiv \sum_i w_i g_i(\widehat{\psi})$.

$$\begin{aligned} 0 = S_n(\widehat{\psi}) &\approx S_n(\psi_0) + \frac{\partial}{\partial \psi} S_n(\psi^*) \left(\widehat{\psi} - \psi_0\right) \\ \Leftrightarrow \sqrt{n}\left(\widehat{\psi} - \psi_0\right) &\approx \left[-\frac{\frac{\partial}{\partial \psi} S_n(\psi^*)}{n}\right]^{-1} \cdot \frac{S_n(\psi_0)}{\sqrt{n}} \end{aligned}$$

Regarding the term $-\frac{\frac{\partial}{\partial \psi} S_n(\psi^*)}{n}$, we see that

$$\begin{aligned} -\frac{\frac{\partial}{\partial \psi} S_n(\psi^*)}{n} &= -\frac{-\sum_i w_i X_i}{n} \\ &= -\frac{\sum_i w_i X_i}{n} \\ &\to_p E\left[-w_i X_i\right] = -\frac{w_i}{\psi \lambda_i} \end{aligned}$$

where w_i is optional to a proportionality constant.

Question Incomplete.

When $w_i = 1$, $\widehat{\psi} = \frac{\sum_i Y_i}{\sum_i X_i}$

3.2.2 Question 2

2. In this problem, we consider the univariate density

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

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

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

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

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

$$(\alpha,\beta)=(\mu\phi,(1-\mu)\phi)\Rightarrow(\mu,\phi)=\left(\frac{\alpha}{\alpha+\beta},\alpha+\beta\right)$$

(b) Let $Y_1,\dots,Y_n$ be a random sample from the density (2). Show that the joint density for $Y_1,\dots,Y_n$ 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).
Solution
The joint likelihood is

$$\begin{aligned}
P\left(\boldsymbol{y}|\mu,\phi\right) &= \prod_{i=1}^{n}P\left(y_i|\mu,\phi\right)=\prod_{i=1}^{n}\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y_i^{\mu\phi-1}(1-y_i)^{(1-\mu)\phi-1}\\
&= \left(\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\right)^n\left(\prod_{i=1}^{n}y_i\right)^{\mu\phi-1}\left(\prod_{i=1}^{n}(1-y_i)\right)^{(1-\mu)\phi-1}\\
&= \exp\left\{n\log\left(\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\right)+(\mu\phi-1)\sum_i\log\left(y_i\right)+((1-\mu)\phi-1)\sum_i\log(1-y_i)\right\}\\
&= \exp\left\{\begin{bmatrix}\sum_i\log\left(\frac{y_i}{1-y_i}\right) & \sum_i\log\left(1-y_i\right)\end{bmatrix}\begin{bmatrix}\mu\phi\\ \phi\end{bmatrix}+n\log\left(\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\right)-\sum_i\log\left(y_i(1-y_i)\right)\right\}
\end{aligned}$$

Since this density belongs to the exponential family, it has rank 2. Two statistics, two parameters. The sufficient and hence complete statistics are

$$\begin{bmatrix}\sum_i\log\left(\frac{y_i}{1-y_i}\right) & \sum_i\log\left(1-y_i\right)\end{bmatrix}$$

(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 X_i is a $p \times 1$ vector of covariates, with $g(\mu_i) = \beta^T X_i$, where β is a $p \times 1$ vector of unknown regression coefficients and $g(\cdot)$ is an arbitrary known link function. Define $\xi = (\beta, \phi)$.

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

Solution

The joint likelihood and log likelihood are

$$\begin{aligned}
P(\boldsymbol{y}|\boldsymbol{\mu}, \xi) &= \prod_{i=1}^{n} P(y_i|\mu_i; \xi) \\
&= \prod_{i=1}^{n} \frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma((1-\mu_i)\phi)} y_i^{\mu_i\phi-1}(1-y_i)^{(1-\mu_i)\phi-1} \\
l_n(\xi) &\equiv \log\left(P(\boldsymbol{y}|\boldsymbol{\mu}, \xi)\right) \\
&= \sum_{i=1}^{n} \left[\log(\Gamma(\phi)) - \log(\Gamma(\mu_i\phi)) - \log(\Gamma((1-\mu_i)\phi)) + (\mu_i\phi - 1)\log(y_i) + ((1-\mu_i)\phi - 1)\log(1-y_i)\right]
\end{aligned}$$

To find the score function, first let $\mu_i = \mu_i(\beta) = g^{-1}(X_i^T\beta)$,

$$\begin{aligned}
\frac{\partial l_n(\xi)}{\partial \beta} &= \sum_{i=1}^{n} \left[-\frac{\Gamma'(\mu_i\phi)}{\Gamma(\mu_i\phi)}\phi\partial_\beta\mu_i + \frac{\Gamma'((1-\mu_i)\phi)}{\Gamma((1-\mu_i)\phi)}\phi\partial_\beta\mu_i + \phi\log(y_i)\,\partial_\beta\mu_i - \phi\log(1-y_i)\,\partial_\beta\mu_i\right] \\
&= \phi\sum_{i=1}^{n}\partial_\beta\mu_i\left[\frac{\Gamma'((1-\mu_i)\phi)}{\Gamma((1-\mu_i)\phi)} - \frac{\Gamma'(\mu_i\phi)}{\Gamma(\mu_i\phi)} + \log\left(\frac{y_i}{1-y_i}\right)\right] \\
&= \phi\sum_{i=1}^{n}\partial_\beta\mu_i\left[\psi((1-\mu_i)\phi) - \psi(\mu_i\phi) + \log\left(\frac{y_i}{1-y_i}\right)\right] \\
\frac{\partial l_n(\xi)}{\partial \phi} &= \sum_{i=1}^{n}\left[\frac{\Gamma'(\phi)}{\Gamma(\phi)} - \frac{\Gamma'(\mu_i\phi)\mu_i}{\Gamma(\mu_i\phi)} - \frac{\Gamma'((1-\mu_i)\phi)(1-\mu_i)}{\Gamma((1-\mu_i)\phi)} + \mu_i\log(y_i) + (1-\mu_i)\log(1-y_i)\right] \\
&= \sum_{i=1}^{n}\left[\frac{\Gamma'(\phi)}{\Gamma(\phi)} - \frac{\Gamma'(\mu_i\phi)\mu_i}{\Gamma(\mu_i\phi)} - \frac{\Gamma'((1-\mu_i)\phi)(1-\mu_i)}{\Gamma((1-\mu_i)\phi)} + \mu_i\log\left(\frac{y_i}{1-y_i}\right) + \log(1-y_i)\right] \\
&= \sum_{i=1}^{n}\left[\frac{\Gamma'(\phi)}{\Gamma(\phi)} - \frac{\Gamma'((1-\mu_i)\phi)}{\Gamma((1-\mu_i)\phi)} + \log(1-y_i) + \mu_i\left(\frac{\Gamma'((1-\mu_i)\phi)}{\Gamma((1-\mu_i)\phi)} - \frac{\Gamma'(\mu_i\phi)}{\Gamma(\mu_i\phi)} + \log\left(\frac{y_i}{1-y_i}\right)\right)\right] \\
&= \sum_{i=1}^{n}\left[\psi(\phi) - \psi((1-\mu_i)\phi) + \log(1-y_i) + \mu_i\left(\psi((1-\mu_i)\phi) - \psi(\mu_i\phi) + \log\left(\frac{y_i}{1-y_i}\right)\right)\right]
\end{aligned}$$

We want to now show that $E\left[\frac{\partial l_n(\xi)}{\partial \xi}\right] = 0$, in other words, $E\left[\frac{\partial l_n(\xi)}{\partial \beta}\right] = 0$ and $E\left[\frac{\partial l_n(\xi)}{\partial \phi}\right] = 0$. The distribution of Y belongs to the exponential family so we know that $E\left[\frac{\partial l_n(\xi)}{\partial \xi}\right] = 0$. But also, one will notice that this task simplifies to showing that

$$E\left[\log\left(\frac{y_i}{1-y_i}\right)\right] = \psi(\mu_i\phi) - \psi((1-\mu_i)\phi) \text{ and } E\left[\log(1-y_i)\right] = \psi((1-\mu_i)\phi) - \psi(\phi)$$

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

$$I_n(\xi) = \begin{pmatrix} I_{\beta\beta} & I_{\beta\phi} \\ I_{\phi\beta} & I_{\phi\phi} \end{pmatrix},$$

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

$$w_j = \phi\left\{\psi'(\mu_j\phi) + \psi'((1-\mu_j)\phi)\right\}\frac{1}{\{g'(\mu_j)\}^2}.$$

Solution

We'll solve each of the three pieces one at a time. For the first element,

$$\begin{aligned}
\frac{\partial^2 l_n(\xi)}{\partial\beta\partial\beta^T} &= \phi\sum_{i=1}^n \left[\partial_\beta\mu_i\left(\partial_\beta\psi((1-\mu_i)\phi) - \partial_\beta\psi(\mu_i\phi)\right) + \left(\psi((1-\mu_i)\phi) - \psi(\mu_i\phi) + \log\left(\frac{y_i}{1-y_i}\right)\right)\partial_\beta^2\mu_i\right] \\
&= \phi\sum_{i=1}^n \left[-\phi\partial_\beta\mu_i^{\otimes 2}\left(\psi'((1-\mu_i)\phi) + \psi'(\mu_i\phi)\right) + \left(\psi((1-\mu_i)\phi) - \psi(\mu_i\phi) + \log\left(\frac{y_i}{1-y_i}\right)\right)\partial_\beta^2\mu_i\right] \\
E\left[-\frac{\partial^2 l_n(\xi)}{\partial\beta\partial\beta^T}\right] &= \phi\sum_{i=1}^n \left[\phi\partial_\beta\mu_i^{\otimes 2}\left(\psi'((1-\mu_i)\phi) + \psi'(\mu_i\phi)\right)\right]
\end{aligned}$$

For the derivative term,

$$\begin{aligned}
g(\mu_i) &= X_i^T\beta \Leftrightarrow \mu_i = g^{-1}(X_i^T\beta) \\
\partial_\beta\mu_i &= \partial_\beta g^{-1}(X_i^T\beta)\cdot X_i \equiv \frac{1}{g'(\mu_i)}X_i \\
\partial_\beta\mu_i^{\otimes 2} &= \partial_\beta\mu_i\cdot\partial_\beta\mu_i^T = \frac{1}{\{g'(\mu_i)\}^2}X_i X_i^T
\end{aligned}$$

And so

$$\begin{aligned}
I_{\beta\beta} &= E\left[-\frac{\partial^2 l_n(\xi)}{\partial\beta\partial\beta^T}\right] \\
&= \phi\sum_{i=1}^n\left[\phi\frac{1}{\{g'(\mu_i)\}^2}X_i X_i^T\left(\psi'((1-\mu_i)\phi) + \psi'(\mu_i\phi)\right)\right] \\
&= \phi\sum_{i=1}^n\left[X_i\phi\left(\psi'((1-\mu_i)\phi) + \psi'(\mu_i\phi)\right)\frac{1}{\{g'(\mu_i)\}^2}X_i^T\right] \\
&= \phi\sum_{i=1}^n\left[X_i w_i X_i^T\right] \\
&= \phi X^T W X
\end{aligned}$$

For the second element,

$$\begin{aligned}
\frac{\partial^2 l_n(\xi)}{\partial\beta\partial\phi} &= \sum_{i=1}^n\left[\phi\partial_\beta\mu_i\left(\psi'((1-\mu_i)\phi)(1-\mu_i) - \psi'(\mu_i\phi)(\mu_i)\right) + \partial_\beta\mu_i\left(\psi((1-\mu_i)\phi) - \psi(\mu_i\phi) + \log\left(\frac{y_i}{1-y_i}\right)\right)\right] \\
E\left[-\frac{\partial^2 l_n(\xi)}{\partial\beta\partial\phi}\right] &= \sum_{i=1}^n\left[\partial_\beta\mu_i\cdot\phi\left\{\psi'(\mu_i\phi)(\mu_i) - \psi'((1-\mu_i)\phi)(1-\mu_i)\right\}\right] \\
&= \sum_{i=1}^n\frac{1}{g'(\mu_i)}X_i\cdot c_i = X^T T c
\end{aligned}$$

For the third and final piece,

$$\begin{aligned}
\frac{\partial^2 l_n(\xi)}{\partial\phi^2} &= \sum_{i=1}^n\left[\psi'(\phi) - \psi'((1-\mu_i)\phi)(1-\mu_i) + \mu_i\left(\psi'((1-\mu_i)\phi)(1-\mu_i) - \psi'(\mu_i\phi)\mu_i\right)\right] \\
&= \sum_{i=1}^n\left[\psi'(\phi) - \psi'((1-\mu_i)\phi)(1-\mu_i)^2 - \psi'(\mu_i\phi)\mu_i^2\right] \\
E\left[-\frac{\partial^2 l_n(\xi)}{\partial\phi^2}\right] &= \sum_{i=1}^n\left[\psi'(\mu_i\phi)\mu_i^2 + \psi'((1-\mu_i)\phi)(1-\mu_i)^2 - \psi'(\phi)\right] \\
&= \sum_{i=1}^n d_i \\
&= \text{trace}\,[D]
\end{aligned}$$

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

Solution

Given the MLE $\widehat{\xi}$ we wish for the score function to equal 0, in other words, $\dot{l}_n(\widehat{\xi}) = \left.\frac{\partial l_n(\xi)}{\partial \xi}\right|_{\xi=\widehat{\xi}} = 0$. We can Taylor expand this expression around the true parameter value denoted ξ_0 and let ξ^* be some element along the path between $\widehat{\xi}$ and ξ_0.

$$\begin{aligned} 0 &\approx \dot{l}_n(\xi_0) + \ddot{l}_n(\xi^*)(\widehat{\xi} - \xi_0) \\ \Leftrightarrow \sqrt{n}\left(\widehat{\xi} - \xi_0\right) &\approx \left[-\frac{\partial_\xi^2 l_n(\xi^*)}{n}\right]^{-1} \frac{\partial_\xi l_n(\xi_0)}{\sqrt{n}} \end{aligned}$$

By the WLLN, we know that $-\frac{\partial_\xi^2 l_n(\xi^*)}{n} \to_p I(\xi)$. So by the CMT, $\left[-\frac{\partial_\xi^2 l_n(\xi^*)}{n}\right]^{-1} \to_p I^{-1}(\xi)$. By the CLT and both parts of (c),

$$\frac{\partial_\xi l_n(\xi_0)}{\sqrt{n}} = \sqrt{n}\left(\frac{\partial_\xi l_n(\xi_0)}{n} - 0\right) \to_d \mathcal{N}\left(0, I(\xi)\right)$$

where $\frac{1}{n} I_n(\xi) \to I(\xi)$. Finally by Slutsky's Theorem, we have

$$\sqrt{n}\left(\widehat{\xi} - \xi_0\right) \to_d \mathcal{N}(0, I^{-1}(\xi))$$

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

Solution

Define the transformation $q_1 = \dfrac{r_1}{s}, \ldots, q_{k-1} = \dfrac{r_{k-1}}{s}$ and $s = \sum_{j=1}^k r_j$. And so $r_1 = sq_1, \ldots, r_{k-1} = sq_{k-1}$ where $\sum_{j=1}^k q_j = 1$. The corresponding Jacobian and it's determinant are

$$J = \begin{vmatrix} \partial_{q_1} r_1 & \cdots & \partial_{q_{k-1}} r_1 & \partial_s r_1 \\ \vdots & \ddots & \ddots & \partial_s r_2 \\ \vdots & \ddots & \ddots & \vdots \\ \partial_{q_1} r_k & \partial_{q_2} r_k & \cdots & \partial_s r_k \end{vmatrix} = \begin{vmatrix} s & 0 & \cdots & 0 & q_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & s & \vdots \\ -s & -s & \cdots & \cdots & q_k \end{vmatrix} = \begin{vmatrix} s & 0 & \cdots & 0 & q_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ddots & 0 & s & q_{k-1} \\ 0 & \cdots & 0 & 0 & 1 \end{vmatrix} = s^{k-1}$$

And so we have

$$\begin{aligned}
f_{q_1,\ldots,q_k,s}(q_1, \ldots, q_k, s) &= f_{r_1,\ldots,r_k}(g_1^{-1}(q), \ldots, g_k^{-1}(q)) \cdot J = f_{r_1}(g_1^{-1}(q)) \cdots f_{r_k}(g_k^{-1}(q)) \cdot J \\
&= J \cdot \prod_{j=1}^k f_{r_j}(r_j) \\
&= J \cdot \prod_{j=1}^k \Gamma(\alpha_j) r_j^{\alpha_j - 1} e^{-r_j} \\
&= s^{k-1} \cdot \prod_{j=1}^k \Gamma(\alpha_j)(sq_j)^{\alpha_j - 1} e^{-sq_j} \\
&= s^{k-1} e^{-s\sum_j q_j} s^{-k+\sum_j \alpha_j} \cdot \prod_{j=1}^k \Gamma(\alpha_j)(q_j)^{\alpha_j-1} \\
&= s^{\sum_j \alpha_j - 1} e^{-s} \cdot \prod_{j=1}^k \Gamma(\alpha_j)(q_j)^{\alpha_j-1} \\
&\quad \text{Note: } \sum_j q_j = 1 \\
f_{q_1,\ldots,q_k}(q_1, \ldots, q_k) &= \int_s f_{q_1,\ldots,q_k,s}(q_1, \ldots, q_k, s) ds \\
&= \int_0^\infty s^{\sum_j \alpha_j - 1} e^{-s} \cdot \prod_{j=1}^k \Gamma(\alpha_j)(q_j)^{\alpha_j-1} ds \\
&= \prod_{j=1}^k \Gamma(\alpha_j) \prod_{j=1}^k (q_j)^{\alpha_j-1} \int_0^\infty \frac{\Gamma\left(\sum_j \alpha_j\right)}{\Gamma\left(\sum_j \alpha_j\right)} \cdot s^{\sum_j \alpha_j - 1} e^{-s} ds \\
&= \frac{\prod_{j=1}^k \Gamma(\alpha_j)}{\Gamma\left(\sum_j \alpha_j\right)} \prod_{j=1}^k (q_j)^{\alpha_j-1} \int_0^\infty \Gamma\left(\sum_j \alpha_j\right) \cdot s^{\sum_j \alpha_j - 1} e^{-s} ds \\
&= \frac{\prod_{j=1}^k \Gamma(\alpha_j)}{\Gamma\left(\sum_j \alpha_j\right)} \prod_{j=1}^k (q_j)^{\alpha_j-1} \\
\Rightarrow f_{q_1,\ldots,q_{k-1}}(q_1, \ldots, q_{k-1}) &= \frac{\prod_{j=1}^k \Gamma(\alpha_j)}{\Gamma\left(\sum_j \alpha_j\right)} \prod_{j=1}^{k-1} q_j^{\alpha_j - 1} \cdot \left(1 - \sum_{j=1}^{k-1} q_j\right)^{\alpha_k - 1}
\end{aligned}$$

3.2.3 Question 3

3. Consider the linear model
$$Y = X\beta + \epsilon, \tag{1}$$
where X is a $n \times p$ covariate matrix, β is a $p \times 1$ vector of regression coefficients, and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$, where the ϵ_i's are i.i.d. $\mathcal{N}(0, \sigma^2), i = 1, \ldots, n$. In this problem, both β and σ^2 are unknown.

   (a) Suppose that $\text{rank}(X) = r \leq p$ and we wish to test
$$H_0 : l^T\beta = \theta_0 \text{ versus } H_1 : l^T\beta \neq \theta_0, \tag{2}$$
where $l \in C(X^T)$ (was a typo), $C(X)$ denotes the column space of X, and θ_0 is a specified constant. Derive a UMPU size α test for the hypotheses in (2). 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.

Solution

Remarks:

- Refer to Jun Shao's *Mathematical Statistics* book, pdf page 431, book page 415.
- Notice that because $l \in C(X^T)$ that $\exists \rho$ such that $l^T = \rho^T X$ and hence that $l^T\beta$ is estimable.
- **Basic Idea**: From BIOS 761, we want to transform the model to get a simpler density function to isolate and identify the parameter of interest Θ, its corresponding sufficient statistic U, nuisance parameters $(\xi_1, \ldots, \xi_k)$, and nuisance sufficient statistics $(T_1, \ldots, T_k)$.

Since we're told that X may not be full rank, the (1) model can be transformed with an orthogonal projection matrix Γ. By construction, let $\Gamma_{n\times n} = \begin{pmatrix}\Gamma_{1,n\times r} & \Gamma_{2,n\times(n-r)}\end{pmatrix}$ where the vectors of Γ_1 form an orthonormal basis for $C(X)$ and the vectors of Γ_2 form an orthonormal basis for $C(X)^\perp$ which is guaranteed by the Gram-Schmidt process. Notice that Γ is full rank (with columns spanning $\mathbb{R}^n$) hence preserving the rank of X after the linear transformation.

Our new model is

$$\begin{aligned}
\Gamma^T Y &= \Gamma^T X\beta + \Gamma^T\epsilon \\
\Leftrightarrow \begin{pmatrix}\Gamma_1^T \\ \Gamma_2^T\end{pmatrix} Y &= \begin{pmatrix}\Gamma_1^T \\ \Gamma_2^T\end{pmatrix} X\beta + \begin{pmatrix}\Gamma_1^T \\ \Gamma_2^T\end{pmatrix}\epsilon \\
\Leftrightarrow \begin{pmatrix}\Gamma_1^T Y \\ \Gamma_2^T Y\end{pmatrix} &= \begin{pmatrix}\Gamma_1^T X\beta \\ \Gamma_2^T X\beta\end{pmatrix} + \begin{pmatrix}\Gamma_1^T\epsilon \\ \Gamma_2^T\epsilon\end{pmatrix} \\
\Leftrightarrow \begin{pmatrix}Z_1 \\ Z_2\end{pmatrix} &= \begin{pmatrix}\eta \\ 0\end{pmatrix} + \begin{pmatrix}\Gamma_1^T\epsilon \\ \Gamma_2^T\epsilon\end{pmatrix} \\
\Leftrightarrow Z &= \mu + e
\end{aligned}$$

To find the distribution of e, we have $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, therefore $e = \Gamma^T\epsilon \sim \mathcal{N}(0, \sigma^2\Gamma^T\Gamma) = \mathcal{N}(0, \sigma^2 I_n)$.
The joint likelihood is therefore

$$\begin{aligned}
P\left(Z \middle| \mu, \sigma^2\right) &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(Z-\mu\right)^T\left(Z-\mu\right)\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(Z_1^T Z_1 - 2\eta^T Z_1 + \eta^T\eta + Z_2^T Z_2\right)\right\} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{\eta^T\eta}{2\sigma^2}\right\} \exp\left\{\frac{\eta^T Z_1}{\sigma^2} - \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma^2}\right\} \\
&= P\left(Z \middle| \eta, \sigma^2\right)
\end{aligned}$$

Notice that given our transformation,

$$\begin{aligned}
l^T\widehat{\beta} &= \rho^T X(X^T X)^- X^T Y \\
&= \rho^T MY \\
&\quad \text{Note: Since } \Gamma \text{ is square, } \Gamma^T\Gamma = \Gamma\Gamma^T = I_n \\
&= \rho^T\Gamma\Gamma^T MY \\
&\quad \text{Note: } \Gamma\Gamma^T M = \begin{pmatrix}\Gamma_1 & \Gamma_2\end{pmatrix}\begin{pmatrix}\Gamma_1^T M \\ \Gamma_2^T M\end{pmatrix} = \begin{pmatrix}\Gamma_1 & \Gamma_2\end{pmatrix}\begin{pmatrix}\Gamma_1^T \\ 0\end{pmatrix} = \Gamma_1\Gamma_1^T \\
&= \rho^T\Gamma_1\Gamma_1^T Y = \rho^T\Gamma_1 Z_1 \\
&\equiv a^T Z_1 \\
l^T\beta &= E\left[l^T\widehat{\beta}\right] = E\left[a^T Z_1\right] = a^T\eta
\end{aligned}$$

where $a = (a_1, \ldots, a_r)$. This means that the sufficient statistic for Θ is contained in the term $\dfrac{\eta^T Z_1}{\sigma^2}$. Without loss of generality regarding

$l^T\beta$, assume $a_1 \neq 0$ since l has to be a nonzero vector. To find the parameters and statistics of the UMPU test, we have

$$\begin{aligned}
\frac{\eta^T Z_1}{\sigma^2} &= \frac{1}{\sigma^2}\left[\eta_1 Z_{11} + \cdots + \eta_r Z_{1r}\right] \\
&= \frac{1}{\sigma^2}\left[\left(\frac{a_1\eta_1}{a_1}\right) Z_{11} + \cdots + \eta_r Z_{1r}\right] \\
&= \frac{1}{\sigma^2}\left[\left(\frac{a^T\eta - a_2\eta_2 - \cdots - a_r\eta_r}{a_1}\right) Z_{11} + \cdots + \eta_r Z_{1r}\right] \\
&= \frac{1}{\sigma^2}\left[\left(\frac{a^T\eta - \theta_0 + \theta_0 - a_2\eta_2 - \cdots - a_r\eta_r}{a_1}\right) Z_{11} + \cdots + \eta_r Z_{1r}\right] \\
&= \frac{a^T\eta - \theta_0}{a_1\sigma^2} Z_{11} + \frac{1}{\sigma^2}\left[\left(\frac{\theta_0 - a_2\eta_2 - \cdots - a_r\eta_r}{a_1}\right) Z_{11} + \cdots + \eta_r Z_{1r}\right] \\
\frac{\eta^T Z_1}{\sigma^2} - \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma^2} &= \frac{a^T\eta - \theta_0}{a_1\sigma^2} Z_{11} + \frac{1}{\sigma^2}\left[\left(\frac{\theta_0 - a_2\eta_2 - \cdots - a_r\eta_r}{a_1}\right) Z_{11} + \cdots + \eta_r Z_{1r}\right] - \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma^2} \\
&= \frac{a^T\eta - \theta_0}{a_1\sigma^2} Z_{11} - \frac{1}{2\sigma^2}\left[Z_1^T Z_1 + Z_2^T Z_2 - 2\theta_0 Z_{11}\right] + \sum_{j=2}^{r} \frac{\eta_j}{\sigma^2}\left(Z_{1j} - \frac{a_j Z_{11}}{a_1}\right)
\end{aligned}$$

Finally we see that the null hypothesis can be re-expressed as $H_0 : \Theta = 0$ where

- $\Theta = \dfrac{a^T\eta - \theta_0}{a_1\sigma^2}$,
- $U = Z_{11}$,
- $\xi = \begin{pmatrix} -\dfrac{1}{2\sigma^2} & \dfrac{\eta_2}{\sigma^2} & \cdots & \dfrac{\eta_r}{\sigma^2} \end{pmatrix}$, and
- $T = \begin{pmatrix} Z_1^T Z_1 + Z_2^T Z_2 - \dfrac{2\theta_0 Z_{11}}{a_1} & Z_{12} - \dfrac{a_2 Z_{11}}{a_1} & \cdots & Z_{1j} - \dfrac{a_j Z_{11}}{a_1} \end{pmatrix}$

This is the case because

$$l^T\beta = \theta_0 \Leftrightarrow l^T\beta - \theta_0 = 0 \Leftrightarrow \Theta \equiv \frac{a^T\eta - \theta_0}{a_1\sigma^2} = 0$$

Intuitively, we should see that the UMPU test is the t-test with test statistic

$$t(X,Y) = \frac{l^T\widehat{\beta} - \theta_0}{\sqrt{Y^T(I-M)Y \cdot l^T(X^TX)^{-}l/(n-r)}} = \frac{a^T Z_1 - \theta_0}{\sqrt{Z_2^T Z_2 \cdot a^T a/(n-r)}} \sim^{H_0} t_{n-r}$$

Reject the null hypothesis if $|t(X,Y)| > t_{n-r,1-\alpha/2}$.

(b) Consider the model in (1) and the hypothesis in (2).

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

Solution

To calculate power at a local alternative, say $H_1 : l^T\beta = \theta_1$, we have

$$
\begin{aligned}
\text{Power} &= P\left(\text{Reject } H_0 | H_1 \text{ is true}\right) \\
&= P\left(|t(X,Y)| \geq t_{n-r,1-\alpha/2} \middle| H_1 \text{ is true}\right) \\
&= P\left(t(X,Y) \geq t_{n-r,1-\alpha/2} \text{ or } t(X,Y) \leq -t_{n-r,1-\alpha/2} \middle| H_1 \text{ is true}\right) \\
&= P\left(t(X,Y) \geq t_{n-r,1-\alpha/2} \middle| H_1 \text{ is true}\right) + P\left(t(X,Y) \leq -t_{n-r,1-\alpha/2} \middle| H_1 \text{ is true}\right) \\
&= 2 \cdot P\left(t(X,Y) \geq t_{n-r,1-\alpha/2} \middle| H_1 \text{ is true}\right) \\
&= 2 \cdot P\left(\frac{l^T\widehat{\beta} - \theta_0}{\sqrt{\widehat{\sigma}^2 \cdot l^T(X^TX)^- l}} \geq t_{n-r,1-\alpha/2} \middle| H_1 \text{ is true}\right) \\
&= 2 \cdot P\left(\frac{l^T\widehat{\beta} - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}} \geq t_{n-r,1-\alpha/2}\sqrt{\frac{\widehat{\sigma}^2 \cdot l^T(X^TX)^- l}{\sigma^2 \cdot l^T(X^TX)^- l}} + \frac{\theta_0 - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}} \middle| H_1 \text{ is true}\right) \\
&\approx 2 \cdot P\left(T \geq t_{n-r,1-\alpha/2} + \frac{\theta_0 - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}}\right) \\
&= 2 \cdot P\left(T < -t_{n-r,1-\alpha/2} - \frac{\theta_0 - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}}\right) \\
&\to 2 \cdot \Phi\left(-z_{1-\alpha/2} - \frac{\theta_0 - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}}\right) \quad \text{as } n \to \infty
\end{aligned}
$$

(ii) Suppose that $p = 2$, X is $n \times 2$ where the first column consists of a vector of ones, $\beta = (\beta_0, \beta_1)^T$, $l = (0, 1)$, $\text{rank}(X) = 2$, and $\sum_{i=1}^n x_i = n/2$, where $(x_1, \ldots, x_n)^T$ denotes the second column of 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.

Solution

Let power be denoted by $1 - \gamma$. Based on what's given,

$$
\begin{aligned}
l^T(X^TX)^- l &= \begin{bmatrix} 0 & 1 \end{bmatrix} \cdot \frac{1}{n\sum_i x_i^2 - n^2/4} \cdot \begin{bmatrix} \sum_i x_i^2 & -n/2 \\ -n/2 & n \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{1}{\sum_i x_i^2 - n/4}
\end{aligned}
$$

Using this, we'll first solve the asymptotic power function for the term above.

$$
\begin{aligned}
1 - \gamma &\approx 2 \cdot \Phi\left(-z_{1-\alpha/2} - \frac{\theta_0 - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}}\right) \\
\Leftrightarrow z_{1-\alpha/2} + z_{1-\gamma/2} &= -\frac{\theta_0 - \theta_1}{\sqrt{\sigma^2 \cdot l^T(X^TX)^- l}} \\
\Leftrightarrow l^T(X^TX)^- l &= \frac{(\theta_1 - \theta_0)^2}{\left(z_{1-\alpha/2} + z_{1-\gamma/2}\right)^2} \\
\Leftrightarrow n &= 4\left(\sum_i x_i^2 - \frac{\left(z_{1-\alpha/2} + z_{1-\gamma/2}\right)^2}{(\theta_1 - \theta_0)^2}\right)
\end{aligned}
$$

(c) Consider the model in (1) and suppose that rank$(X) = p$. We wish to test

$$H_0 : R\beta = b_0 \text{ versus } H_1 : R\beta \neq b_0 \tag{3}$$

where R is an $s \times p$ specified matrix of constants of rank $s \leq p$ and 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.

Solution

To find the MLE of β under the null, maximize the likelihood with respect to the linear constraint. Let $f(\beta)$ be the function being optimized with Lagrange multipliers.

$$\begin{aligned}
l_n(\beta, \sigma^2) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{(Y - X\beta)^T(Y - X\beta)}{2\sigma^2} \\
f(\beta, \Lambda) &= l_n(\beta, \sigma^2) + \Lambda(R\beta - b_0) \\
\frac{\partial f(\beta, \Lambda)}{\partial \beta} &= -\frac{1}{2\sigma^2}\left(-2X^TY + 2X^TX\beta\right) + R^T\Lambda^T = 0 \\
&\Leftrightarrow X^TX\beta = \sigma^2 R^T\Lambda^T + X^TY \\
&\quad \text{Note: We're told } X \text{ is full rank now} \\
&\Leftrightarrow \beta = \sigma^2(X^TX)^{-1}R^T\Lambda^T + (X^TX)^{-1}X^TY \\
\frac{\partial f(\beta, \Lambda)}{\partial \Lambda} &= \beta^TR^T - b_0^T = 0 \\
&\Leftrightarrow R\beta = b_0 \\
&\Leftrightarrow R\left[\sigma^2(X^TX)^{-1}R^T\Lambda^T + (X^TX)^{-1}X^TY\right] = b_0 \\
&\quad \text{Note: Let } R = P^TX \\
&\Leftrightarrow \Lambda^T = \frac{1}{\sigma^2}(P^TMP)^{-1}[b_0 - P^TMY] \\
&\quad \text{Note: Plugging this back into the first equation.} \\
&\Rightarrow \tilde{\beta} = \widehat{\beta} - \left(X^TX\right)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right) \\
\frac{\partial l_n(\tilde{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})}{2\sigma^4} = 0 \\
&\Leftrightarrow \tilde{\sigma}^2 = \frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})}{n}
\end{aligned}$$

Under the full parameter space, we know that

$$\begin{aligned}
\widehat{\beta} &= (X^TX)^{-1}X^TY \\
\widehat{\sigma}^2 &= \frac{(Y - X\widehat{\beta})^T(Y - X\widehat{\beta})}{n}
\end{aligned}$$

The likelihood ratio test is $\phi(\boldsymbol{y}, \boldsymbol{x}) = \begin{cases} 1 & \text{if } \Lambda < k \\ 0 & \text{if } \Lambda \geq k \end{cases}$ that rejects H_0. (Λ here is for the LRT, not the Lagrange multiplier.)

$$\begin{aligned}
\Lambda &= \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)} = \frac{(2\pi\tilde{\sigma}^2)^{-n/2}\exp\left\{-\frac{1}{2\tilde{\sigma}^2}(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})\right\}}{(2\pi\widehat{\sigma}^2)^{-n/2}\exp\left\{-\frac{1}{2\widehat{\sigma}^2}(Y - X\widehat{\beta})^T(Y - X\widehat{\beta})\right\}} \\
&= \frac{(\tilde{\sigma}^2)^{-n/2}\exp\{-n/2\}}{(\widehat{\sigma}^2)^{-n/2}\exp\{-n/2\}} = \left(\frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})/n}{(Y - X\widehat{\beta})^T(Y - X\widehat{\beta})/n}\right)^{-n/2} \\
&= \left(\frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})}{Y^T(I - M)Y}\right)^{-n/2}
\end{aligned}$$

Let's expand the numerator into something more familiar.

$$\begin{aligned}
\tilde{\beta} &= \widehat{\beta} - \left(X^TX\right)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right) \\
X\tilde{\beta} &= X\widehat{\beta} - X\left(X^TX\right)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right) \\
Y - X\tilde{\beta} &= Y - X\widehat{\beta} + X\left(X^TX\right)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right) \\
&= (I - M)Y + X\left(X^TX\right)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right) \\
(Y - X\tilde{\beta})^T(Y - X\tilde{\beta}) &= Y^T(I - M)Y + \left(R\widehat{\beta} - b_0\right)^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right)
\end{aligned}$$

The middle term of the inner product is 0 since $(I-M)X = 0$.
And so the rejection region can be expressed as

$$\begin{aligned}
\Lambda < k \quad &\Leftrightarrow \quad \left(\frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})}{Y^T(I-M)Y}\right)^{-n/2} < k \\
&\Leftrightarrow \quad -\frac{n}{2}\log\left(\frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})}{Y^T(I-M)Y}\right) < \log(k) \\
&\Leftrightarrow \quad \frac{(Y - X\tilde{\beta})^T(Y - X\tilde{\beta})}{Y^T(I-M)Y} > k' \\
&\Leftrightarrow \quad \frac{Y^T(I-M)Y + \left(R\widehat{\beta} - b_0\right)^T \left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right)}{Y^T(I-M)Y} > k' \\
&\Leftrightarrow \quad \frac{\left(R\widehat{\beta} - b_0\right)^T \left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right)}{Y^T(I-M)Y} > k' - 1 \equiv k'' \\
&\Leftrightarrow \quad F^* \equiv \frac{\left(R\widehat{\beta} - b_0\right)^T \left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta} - b_0\right)/s}{Y^T(I-M)Y/(n-p)} > k''' \\
&\Leftrightarrow \quad F^* \sim F\left(s, n-p, \gamma = \frac{(R\beta - b_0)^T \left[R(X^TX)^{-1}R^T\right]^{-1}(R\beta - b_0)}{2\sigma^2}\right) \\
&\Rightarrow \quad k''' = F(s, n-p, 1-\alpha)
\end{aligned}$$

The numerator and denominator are two independent χ^2 distributed random variables divided by their degrees of freedom. They're independent because the numerator is a function of $X\widehat{\beta} = MY$ and the denominator is a function of $(I-M)Y$ which are orthogonal. This answer agrees with the results of BIOS 762 hypothesis tests. Reject H_0 when $F^* > F(s, n-p, 1-\alpha)$.

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

Solution

Recall that the Score test statistic is

$$SC_n = \left[\partial_\xi l_n(\tilde{\xi})\right]^T \cdot E\left[-\frac{\partial^2 l_n(\xi)}{\partial\xi\partial\xi^T}\right]^{-1}\Bigg|_{\xi=\tilde{\xi}} \cdot \left[\partial_\xi l_n(\tilde{\xi})\right]$$

where $\tilde{\xi}$ is the MLE under H_0.

Let $\xi = (\beta, \sigma^2)$. From part (c), we've already obtained $\tilde{\beta}$ and $\tilde{\sigma}^2$. We just need to obtain the score and Fisher's Information.

$$\begin{aligned}
l_n(\xi) &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(Y-X\beta)^T(Y-X\beta)}{2\sigma^2} \\
\partial_\beta l_n(\xi) &= -\frac{1}{\sigma^2}(-X^TY + X^TX\beta) \\
\partial_\beta^2 l_n(\xi) &= -\frac{1}{\sigma^2}X^TX \Rightarrow E\left[-\partial_\beta^2 l_n(\xi)\right] = \frac{1}{\sigma^2}X^TX \\
\partial_{\beta,\sigma^2}^2 l_n(\xi) &= \frac{1}{\sigma^4}(-X^TY + X^TX\beta) \Rightarrow E\left[-\partial_{\beta,\sigma^2}^2 l_n(\xi)\right] = 0 \\
\partial_{\sigma^2} l_n(\xi) &= -\frac{n}{2\sigma^2} + \frac{(Y-X\beta)^T(Y-X\beta)}{2\sigma^4} \\
\partial_{\sigma^2}^2 l_n(\xi) &= \frac{n}{2\sigma^4} - \frac{(Y-X\beta)^T(Y-X\beta)}{\sigma^6} \Rightarrow E\left[-\partial_{\sigma^2}^2 l_n(\xi)\right] = \frac{n}{2\sigma^4} \\
I_n(\xi) &= \begin{bmatrix} \frac{1}{\sigma^2}X^TX & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \\
\Rightarrow I_n^{-1}(\xi) = E\left[-\frac{\partial^2 l_n(\xi)}{\partial\xi\partial\xi^T}\right]^{-1} &= \begin{bmatrix} \sigma^2(X^TX)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \\
\partial_\xi l_n(\xi) &= \begin{bmatrix} -\frac{(-X^TY+X^TX\beta)}{\sigma^2} \\ -\frac{n}{2\sigma^2} + \frac{(Y-X\beta)^T(Y-X\beta)}{2\sigma^4} \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} X^T(Y-X\beta) \\ \frac{1}{2}\left(-n + \frac{(Y-X\beta)^T(Y-X\beta)}{\sigma^2}\right) \end{bmatrix} \\
\partial_\xi l_n(\tilde{\xi}) &= \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}X^T(Y-X\tilde{\beta}) \\ 0 \end{bmatrix}
\end{aligned}$$

So the Score test statistic is now (with some algebra)

$$\begin{aligned}
SC_n &= \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}(Y-X\tilde{\beta})^TX & 0 \end{bmatrix}^T \begin{bmatrix} \tilde{\sigma}^2(X^TX)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}X^T(Y-X\tilde{\beta}) \\ 0 \end{bmatrix} \\
&= \frac{1}{\tilde{\sigma}^2}(Y-X\tilde{\beta})^TM(Y-X\tilde{\beta}) \\
&\quad \text{Note: } M(Y-X\tilde{\beta}) = X(X^TX)^{-1}R^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}-b_0\right) \\
&= \frac{nsF^*}{n-p+sF^*} \text{ where } F^* = \frac{\left(R\widehat{\beta}-b_0\right)^T\left[R(X^TX)^{-1}R^T\right]^{-1}\left(R\widehat{\beta}-b_0\right)/s}{Y^T(I-M)Y/(n-p)} \\
&\to_d^{H_0} \chi_s^2
\end{aligned}$$

We reject H_0 if $SC_n > \chi^2_{s,1-\alpha}$.

(e) Consider the model in (1) and suppose that $\text{rank}(X) = r \le p$. Derive an exact 95% confidence region for $R\beta$, where R is an $s \times p$ matrix of constants of rank $s \le r$, and all rows of R are contained in $C(X)$.

Solution

Using the Wald Test/LRT statistic from part (c), the exact 95% confidence region for $R\beta$ is

$$CR = \left\{ R\beta : \frac{\left(R\widehat{\beta}-R\beta\right)^T\left[R(X^TX)^-R^T\right]^-\left(R\widehat{\beta}-R\beta\right)/s}{Y^T(I-M)Y/(n-r)} \le F(s, n-r, 1-\alpha) \right\}$$

for $\alpha = 0.05$.

(f) Consider the model in (1) and suppose that $\text{rank}(X) = r \leq p$. Derive a UMPU size α test for testing $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$, 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.

Solution

Remark: The framework of the solution can be found in Jun Shao's *Exercises and Solutions* book/pdf Exercise 30 (# 6.53).

Using the same approach as in part (a), we start with the orthogonal/linear transformation of model (1). We arrive at the density

$$
\begin{aligned}
P\left(Z \middle| \eta, \sigma^2\right) &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{\eta^T\eta}{2\sigma^2}\right\} \exp\left\{\frac{\eta^T Z_1}{\sigma^2} - \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma^2}\right\} \\
&\propto \exp\left\{\xi^T Z_1 - \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma^2} + \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma_0^2} - \frac{Z_1^T Z_1 + Z_2^T Z_2}{2\sigma_0^2}\right\} \\
&\propto \exp\left\{\left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma^2}\right)\left(Z_1^T Z_1 + Z_2^T Z_2\right) + \xi^T Z_1\right\} \\
&= \exp\left\{\Theta U(Z) + \xi^T Z_1\right\}
\end{aligned}
$$

As one can see,

- $\Theta = \dfrac{1}{2\sigma_0^2} - \dfrac{1}{2\sigma^2}$
- $U(Z) = Z_1^T Z_1 + Z_2^T Z_2$
- $\xi^T = \eta^T/\sigma^2$
- $T(Z) = Z_1$

To double check this, notice that the null can be restated as

$$
H_0 : \sigma^2 \leq \sigma_0^2 \Leftrightarrow \frac{1}{2\sigma_0^2} - \frac{1}{2\sigma^2} \leq 0 \Leftrightarrow \Theta \leq 0
$$

The UMPU test can be unconditional if we construct a statistic $V \equiv h(U, T) = a(t)U + b(t)$ with $a(t) > 0$ and if V is independent of T on the boundary. In this case, we see that letting

$$
\begin{aligned}
V &= \frac{1}{\sigma_0^2}\left(U - T^T T\right) \\
&= \frac{1}{\sigma_0^2}\left(Z_1^T Z_1 + Z_2^T Z_2 - Z_1^T Z_1\right) \\
&= \frac{1}{\sigma_0^2}\left(Z_2^T Z_2\right) \\
&= \frac{1}{\sigma_0^2}\left(\left(\Gamma_2^T Y\right)^T \Gamma_2^T Y\right) \\
&= \frac{1}{\sigma_0^2}\left(Y^T \Gamma_2 \Gamma_2^T Y\right) \\
&\quad \text{Note: } Y^T(I - M)Y = Y^T \Gamma\Gamma^T (I - M)Y = Y^T \Gamma_2 \Gamma_2^T Y \\
&= \frac{1}{\sigma_0^2} Y^T (I - M) Y \sim^{H_0} \chi^2_{n-r}
\end{aligned}
$$

is the solution. The UMPU test of size α is

$$
\phi(X, Y) = \begin{cases} 1 & \text{if } \dfrac{1}{\sigma_0^2} Y^T (I - M) Y > \chi^2_{n-r, 1-\alpha} \\ 0 & \text{if } \dfrac{1}{\sigma_0^2} Y^T (I - M) Y \leq \chi^2_{n-r, 1-\alpha} \end{cases}
$$

as we might expect.

4 Theory 2012

4 Part 1

4.1.1 Question 1

1. Let N be a Poisson random variable with parameter μ, and let $X_1, X_2, \ldots,$ be a sequence of i.i.d Poisson random variables with parameter λ, where $0 < \mu, \lambda < \infty$. Define

$$U = \mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i,$$

where $\mathbf{1}\{A\}$ is the indicator of the event A. Do the following:

(a) Show that $E[U] = \mu\lambda$ and $\text{var}(U) = \mu\lambda(1+\lambda)$.

Solution

For the expectation,

$$\begin{aligned}
E[U] &= E\left[\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i\right] = E\left[E\left[\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i \middle| N\right]\right] \\
&= E[\mathbf{1}\{N > 0\} N \cdot E[X_i]] = \lambda E[\mathbf{1}\{N > 0\} N] \\
&\quad \text{Note: If one thinks about it, } E[\mathbf{1}\{N > 0\} N] = E[N] \\
&= \lambda E[N] \\
&= \lambda\mu
\end{aligned}$$

For the variance,

$$\begin{aligned}
V[U] &= E\left[V\left[\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i \middle| N\right]\right] + V[E[U|N]] \\
&= E[\mathbf{1}\{N > 0\} N \cdot V[X_i]] + V[E[U|N]] \\
&= \lambda\mu + V[E[U|N]] \\
&= \lambda\mu + V\left[E\left[\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i \middle| N\right]\right] \\
&= \lambda\mu + V[\mathbf{1}\{N > 0\} N \cdot E[X_i]] \\
&= \lambda\mu + \lambda^2 V[\mathbf{1}\{N > 0\} N] \\
&\quad \text{Note: If one thinks about it, } V[\mathbf{1}\{N > 0\} N] = V[N] = \mu \\
&= \lambda\mu + \lambda^2\mu \\
&= \lambda\mu(1+\lambda)
\end{aligned}$$

(b) In this part, we add a subscript k to the Poisson parameters μ and λ defined above to denote dependence on an integer $k \geq 1$. Specifically let $\mu = \mu_k = k$ and $\lambda = \lambda_k = h/k$, where $0 < h < \infty$ is a fixed scalar. We want to study what happens to U as $k \to \infty$. Let $D_i = \mathbf{1}\{X_i = 1\}$, for all $i \geq 1$, and define

$$T = \mathbf{1}\{N > 0\} \sum_{i=1}^{N} D_i.$$

Do the following:

(i) Derive the limits of $E[U]$ and $\mathrm{var}(U)$ as $k \to \infty$.

Solution

Using part (a),

$$\begin{aligned} E[U] &= \lambda_k \mu_k \\ &= \frac{h}{k} \cdot k = h \\ V[U] &= \lambda_k \mu_k (1 + \lambda_k) \\ &= \frac{h}{k} \cdot k \left(1 + \frac{h}{k}\right) \\ &= h \end{aligned}$$

So $E[U] = V[U] \to h$ as $k \to \infty$.

(ii) Show that $P(X_i \neq D_i) = \lambda_k^2 \{1 + o(\lambda_k)\}$ as $k \to \infty$.

Solution

Through expanding the LHS,

$$\begin{aligned} P(X_i \neq D_i) &= 1 - P(X_i = D_i) \\ &= 1 - P(X_i = 0) - P(X_i = 1) \\ &= 1 - e^{-\lambda_k} - \lambda_k e^{-\lambda_k} \end{aligned}$$

Using a Maclaurin expansion of e^{-x}, we have

$$\begin{aligned} e^{-x} &= \sum_{j=0}^{\infty} \frac{(-1)^j x^j}{j!} = 1 - x + \frac{x^2}{2} + o(x^3) \\ e^{-x} + xe^{-x} &= (1+x)e^{-x} = (1+x)\left(1 - x + \frac{x^2}{2} + o(x^3)\right) \\ &= 1 - x^2 + (1+x)\left(\frac{x^2}{2} + o(x^3)\right) \\ &\quad \text{Note: } x = o(1), \frac{x^2}{2} = o(1), 1 + x = 1 + o(1) \\ &= 1 - x^2 + o(x^3) \\ 1 - e^{-x} - xe^{-x} &= 1 - \left(1 - x^2 + o(x^3)\right) \\ &= x^2 + o(x^3) = x^2 + x^2 o(x) \\ &= x^2 \{1 + o(x)\} \end{aligned}$$

And so, as $k \to \infty$, we have

$$P(X_i \neq D_i) = 1 - e^{-\lambda_k} - \lambda_k e^{-\lambda_k} = \lambda_k^2 \{1 + o(\lambda_k)\}$$

(iii) Show that

$$\mathbf{1}\{U \neq T\} \leq \mathbf{1}\{N > 0\} \sum_{i=1}^{N} \mathbf{1}\{X_i \neq D_i\};$$

and thus $U - T \to 0$, in probability, as $k \to \infty$.

Solution

Looking at the LHS, recalling that $D_i \equiv \mathbf{1}\{X_i = 1\}$,

$$\begin{aligned}
\mathbf{1}\{U \neq T\} &= \mathbf{1}\left\{\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i \neq \mathbf{1}\{N > 0\} \sum_{i=1}^{N} D_i\right\} \\
&= \mathbf{1}\left\{\mathbf{1}\{N > 0\} \sum_{i=1}^{N} [X_i - D_i] \neq 0\right\} \\
&= \mathbf{1}\left\{\mathbf{1}\{N > 0\} \sum_{i=1}^{N} [X_i - \mathbf{1}\{X_i = 1\}] \neq 0\right\}
\end{aligned}$$

We know that $X_i \neq D_i$ doesn't occur for $X_i = 0, 1$.
Looking at the different cases,

- if $\forall i, X_i \in \{0, 1\}$, then

$$\mathbf{1}\{U \neq T\} = \mathbf{1}\{N > 0\} \sum_{i=1}^{N} \mathbf{1}\{X_i \neq D_i\} = 0$$

- else if $\exists i^*$ such that $X_{i^*} \notin \{0, 1\}$, then

$$\mathbf{1}\{U \neq T\} = 1 \text{ and } \mathbf{1}\{N > 0\} \sum_{i=1}^{N} \mathbf{1}\{X_i \neq D_i\} \geq 1$$

- else if $\forall i, X_i \notin \{0, 1\}$, then

$$\mathbf{1}\{U \neq T\} = 1 \text{ and } \mathbf{1}\{N > 0\} \sum_{i=1}^{N} \mathbf{1}\{X_i \neq D_i\} > 1$$

Hence we've proven the inequality. Now take the expectation of both sides.

$$\begin{aligned}
E[\mathbf{1}\{U \neq T\}] &\leq E\left[\mathbf{1}\{N > 0\} \sum_{i=1}^{N} \mathbf{1}\{X_i \neq D_i\}\right] \\
P(U \neq T) &\leq E\left[E\left[\mathbf{1}\{N > 0\} \sum_{i=1}^{N} \mathbf{1}\{X_i \neq D_i\} \middle| N\right]\right] \\
&= E[\mathbf{1}\{N > 0\} \cdot N \cdot E[\mathbf{1}\{X_i \neq D_i\}]] \\
&= P(X_i \neq D_i) E[\mathbf{1}\{N > 0\} N] \\
&= \lambda_k^2 \{1 + o(\lambda_k)\} \mu_k \\
\Leftrightarrow P(U \neq T) &\leq \left(\frac{h}{k}\right)^2 \{1 + o(\lambda_k)\} k \\
&= \frac{h^2}{k} \{1 + o(\lambda_k)\} \\
&\to 0 \\
\Rightarrow U - T &\to_p 0
\end{aligned}$$

(iv) Show that $T - \sum_{i=1}^{k} D_i \to 0$, in probability, as $k \to \infty$.

Solution

It's simple to show that the expected difference is 0 for fixed k. And so, by Chebyshev's Inequality,

$$\begin{aligned} P\left(\left|T - \sum_{i=1}^{k} D_i - 0\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2} V\left[T - \sum_{i=1}^{k} D_i\right] \\ &= \frac{1}{\epsilon^2}\left(E\left[V\left[T - \sum_{i=1}^{k} D_i \middle| N\right]\right] + V\left[E\left[T - \sum_{i=1}^{k} D_i \middle| N\right]\right]\right) \end{aligned}$$

For the first term,

$$\begin{aligned} E\left[V\left[T - \sum_{i=1}^{k} D_i \middle| N\right]\right] &= E\left[V\left[\sum_{i=k+1}^{N} D_i \middle| N\right]\right] \\ &= E\left[(N-k)V\left[D_i\right]\right] \\ &= (\mu_k - k)V\left[D_i\right] \\ &= 0 \end{aligned}$$

For the second term,

$$\begin{aligned} V\left[E\left[T - \sum_{i=1}^{k} D_i \middle| N\right]\right] &= V\left[E\left[\sum_{i=k+1}^{N} D_i \middle| N\right]\right] \\ &= V\left[(N-k)E\left[D_i\right]\right] \\ &= \left(\lambda_k e^{-\lambda_k}\right)^2 V\left[N-k\right] \\ &= \lambda_k^2 e^{-2\lambda_k}\mu_k \\ &= \frac{h^2}{k}e^{-2h/k} \end{aligned}$$

So this means $V\left[T - \sum_{i=1}^{k} D_i\right] \to 0$, hence $T - \sum_{i=1}^{k} D_i \to_p 0$.

(v) Show that U converges in distribution to a Poisson random variable with parameter h, as $k \to \infty$.

Solution

Let's begin with realizing that

$$U = U - T + T - \sum_{i=1}^{k} D_i + \sum_{i=1}^{k} D_i$$

When $k \to \infty$,

- From part (iii), we showed that $U - T \to_p 0$.
- From part (iv), it was shown that $T - \sum_{i=1}^{k} D_i \to_p 0$.

We can now focus on $\sum_{i=1}^{k} D_i$. We know that $D_i = \mathbf{1}\left\{X_i = 1\right\}$ is Bernoulli distributed. The sum of k i.i.d Bernoulli RV is binomial or

$$\begin{aligned} \sum_{i=1}^{k} D_i &\sim Bin\left(k, P\left(X_i = 1\right)\right) \\ &= Bin\left(k, \lambda_k e^{-\lambda_k}\right) \\ &= Bin\left(k, \frac{h}{k}e^{-h/k}\right) \end{aligned}$$

Recall the relationship between Binomial and Poisson. If $X \sim Bin(n,p)$ and as $n \to \infty$, (1) $np \to \lambda$, a scalar, and (2) $p \to 0$, then $X \to_d Poisson(\lambda)$. Since this is the case, we conclude that

$$\sum_{i=1}^{k} D_i \to_d Poisson(h)$$

Finally through Slutsky's Theorem, we can conclude that $U \to_d Poisson(h)$.

(c) We now modify the setting in (b) so that $\mu = \mu_k = h/k$ and $\lambda = \lambda_k = k$. Do the following:

(i) Derive the limits of $E[U]$ and $\text{var}(U)$ as $k \to \infty$.

Solution

Looking at the two terms in question ...

$$\begin{aligned}
E[U] &= \lambda_k \mu_k \\
&= k \cdot \frac{h}{k} \\
&= h \\
V[U] &= \lambda_k \mu_k (1 + \lambda_k) \\
&= k \cdot \frac{h}{k} \cdot (1 + k) \\
&= h(1 + k) \\
&\to \infty
\end{aligned}$$

(ii) Show that $U \to 0$ in distribution as $k \to \infty$.

Solution

Notice that to show this, we want to show that $P(U = 0) \to 1$ because $U \to_d c$, a constant, implies $U \to_p c$.

One approach:

$$\begin{aligned}
P(U = 0) &= P\left(\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i = 0\right) \\
&= P(N = 0) + P\left(\sum_{i=1}^{N} X_i = 0\right) \\
&= e^{-\mu_k} + \sum_{j=1}^{\infty} P\left(\sum_{i=1}^{N} X_i = 0, N = j\right) \\
&= e^{-\mu_k} + \sum_{j=1}^{\infty} P\left(\sum_{i=1}^{N} X_i = 0 \middle| N = j\right) P(N = j) \\
&= e^{-\mu_k} + \sum_{j=1}^{\infty} e^{-j\lambda_k} \cdot \frac{e^{-\mu_k} \mu_k^j}{j!} \\
&= e^{-\mu_k} + e^{-\mu_k} \sum_{j=1}^{\infty} \frac{\left(e^{-\lambda_k} \mu_k\right)^j}{j!} \\
&= e^{-\mu_k} + e^{-\mu_k} \left(\sum_{j=0}^{\infty} \frac{\left(e^{-\lambda_k} \mu_k\right)^j}{j!} - 1\right) \\
&= e^{-\mu_k} + e^{-\mu_k} \left(\exp\left\{e^{-\lambda_k} \mu_k\right\} - 1\right) \\
&= e^{-h/k} + e^{-h/k} \left(\exp\left\{e^{-k} h/k\right\} - 1\right) \\
&\to 1
\end{aligned}$$

Second approach:

Start with $P(U = 0) = 1 - P(U \neq 0)$.

$$\begin{aligned}
P(U \neq 0) &= P(U > 0) \\
&= P\left(\mathbf{1}\{N > 0\} \sum_{i=1}^{N} X_i > 0\right) \\
&\leq P(N > 0) \\
&= 1 - e^{-\mu_k} \\
&= 1 - e^{-h/k} \\
&\to 0
\end{aligned}$$

Regardless of the approach, we see that $U \to_p 0$.

4.1.2 Question 2

2. Let $Y_1, \ldots, Y_n$ be i.i.d random variables from a distribution with mean μ and finite variance. Due to non-response, we may not be able to observe all the Y_i's for these n subjects. Let $R_1, \ldots, R_n$ denote indicator of response, i.e., $R_i = 1$ means that Y_i observed and $R_i = 0$ otherwise. Suppose that we also collect additional information $X_1, \ldots, X_n$, which are i.i.d random variables, from these n subjects. Assume that R_i and Y_i are independent given X_i and that the random vectors (Y_i, R_i, X_i) are i.i.d for $i = 1, \ldots, n$. Define $\pi(x) = P(R_i = 1 | X_i = x)$ and assume $\pi(x)$ is known and bounded by a positive constant from below for any x in the support of X_i.

   (a) A simple estimator for μ is the average of the observed Y_i's:

$$\widehat{\mu}_1 = \sum_{i=1}^{n} R_i Y_i / \sum_{i=1}^{n} R_i.$$

   Derive the asymptotic limit of $\widehat{\mu}_1$, denoted by μ^*, and give the asymptotic distribution of $\sqrt{n}\left(\widehat{\mu}_1 - \mu^*\right)$. Leave expressions in the result.

   Solution

   First, we'll create some notations through finding expectations, variances, and covariances.

$$\begin{aligned}
E\left[R_i Y_i\right] &= E\left[E\left[R_i Y_i | X_i\right]\right] \\
&= E\left[E\left[R_i | X_i\right] E\left[Y_i | X_i\right]\right] \\
&\equiv E\left[\pi_X \mu_X\right] \\
V\left[R_i Y_i\right] &= E\left[(R_i Y_i)^2\right] - \left(E\left[R_i Y_i\right]\right)^2 \\
&= E\left[R_i Y_i^2\right] - \left(E\left[\pi_X \mu_X\right]\right)^2 \\
&\equiv E\left[\pi_X \mu_{2X}\right] - \left(E\left[\pi_X \mu_X\right]\right)^2 \\
E\left[R_i\right] &= E\left[\pi_X\right] \\
V\left[R_i\right] &= E\left[\pi_X\right]\left(1 - E\left[\pi_X\right]\right) \\
Cov(R_i Y_i, R_i) &= E\left[R_i R_i Y_i\right] - E\left[R_i\right] E\left[R_i Y_i\right] \\
&= E\left[R_i Y_i\right]\left(1 - E\left[R_i\right]\right) \\
&\equiv E\left[\pi_X \mu_X\right]\left(1 - E\left[\pi_X\right]\right)
\end{aligned}$$

   To find the asymptotic limit,

$$\begin{aligned}
\widehat{\mu}_1 &= \frac{\sum_{i=1}^{n} R_i Y_i}{\sum_{i=1}^{n} R_i} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} R_i Y_i}{\frac{1}{n}\sum_{i=1}^{n} R_i} \\
& \quad \text{Note: Apply } \textbf{WLLN} \text{ and } \textbf{CMT}. \\
&\to_p \frac{E\left[\pi_X \mu_X\right]}{E\left[\pi_X\right]} \\
&\equiv \mu^*
\end{aligned}$$

   To find the asymptotic distribution, use **CLT** and **Delta Method**.

$$\begin{aligned}
\sqrt{n}\left(\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} R_i Y_i \\ \frac{1}{n}\sum_{i=1}^{n} R_i \end{bmatrix} - \begin{bmatrix} E\left[\pi_X \mu_X\right] \\ E\left[\pi_X\right] \end{bmatrix}\right) &\to_d \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} E\left[\pi_X \mu_{2X}\right] - \left(E\left[\pi_X \mu_X\right]\right)^2 & E\left[\pi_X \mu_X\right]\left(1 - E\left[\pi_X\right]\right) \\ E\left[\pi_X \mu_X\right]\left(1 - E\left[\pi_X\right]\right) & E\left[\pi_X\right]\left(1 - E\left[\pi_X\right]\right) \end{bmatrix}\right) \\
\sqrt{n}\left(\widehat{\mu}_1 - \mu^*\right) &= \sqrt{n}\left(g\left(\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} R_i Y_i \\ \frac{1}{n}\sum_{i=1}^{n} R_i \end{bmatrix}\right) - g\left(\begin{bmatrix} E\left[\pi_X \mu_X\right] \\ E\left[\pi_X\right] \end{bmatrix}\right)\right) \\
&\to_d \nabla g\left(\begin{bmatrix} E\left[\pi_X \mu_X\right] \\ E\left[\pi_X\right] \end{bmatrix}\right) \cdot \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} E\left[\pi_X \mu_{2X}\right] - \left(E\left[\pi_X \mu_X\right]\right)^2 & E\left[\pi_X \mu_X\right]\left(1 - E\left[\pi_X\right]\right) \\ E\left[\pi_X \mu_X\right]\left(1 - E\left[\pi_X\right]\right) & E\left[\pi_X\right]\left(1 - E\left[\pi_X\right]\right) \end{bmatrix}\right) \\
& \quad \text{Note: I'd recommend using a temporary shorthand notation.} \\
&= \mathcal{N}\left(0, \frac{E\left[\pi_X \mu_{2X}\right] E\left[\pi_X\right] - \left(E\left[\pi_X \mu_X\right]\right)^2}{\left(E\left[\pi_X\right]\right)^3}\right)
\end{aligned}$$

(b) A Horwitz-Thompson estimator for μ is given by

$$\widehat{\mu}_2 = n^{-1} \sum_{i=1}^{n} R_i Y_i / \pi(X_i).$$

Show that $\widehat{\mu}_2$ is a consistent estimator for μ and derive the asymptotic distribution of $\sqrt{n}\,(\widehat{\mu}_2 - \mu)$. Leave expressions in the result.

Solution

To show $\widehat{\mu}_2$ is a consistent estimator for μ, notice that

$$\begin{aligned}
E\left[\frac{R_i Y_i}{\pi(X_i)}\right] &= E\left[E\left[\frac{R_i Y_i}{\pi(X_i)}\middle| X_i\right]\right] \\
&= E\left[\frac{1}{\pi(X_i)}\pi_X \mu_X\right] \\
&= E\left[\mu_X\right] = E\left[E\left[Y|X\right]\right] \\
&= E\left[Y\right] \\
&= \mu
\end{aligned}$$

Now to calculating the variance.

$$\begin{aligned}
V\left[\frac{R_i Y_i}{\pi(X_i)}\right] &= E\left[\left(\frac{R_i Y_i}{\pi(X_i)}\right)^2\right] - \left(E\left[\frac{R_i Y_i}{\pi(X_i)}\right]\right)^2 \\
&= E\left[\frac{R_i Y_i^2}{\pi(X_i)^2}\right] - \mu^2 \\
&= E\left[E\left[\frac{R_i Y_i^2}{\pi(X_i)^2}\middle| X_i\right]\right] - \mu^2 \\
&= E\left[\frac{1}{\pi(X_i)^2} E\left[R_i Y_i^2 | X_i\right]\right] - \mu^2 \\
&= E\left[\frac{\mu_{2X}}{\pi_X}\right] - \mu^2
\end{aligned}$$

By **CLT**,

$$\sqrt{n}\,(\widehat{\mu}_2 - \mu) \to_d \mathcal{N}\left(0, E\left[\frac{\mu_{2X}}{\pi_X}\right] - \mu^2\right)$$

(c) For any measurable function $g(X_i)$ with finite second moment, we define

$$\widehat{\mu}_g = n^{-1}\left\{\sum_{i=1}^{n} R_iY_i/\pi(X_i) + \sum_{i=1}^{n}\left(1 - R_i/\pi_i(X_i)\right)g(X_i)\right\}$$

Show that $\widehat{\mu}_g$ is a consistent estimator for μ and derive the asymptotic distribution of $\sqrt{n}\,(\widehat{\mu}_g - \mu)$. Leave expressions in the result.

Solution

To show consistency, we just need to show the second term's expectation equals 0.

$$\begin{aligned}
E\left[\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right] &= E\left[E\left[\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\middle|X_i\right]\right]\\
&= E\left[g(X_i)E\left[1-\frac{R_i}{\pi(X_i)}\middle|X_i\right]\right]\\
&= 0
\end{aligned}$$

Let's look at the variance of the second term.

$$\begin{aligned}
V\left[\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right] &= E\left[\left(\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right)^2\right] - 0 = E\left[\left(1-\frac{R_i}{\pi(X_i)}\right)^2 g(X_i)^2\right]\\
&= E\left[E\left[\left(1-\frac{R_i}{\pi(X_i)}\right)^2 g(X_i)^2\middle|X_i\right]\right]\\
&= E\left[g(X_i)^2E\left[\left(1-\frac{R_i}{\pi(X_i)}\right)^2\middle|X_i\right]\right]\\
&= E\left[g(X_i)^2E\left[1-\frac{2R_i}{\pi(X_i)}+\frac{R_i^2}{\pi(X_i)^2}\middle|X_i\right]\right]\\
&= E\left[g(X_i)^2\left(1-2+E\left[\frac{R_i}{\pi(X_i)^2}\middle|X_i\right]\right)\right]\\
&= E\left[g(X_i)^2\left(1-2+\frac{1}{\pi(X_i)}\right)\right]\\
&= E\left[g(X_i)^2\left(\frac{1-\pi_X}{\pi_X}\right)\right]
\end{aligned}$$

Let's look at the covariance between the first and second terms.

$$\begin{aligned}
Cov\left(\frac{R_iY_i}{\pi(X_i)}, \left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right) &= E\left[\frac{R_iY_i}{\pi(X_i)}\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right] - E\left[\frac{R_iY_i}{\pi(X_i)}\right]E\left[\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right]\\
&= E\left[\frac{R_iY_i}{\pi(X_i)}\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right] = E\left[\left(\frac{\pi(X_i)R_iY_i - R_i^2Y_i}{\pi(X_i)^2}\right)g(X_i)\right]\\
&= E\left[E\left[\left(\frac{\pi(X_i)R_iY_i - R_i^2Y_i}{\pi(X_i)^2}\right)g(X_i)\middle|X_i\right]\right]\\
&= E\left[\frac{g(X_i)}{\pi(X_i)^2}\left(\pi_X^2\mu_X - \pi_X\mu_X\right)\right]\\
&= -E\left[g(X_i)\cdot\frac{\mu_X(1-\pi_X)}{\pi_X}\right]
\end{aligned}$$

Using **CLT** and **Delta Method**, we can find the asymptotic distribution.

$$\begin{aligned}
\sqrt{n}\left(\begin{bmatrix}\frac{1}{n}\sum_i\frac{R_iY_i}{\pi(X_i)}\\ \frac{1}{n}\sum_i\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\end{bmatrix} - \begin{bmatrix}\mu\\0\end{bmatrix}\right) &\to_d \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}E\left[\frac{\mu_{2X}}{\pi_X}\right]-\mu^2 & -E\left[g(X_i)\cdot\frac{\mu_X(1-\pi_X)}{\pi_X}\right]\\ -E\left[g(X_i)\cdot\frac{\mu_X(1-\pi_X)}{\pi_X}\right] & E\left[g(X_i)^2\left(\frac{1-\pi_X}{\pi_X}\right)\right]\end{bmatrix}\right)\\
\sqrt{n}\,(\widehat{\mu}_2 - \mu) &= \sqrt{n}\left(g\left(\begin{bmatrix}\frac{1}{n}\sum_i\frac{R_iY_i}{\pi(X_i)}\\ \frac{1}{n}\sum_i\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\end{bmatrix}\right) - g\left(\begin{bmatrix}\mu\\0\end{bmatrix}\right)\right)\\
&\to_d \nabla g\left(\begin{bmatrix}\mu\\0\end{bmatrix}\right)\cdot\mathcal{N}\left(\mathbf{0}, \begin{bmatrix}E\left[\frac{\mu_{2X}}{\pi_X}\right]-\mu^2 & -E\left[g(X_i)\cdot\frac{\mu_X(1-\pi_X)}{\pi_X}\right]\\ -E\left[g(X_i)\cdot\frac{\mu_X(1-\pi_X)}{\pi_X}\right] & E\left[g(X_i)^2\left(\frac{1-\pi_X}{\pi_X}\right)\right]\end{bmatrix}\right)\\
&= \mathcal{N}\left(0, E\left[\frac{\mu_{2X}}{\pi_X}\right] - \mu^2 - E\left[2g(X_i)\cdot\frac{\mu_X(1-\pi_X)}{\pi_X}\right] + E\left[g(X_i)^2\cdot\frac{1-\pi_X}{\pi_X}\right]\right)
\end{aligned}$$

(d) Determine a function g which minimizes the asymptotic variance of $\widehat{\mu}_g$. Denote this function by $g^*(x)$.

Solution

Looking at the variance in part (c), notice that it is a quadratic function with respect to $g(X_i)$ of the form ax^2+bx+c. To minimize the variance, which is a large expectation, we want to minimize the expression within the expectation. The minimum of the quadratic function is where $x=-\dfrac{b}{2a}$ or in this case,

$$\begin{aligned} g^*(x) &= -\frac{-2\dfrac{\mu_X(1-\pi_X)}{\pi_X}}{2\dfrac{1-\pi_X}{\pi_X}} \\ &= \mu_X \\ &= E[Y|X] \end{aligned}$$

(e) Suppose that X_i is a discrete random variable with K categories. Suggest a consistent estimator for $g^*(x)$, denoted by $\widehat{g}$. Justify your answer.

Solution

Let's assume $X_i \in \{1,\dots,K\}$. An intuitive approach would be

$$\begin{aligned} \widehat{g}(x) &= \widehat{E}[Y|X=x] \\ &= \frac{\sum_{i=1}^{n}\mathbf{1}\{X_i=x\}Y_i}{\sum_{i=1}^{n}\mathbf{1}\{X_i=x\}} \end{aligned}$$

because, with the numerator and denominator divided by n, $\widehat{g}(x)$ converges in probability by **WLLN** and **CMT** to $E[Y|X=x]$.

(f) Following (e), derive the asymptotic distribution of $\sqrt{n}\,(\widehat{\mu}_{\widehat{g}}-\mu)$.

Solution

First, re-express the expression.

$$\sqrt{n}\,(\widehat{\mu}_{\widehat{g}}-\mu)=\sqrt{n}\,(\widehat{\mu}_{\widehat{g}}-\widehat{\mu}_{g^*}+\widehat{\mu}_{g^*}-\mu)$$

Looking at the first difference, we have

$$\begin{aligned} \sqrt{n}\,(\widehat{\mu}_{\widehat{g}}-\widehat{\mu}_{g^*}) &= n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{R_i}{\pi(X_i)}\right)(\widehat{g}(X_i)-g^*(X_i)) \\ &\le n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{R_i}{\pi(X_i)}\right)\sup_{x\in\{X_1,\dots,X_n\}}(\widehat{g}(X_i)-g^*(X_i)) \\ &= \sup_{x\in\{X_1,\dots,X_n\}}(\widehat{g}(X_i)-g^*(X_i))\cdot n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{R_i}{\pi(X_i)}\right) \\ &= o_p(1)O_p(1) \\ &= o_p(1) \end{aligned}$$

For the second difference,

$$\sqrt{n}\,(\widehat{\mu}_{g^*}-\mu) \quad \to_d \quad \mathcal{N}\left(0, E\left[\frac{\mu_{2X}}{\pi_X}\right]-\mu^2-E\left[\frac{\mu_X^2(1-\pi_X)}{\pi_X}\right]\right)$$

And so, by **Slutsky's Theorem**,

$$\sqrt{n}\,(\widehat{\mu}_{\widehat{g}}-\mu)\to_d \mathcal{N}\left(0, E\left[\frac{\mu_{2X}}{\pi_X}\right]-\mu^2-E\left[\frac{\mu_X^2(1-\pi_X)}{\pi_X}\right]\right)$$

(g) How would you estimate g^* if X_i is a continuous variable?

Solution

Kernel Estimation is one possibility. If we define $g(x)=E[Y|X=x]$, then

$$\widehat{g}(x)=\frac{\sum_{i=1}^{n}Y_i\cdot K\left(\dfrac{X_i-x}{h}\right)}{\sum_{i=1}^{n}K\left(\dfrac{X_i-x}{h}\right)}$$

4.1.3 Question 3, incomplete

3. (a) In this part, let T_0 be an unbiased estimator of an unknown parameter θ and consider the properties of T_0 under squared error loss.

   i. Show that $T_0 + c$ is not a minimax estimator under squared error loss, where $c \neq 0$ is a known constant.
   Solution

   ii. Show that the estimator cT_0 is not minimax under squared error loss unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T of θ, where $c \in (0, 1)$ is a known constant and $R_T(\theta)$ is the frequentist risk function for T.
   Solution

   (b) In this part, let $X = 1$ or 0 with probabilities p and q respectively, and consider the estimation of p with loss function $L(p, a)$ equal to 1 when $|a - p| \geq 0.25$ and equal to 0 otherwise. The most general randomized estimator is $T_0 = U$ when $X = 0$ and $T_0 = V$ when $X = 1$, where U and V are two random variables with known distributions.

   i. Evaluate the risk function and the maximum risk of T_0 when U and V are uniform on $(0, 0.5)$ and $(0.5, 1)$, respectively.
   Solution

   ii. Is T_0 minimax? Justify your answer rigorously.
   Solution

   (c) In this part, one has a sample of n iid normal random variables with mean θ and variance σ^2, $X_1, \ldots, X_n$.

   i. Assume $0 < \sigma^2 < K$ is known, where K is a finite positive constant. Is the sample mean $\bar{X}$ minimax with respect to the loss function $L(\theta, a) = \{\theta - a\}^2 / \sigma^2$? Justify your answer rigorously.
   Solution

   ii. Redo part (i) without assuming σ^2 is known.
   Solution

4 Part 2

4.2.1 Question 1

1. After a certain surgical procedure, some patients develop a wound infection. Typically, the infection is treated and cleared. However, some patients develop another wound infection. The first infection is called the "primary infection", while the second is called a "secondary infection". An investigator is interested in the question whether the risk of a secondary infection in those who have had a primary infection is the same as the risk of a primary infection.

   Data are collected on a random sample of n patients. Assume that the n responses are independent and identically distributed. For the i-th patient, $1 \leq i \leq n$, let Y_{i1} denote a binary indicator of a primary infection and Y_{i2} a binary indicator of a secondary infection, both coded as 0 for "no", and 1 for "yes". Define $\alpha = P(Y_{i1} = 1)$ and $\beta = P(Y_{i2} = 1 | Y_{i1} = 1)$. Both α and β take values in $(0, 1)$. Suppose there are X_1 patients with $Y_{i1} = 1, Y_{i2} = 1$; X_2 patients with $Y_{i1} = 1, Y_{i2} = 0$; X_3 patients with $Y_{i1} = 0, Y_{i2} = 0$. Note: $X_1 + X_2 + X_3 = n$. By definition, a secondary infection can occur only in patients who have had a primary infection.

   (a) Does the distribution of the data have the form of the exponential family? Give details.

   Solution

   Remarks:

   - If a patient has a primary and secondary infection, then

   $$P(Y_{i1} = 1, Y_{i2} = 1) = P(Y_{i2} = 1 | Y_{i1} = 1)\, P(Y_{i1} = 1) = \beta\alpha$$

   - Else if a patient has only a primary infection, then

   $$P(Y_{i1} = 1, Y_{i2} = 0) = P(Y_{i2} = 0 | Y_{i1} = 1)\, P(Y_{i1} = 1) = (1 - \beta)\alpha$$

   - Else a patient has neither, then

   $$P(Y_{i1} = 0, Y_{i2} = 0) = P(Y_{i1} = 0) = 1 - \alpha$$

   - We see that these probabilities sum to 1.

   To determine if the data arises from an exponential family, let's look at the joint likelihood.

$$\begin{aligned}
P(X_1, X_2, X_3 | \alpha, \beta; n) &= \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \\
&= \frac{n!}{x_1! x_2! x_3!} [\beta\alpha]^{x_1} [(1-\beta)\alpha]^{x_2} [1-\alpha]^{x_3} \\
&= \exp\left\{ \log\left(\frac{n!}{x_1! x_2! x_3!} [\beta\alpha]^{x_1} [(1-\beta)\alpha]^{x_2} [1-\alpha]^{x_3} \right) \right\} \\
& \quad \text{Note: } \sum_{j=1}^{3} x_j = n \\
&= \exp\left\{ x_1 \log\left(\frac{\beta\alpha}{1-\alpha} \right) + x_2 \log\left(\frac{(1-\beta)\alpha}{1-\alpha} \right) + n \log(1-\alpha) + \log\left(\frac{n!}{x_1! x_2! x_3!} \right) \right\} \\
&= \exp\left\{ \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \log\left(\frac{\beta\alpha}{1-\alpha} \right) \\ \log\left(\frac{(1-\beta)\alpha}{1-\alpha} \right) \end{bmatrix} + n \log(1-\alpha) + \log\left(\frac{n!}{x_1! x_2! x_3!} \right) \right\}
\end{aligned}$$

   At this point, let's reparameterize: $\theta_1 \equiv \log\left(\frac{\beta\alpha}{1-\alpha} \right)$ and $\theta_2 \equiv \log\left(\frac{(1-\beta)\alpha}{1-\alpha} \right)$. Hence, we'll see that

$$\log(1-\alpha) = -\log\left(1 + e^{\theta_1} + e^{\theta_2}\right)$$

   The likelihood is now

$$\begin{aligned}
P(X_1, X_2 | \theta_1, \theta_2; n) &= \exp\left\{ \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} - n \log\left(1 + e^{\theta_1} + e^{\theta_2}\right) + \log\left(\frac{n!}{x_1! x_2! x_3!} \right) \right\} \\
&\equiv \exp\left\{ Q(\boldsymbol{x})^T \Theta - b(\Theta) + c(\boldsymbol{x}) \right\}
\end{aligned}$$

   So yes, the distribution belongs to the exponential family.

(b) Derive the maximum-likelihood estimators of α and β.

Solution

Let $\xi = (\alpha, \beta)$. Using the joint likelihood from part (a), we can obtain the log likelihood to maximize.

$$\begin{aligned}
l_n(\xi) &= x_1 \log\left(\frac{\beta\alpha}{1-\alpha}\right) + x_2 \log\left(\frac{(1-\beta)\alpha}{1-\alpha}\right) + n\log(1-\alpha) + \log\left(\frac{n!}{x_1!x_2!x_3!}\right) \\
&= x_1\left[\log(\beta) + \log(\alpha) - \log(1-\alpha)\right] + x_2\left[\log(1-\beta) + \log(\alpha) - \log(1-\alpha)\right] + n\log(1-\alpha) + \log\left(\frac{n!}{x_1!x_2!x_3!}\right) \\
\frac{\partial l_n(\xi)}{\partial \alpha} &= x_1\left[\frac{1}{\alpha} + \frac{1}{1-\alpha}\right] + x_2\left[\frac{1}{\alpha} + \frac{1}{1-\alpha}\right] - \frac{n}{1-\alpha} = 0 \\
&\Leftrightarrow \frac{x_1+x_2}{\alpha(1-\alpha)} = \frac{n}{1-\alpha} \\
&\Leftrightarrow \widehat{\alpha} = \frac{x_1+x_2}{n} \\
\frac{\partial l_n(\xi)}{\partial \beta} &= \frac{x_1}{\beta} - \frac{x_2}{1-\beta} = 0 \\
&\Leftrightarrow \widehat{\beta} = \frac{x_1}{x_1+x_2}
\end{aligned}$$

(c) Derive the asymptotic covariance matrix of the estimators derived above.

Solution

We need to calculate the Fisher Information matrix $I_n(\xi)$.

$$\begin{aligned}
\frac{\partial^2 l_n(\xi)}{\partial \alpha^2} &= (x_1+x_2)\left[-\frac{1}{\alpha^2} + \frac{1}{(1-\alpha)^2}\right] - \frac{n}{(1-\alpha)^2} \\
E\left[-\frac{\partial^2 l_n(\xi)}{\partial \alpha^2}\right] &= E\left[x_1+x_2\right]\left[-\frac{1}{\alpha^2} + \frac{1}{(1-\alpha)^2}\right] + \frac{n}{(1-\alpha)^2} \\
&\quad \text{Note: } E\left[x_1+x_2\right] = np_1 + np_2 = n(\beta\alpha + (1-\beta)\alpha) = n\alpha \\
&= \frac{n}{\alpha(1-\alpha)} \\
\frac{\partial^2 l_n(\xi)}{\partial\alpha\partial\beta} &= 0 \\
E\left[-\frac{\partial^2 l_n(\xi)}{\partial\alpha\partial\beta}\right] &= 0 \\
\frac{\partial^2 l_n(\xi)}{\partial \beta^2} &= -\frac{x_1}{\beta^2} - \frac{x_2}{(1-\beta)^2} \\
E\left[-\frac{\partial^2 l_n(\xi)}{\partial \beta^2}\right] &= \frac{E\left[x_1\right]}{\beta^2} + \frac{E\left[x_2\right]}{(1-\beta)^2} \\
&= \frac{n\alpha}{\beta} + \frac{n\alpha}{1-\beta} \\
&= \frac{n\alpha}{\beta(1-\beta)}
\end{aligned}$$

The Fisher Information matrix is

$$I_n(\xi) = \begin{bmatrix} \frac{n}{\alpha(1-\alpha)} & 0 \\ 0 & \frac{n\alpha}{\beta(1-\beta)} \end{bmatrix}$$

with $\frac{1}{n}I_n(\xi) \to I(\xi) = \begin{bmatrix} \frac{1}{\alpha(1-\alpha)} & 0 \\ 0 & \frac{\alpha}{\beta(1-\beta)} \end{bmatrix}$. By MLE theory, with regularity conditions holding,

$$\sqrt{n}\left(\widehat{\xi} - \xi\right) \to_d \mathcal{N}(0, I^{-1}(\xi))$$

where $I^{-1}(\xi) = \begin{bmatrix} \alpha(1-\alpha) & 0 \\ 0 & \frac{\beta(1-\beta)}{\alpha} \end{bmatrix}$, the asymptotic covariance matrix.

(d) Does there exist a UMP test for testing

$$H_0 : \beta = 0.5 \text{ versus } H_1 : \beta > 0.5?$$

If so, then please find it. If not, then explain why such a test does not exist.

Solution

- One approach: We've shown that the distribution belongs to a 2-parameter exponential family and only UMPU tests exist in that case. Hence a UMP test doesn't exist because α is a nuisance parameter.
- Second approach: We can start by applying the Neyman Pearson Lemma for the most powerful α level test of $H_0 : \beta = \beta_0$ vs. $H_1 : \beta = \beta_1$. If the rejection region isn't a function of β_1, then there exists a UMP (uniformly most powerful) test. According to NP Lemma, there exists a constant k and critical function ϕ of the form

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \dfrac{p_1(\boldsymbol{x})}{p_0(\boldsymbol{x})} > k \\ 0 & \text{if } \dfrac{p_1(\boldsymbol{x})}{p_0(\boldsymbol{x})} \leq k \end{cases}$$

such that $E_{\theta_0}[\phi(\boldsymbol{x})] = \alpha$. Looking at the ratio of densities under the alternative and null,

$$\begin{aligned}
\frac{p_1(\boldsymbol{x})}{p_0(\boldsymbol{x})} &= \frac{\exp\left\{x_1 \log\left(\frac{\beta_1\alpha}{1-\alpha}\right) + x_2 \log\left(\frac{(1-\beta_1)\alpha}{1-\alpha}\right) + n\log(1-\alpha) + \log\left(\frac{n!}{x_1!x_2!x_3!}\right)\right\}}{\exp\left\{x_1 \log\left(\frac{\beta_0\alpha}{1-\alpha}\right) + x_2 \log\left(\frac{(1-\beta_0)\alpha}{1-\alpha}\right) + n\log(1-\alpha) + \log\left(\frac{n!}{x_1!x_2!x_3!}\right)\right\}} \\
&= \frac{\exp\left\{x_1 \log\left(\frac{\beta_1\alpha}{1-\alpha}\right) + x_2 \log\left(\frac{(1-\beta_1)\alpha}{1-\alpha}\right)\right\}}{\exp\left\{x_1 \log\left(\frac{\beta_0\alpha}{1-\alpha}\right) + x_2 \log\left(\frac{(1-\beta_0)\alpha}{1-\alpha}\right)\right\}} \\
&= \left(\frac{\beta_1}{\beta_0}\right)^{x_1} \left(\frac{1-\beta_1}{1-\beta_0}\right)^{x_2}
\end{aligned}$$

Unfortunately, the ratio is a function of β_1 implying that no UMP test exists.

(e) Derive the likelihood-ratio test statistic for testing

$$H_0 : \alpha - \beta = 0 \text{ versus } H_1 : \alpha - \beta \neq 0$$

Solution

The likelihood ratio test statistic is

$$\Lambda_n = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)}.$$

Starting with the log likelihood, under the null, let $\theta_0 = \alpha = \beta$ and maximize.

$$\begin{aligned}
l_n(\theta_0) &= x_1 \log\left(\frac{\theta_0^2}{1-\theta_0}\right) + x_2 \log(\theta_0) + n\log(1-\theta_0) + C \\
&= x_1 \left[2\log(\theta_0) - \log(1-\theta_0)\right] + x_2 \log(\theta_0) + n\log(1-\theta_0) + C \\
\frac{\partial l_n(\theta_0)}{\partial \theta_0} &= x_1 \left[\frac{2}{\theta_0} + \frac{1}{1-\theta_0}\right] + \frac{x_2}{\theta_0} - \frac{n}{1-\theta_0} = 0 \\
&\Leftrightarrow \widehat{\theta}_0 = \frac{2x_1 + x_2}{n + x_1 + x_2} \text{ and } 1 - \widehat{\theta}_0 = \frac{n - x_1}{n + x_1 + x_2}
\end{aligned}$$

Under the full parameter space, we know that

$$\widehat{\alpha} = \frac{x_1 + x_2}{n} \text{ and } \widehat{\beta} = \frac{x_1}{x_1 + x_2}.$$

And so the LRT statistic is

$$\begin{aligned}
\Lambda_n &= \frac{\dfrac{n!}{x_1!x_2!x_3!} \widehat{\theta}_0^{2x_1} (1-\widehat{\theta}_0)^{x_2+x_3} \widehat{\theta}_0^{x_2}}{\dfrac{n!}{x_1!x_2!x_3!} \left[\widehat{\beta}\widehat{\alpha}\right]^{x_1} \left[(1-\widehat{\beta})\widehat{\alpha}\right]^{x_2} \left[1-\widehat{\alpha}\right]^{x_3}} \\
&= \frac{\left[\dfrac{2x_1 + x_2}{n + x_1 + x_2}\right]^{2x_1+x_2} \left[\dfrac{n - x_1}{n + x_1 + x_2}\right]^{x_2+x_3}}{\left[\dfrac{x_1}{n}\right]^{x_1} \left[\dfrac{x_2}{n}\right]^{x_2} \left[\dfrac{x_3}{n}\right]^{x_3}}
\end{aligned}$$

We know that $-2\log(\Lambda_n)$ is asymptotically χ_1^2 under the null hypothesis. We reject H_0 if $-2\log(\Lambda_n) > \chi^2_{1,1-\alpha}$.

(f) Derive the score test of the hypotheses in part (e).

Solution

Let $\tilde{\xi} = (\tilde{\alpha}, \tilde{\beta})$ be the MLE under H_0 which is $\widehat{\theta}_0 \equiv \tilde{\alpha} = \tilde{\beta}$. The Score test statistic is defined as

$$SC_n = \left\{ [\partial_\xi l_n(\xi)]^T \left(E\left[-\partial_\xi^2 l_n(\xi)\right]\right)^{-1} \partial_\xi l_n(\xi) \right\}\Big|_{\xi=\tilde{\xi}}$$

From part (b), the score vector is

$$\partial_\xi l_n(\xi) = \begin{bmatrix} \partial_\alpha l_n(\xi) \\ \partial_\beta l_n(\xi) \end{bmatrix} = \begin{bmatrix} \dfrac{x_1+x_2}{\alpha(1-\alpha)} - \dfrac{n}{1-\alpha} \\ \dfrac{x_1}{\beta} - \dfrac{x_2}{1-\beta} \end{bmatrix}.$$

Using part (c), the inverse of Fisher information matrix is

$$\left(E\left[-\partial_\xi^2 l_n(\xi)\right]\right)^{-1} = \begin{bmatrix} \dfrac{\alpha(1-\alpha)}{n} & 0 \\ 0 & \dfrac{\beta(1-\beta)}{n\alpha} \end{bmatrix}$$

And so,

$$\begin{aligned}
[\partial_\xi l_n(\xi)]^T \left(E\left[-\partial_\xi^2 l_n(\xi)\right]\right)^{-1} \partial_\xi l_n(\xi) &= \begin{bmatrix} \dfrac{x_1+x_2}{\alpha(1-\alpha)} - \dfrac{n}{1-\alpha} & \dfrac{x_1}{\beta} - \dfrac{x_2}{1-\beta} \end{bmatrix} \begin{bmatrix} \dfrac{\alpha(1-\alpha)}{n} & 0 \\ 0 & \dfrac{\beta(1-\beta)}{n\alpha} \end{bmatrix} \begin{bmatrix} \dfrac{x_1+x_2}{\alpha(1-\alpha)} - \dfrac{n}{1-\alpha} \\ \dfrac{x_1}{\beta} - \dfrac{x_2}{1-\beta} \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{x_1+x_2}{n} - \alpha & \dfrac{1}{n\alpha}\left[x_1(1-\beta) - x_2\beta\right] \end{bmatrix} \begin{bmatrix} \dfrac{x_1+x_2}{\alpha(1-\alpha)} - \dfrac{n}{1-\alpha} \\ \dfrac{x_1}{\beta} - \dfrac{x_2}{1-\beta} \end{bmatrix} \\
&= \frac{n}{\alpha(1-\alpha)} \left(\frac{x_1+x_2}{n} - \alpha \right)^2 + \frac{x_1(1-\beta) - x_2\beta}{n\alpha\beta(1-\beta)} \\
SC_n &= \frac{n}{\widehat{\theta}_0\left(1-\widehat{\theta}_0\right)} \left(\frac{x_1+x_2}{n} - \widehat{\theta}_0 \right)^2 + \frac{x_1\left(1-\widehat{\theta}_0\right) - x_2\widehat{\theta}_0}{n\widehat{\theta}_0^2\left(1-\widehat{\theta}_0\right)} \\
& \quad \text{Note: Recall } \widehat{\theta}_0 = \frac{2x_1+x_2}{n+x_1+x_2}
\end{aligned}$$

The score test statistic is asymptotically χ_1^2 under H_0 and so we reject H_0 if $SC_n > \chi^2_{1,1-\alpha}$.

(g) Derive the Wald test statistic for the hypotheses in part (e).

Solution

Since the hypothesis is linear in ξ, the Wald test statistic simplifies to

$$W_n = \left\{ (R\xi - b_0)^T \left[R\left(E\left[-\partial_\xi^2 l_n(\xi)\right]\right)^{-1} R^T \right]^{-1} (R\xi - b_0) \right\}\Big|_{\xi=\widehat{\xi}}$$

where $\widehat{\xi}$ is the MLE under the full parameter space, $\xi = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, $b_0 = 0$, $R = \begin{bmatrix} 1 & -1 \end{bmatrix}$. To simplify, we have

$$\begin{aligned}
R\xi - b_0 &= \alpha - \beta \\
\left(E\left[-\partial_\xi^2 l_n(\xi)\right]\right)^{-1} &= \begin{bmatrix} \dfrac{\alpha(1-\alpha)}{n} & 0 \\ 0 & \dfrac{\beta(1-\beta)}{n\alpha} \end{bmatrix} \\
R\left(E\left[-\partial_\xi^2 l_n(\xi)\right]\right)^{-1} R^T &= \frac{\alpha(1-\alpha)}{n} + \frac{\beta(1-\beta)}{n\alpha} \\
W_n &= \frac{(\widehat{\alpha} - \widehat{\beta})^2}{\dfrac{\widehat{\alpha}(1-\widehat{\alpha})}{n} + \dfrac{\widehat{\beta}(1-\widehat{\beta})}{n\widehat{\alpha}}} \\
& \quad \text{Note: Recall } \widehat{\alpha} = \frac{x_1+x_2}{n} \text{ and } \widehat{\beta} = \frac{x_1}{x_1+x_2}
\end{aligned}$$

Wald's test statistic is also asymptotically χ_1^2 under H_0 and so we reject H_0 if $W_n > \chi^2_{1,1-\alpha}$.

(h) Now, suppose we are interested in inference about β only, while considering α as a nuisance parameter. Derive a conditional likelihood for β which does not depend on α. Compute the maximum likelihood estimator for β and compare with the estimator for β in part (b). Is the result intuitive?

Solution

To eliminate the nuisance parameter, we need to condition on the complete sufficient statistic for α. To find it, write out the joint likelihood once more in the multiparameter exponential form.

$$\begin{aligned}
P\left(X_1, X_2, X_3|\alpha, \beta; n\right) &= \frac{n!}{x_1!x_2!x_3!}\left[\beta\alpha\right]^{x_1}\left[(1-\beta)\alpha\right]^{x_2}(1-\alpha)^{x_3} \\
&= \frac{n!}{x_1!x_2!(n-x_1-x_2)!}\beta^{x_1}\left[1-\beta\right]^{x_2}\left(\frac{\alpha}{1-\alpha}\right)^{x_1+x_2}(1-\alpha)^n \\
&= P\left(X_1, X_2|\alpha, \beta; n\right)
\end{aligned}$$

At this point, we see that $S \equiv X_1 + X_2$ is the complete sufficient statistic for $\dfrac{\alpha}{1-\alpha}$. But what is its distribution of S?

$$\begin{aligned}
P\left(S|\alpha; n\right) &= \frac{n!}{s!(n-s)!}\left[\beta\alpha + (1-\beta)\alpha\right]^s(1-\alpha)^{n-s} \\
&= \frac{n!}{s!(n-s)!}\alpha^s(1-\alpha)^{n-s} \\
&\quad \text{Note: We see that } S = X_1 + X_2 \sim Bin(n, \alpha) \\
&= \frac{n!}{(x_1+x_2)!(n-x_1-x_2)!}\left(\frac{\alpha}{1-\alpha}\right)^{x_1+x_2}(1-\alpha)^n
\end{aligned}$$

And now to find the conditional likelihood.

$$\begin{aligned}
P\left(X_1|S, \alpha, \beta; n\right) &= \frac{P\left(X_1, S|\alpha, \beta; n\right)}{P\left(S|\alpha; n\right)} \\
&= \frac{P\left(X_1, X_2|\alpha, \beta; n\right)}{P\left(S|\alpha; n\right)} \\
&= \frac{\dfrac{n!}{x_1!x_2!(n-x_1-x_2)!}\beta^{x_1}\left[1-\beta\right]^{x_2}\left(\dfrac{\alpha}{1-\alpha}\right)^{x_1+x_2}(1-\alpha)^n}{\dfrac{n!}{(x_1+x_2)!(n-x_1-x_2)!}\left(\dfrac{\alpha}{1-\alpha}\right)^{x_1+x_2}(1-\alpha)^n} \\
&= \frac{(x_1+x_2)!}{x_1!x_2!}\beta^{x_1}(1-\beta)^{x_2} \\
&= \frac{s!}{x_1!(s-x_1)!}\beta^{x_1}(1-\beta)^{s-x_1}
\end{aligned}$$

Since the conditional likelihood also follows the binomial distribution, we know that the conditional MLE for β is

$$\widehat{\beta}_c = \frac{X_1}{S} = \frac{X_1}{X_1 + X_2}$$

which equals the unconditional MLE for β from earlier, seems intuitive.

4.2.2 Question 2, (c)(iii) incomplete

2. Consider the linear model

$$Y = X\beta + \epsilon, \quad (0.1)$$

where

$$E[\epsilon] = 0, \mathrm{Cov}(\epsilon) = \Sigma, \quad (0.2)$$

X is $n \times p$ of rank $r \leq p$, β is $p \times 1$, Y is $n \times 1$, ϵ is $n \times 1$, (β, Σ) are both unknown and Σ is an unstructured positive semidefinite matrix. Let $\widehat{\beta}$ be a least squares estimate (LSE) of β.

In the sequel, let $C(A)$ denote the column space of a matrix A, let $\|a\| = \sqrt{a^T a}$ for a column vector a, let $N_n(a, b)$ denote the n dimensional multivariate normal distribution with mean vector a and covariance matrix b, and let I_s be the s dimensional identity matrix.

(a) Let λ be a $p \times 1$ vector of scalars and let η be an $n \times 1$ vector of scalars. Let U be a conformable matrix such that $X^T U = 0$ and $C(U) \cup C(X) = \mathbb{R}^n$. Show than the following statements are all equivalent:

(i) $\lambda^T \widehat{\beta}$ is the best linear unbiased estimator (BLUE) of $\lambda^T \beta$ for any $\lambda \in C(X^T)$.

(ii) $E\left[\lambda^T \widehat{\beta} \eta^T Y\right] = 0$ for any $\lambda \in C(X^T)$ and any η such that $E\left[\eta^T Y\right] = 0$.

(iii) $X^T \Sigma U = 0$.

(iv) $\Sigma = X V_1 X^T + U V_2 U^T$ for some matrices V_1 and V_2.

(v) The matrix $X(X^T X)^- X^T \Sigma$ is symmetric, where $(X^T X)^-$ denotes an arbitrary generalized inverse of $X^T X$.

Solution

To prove the five statements are equivalent, our plan is to show (i) $\Leftrightarrow$ (ii), next (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (ii).

- Prove (ii) $\Rightarrow$ (i):
  Let's suppose any other unbiased linear estimator of $\lambda^T \beta$ is characterized by $a^T Y$ such that $E\left[a^T Y\right] = \lambda^T \beta$. Looking at the variance,

$$\begin{aligned} V\left[a^T Y\right] &= V\left[a^T Y - \lambda^T \widehat{\beta} + \lambda^T \widehat{\beta}\right] \\ &= V\left[a^T Y - \lambda^T \widehat{\beta}\right] + V\left[\lambda^T \widehat{\beta}\right] - 2 \cdot \mathrm{Cov}\left(a^T Y - \lambda^T \widehat{\beta}, \lambda^T \widehat{\beta}\right) \end{aligned}$$

  Looking at the covariance term,

$$\begin{aligned} \mathrm{Cov}\left(a^T Y - \lambda^T \widehat{\beta}, \lambda^T \widehat{\beta}\right) &= E\left[\left(a^T Y - \lambda^T \widehat{\beta}\right) \lambda^T \widehat{\beta}\right] - E\left[a^T Y - \lambda^T \widehat{\beta}\right] E\left[\lambda^T \widehat{\beta}\right] \\ & \quad \text{Note: From our first statement, } E\left[a^T Y - \lambda^T \widehat{\beta}\right] = 0 \\ &= E\left[\left(a^T Y - \lambda^T \widehat{\beta}\right) \lambda^T \widehat{\beta}\right] \\ & \quad \text{Note: Recall that } \lambda^T \widehat{\beta} = \rho^T M Y \\ &= E\left[\left(a^T Y - \rho^T M Y\right) \lambda^T \widehat{\beta}\right] = E\left[\left(a^T - \rho^T M\right) Y \lambda^T \widehat{\beta}\right] \\ & \quad \text{Note: From (ii), } E\left[\lambda^T \widehat{\beta} \eta^T Y\right] = 0 \text{ if } E\left[\eta^T Y\right] = 0 \text{ where } \eta^T = a^T - \rho^T M \\ &= E\left[\eta^T Y \lambda^T \widehat{\beta}\right] \\ &= 0 \end{aligned}$$

  Hence we've just shown that

$$V\left[a^T Y\right] = V\left[a^T Y - \lambda^T \widehat{\beta}\right] + V\left[\lambda^T \widehat{\beta}\right] \Rightarrow V\left[a^T Y\right] \geq V\left[\lambda^T \widehat{\beta}\right]$$

  meaning that $\lambda^T \widehat{\beta}$, an **unbiased** estimator, has the **smallest variance** among all **linear** estimators. Therefore $\lambda^T \widehat{\beta}$ is BLUE.

- Prove (i) $\Rightarrow$ (ii):
  Let's suppose that $E\left[\eta^T Y\right] = 0$. First, given that $\lambda^T \widehat{\beta}$ is BLUE, then we know it's unbiased. Let any other linear unbiased estimator be characterized by $\lambda^T \widehat{\beta} + a \eta^T Y$ for a scalar $a \geq 0$ because $E\left[\lambda^T \widehat{\beta} + a \eta^T Y\right] = \lambda^T \beta + a E\left[\eta^T Y\right] = \lambda^T \beta$. Second, we have the inequality and notice that

$$\begin{aligned} V\left[\lambda^T \widehat{\beta}\right] &\leq V\left[\lambda^T \widehat{\beta} + a \eta^T Y\right] = V\left[\lambda^T \widehat{\beta}\right] + a^2 V\left[\eta^T Y\right] + 2aE\left[\lambda^T \widehat{\beta} \eta^T Y\right] \\ \Leftrightarrow 0 &\leq a^2 V\left[\eta^T Y\right] + 2aE\left[\lambda^T \widehat{\beta} \eta^T Y\right] = a\left(aV\left[\eta^T Y\right] + 2E\left[\lambda^T \widehat{\beta} \eta^T Y\right]\right) \\ \Leftrightarrow 0 &\leq aV\left[\eta^T Y\right] + 2E\left[\lambda^T \widehat{\beta} \eta^T Y\right] \Leftrightarrow -\frac{a}{2} V\left[\eta^T Y\right] \leq E\left[\lambda^T \widehat{\beta} \eta^T Y\right] \\ & \quad \text{Note: This inequality needs to hold for all } \eta. \text{ So substitute } \eta \text{ with } -\eta. \\ \Leftrightarrow \frac{a}{2} V\left[\eta^T Y\right] &\geq E\left[\lambda^T \widehat{\beta} \eta^T Y\right] \\ \Rightarrow \left|E\left[\lambda^T \widehat{\beta} \eta^T Y\right]\right| &\leq \frac{a}{2} V\left[\eta^T Y\right] \end{aligned}$$

  Lastly, the above inequality needs to hold for all $a \geq 0$. The only way this is possible is if $E\left[\lambda^T \widehat{\beta} \eta^T Y\right] = 0$.

- Prove (ii) $\Rightarrow$ (iii):
  We're told that $X^TU = 0$ and $C(U) \cup C(X) = \mathbb{R}^n$. Hence $C(U) = C(X)^\perp$. For any η such that $E\left[\eta^T Y\right] = 0$, we see that $\eta^T X\beta = 0$, for all β, and hence $\eta \in C(X)^\perp$. To show $X^T\Sigma U = 0$, pick any $a \in C(X^T)$ and any $u \in C(U) = C(X)^\perp$ and expand.

$$\begin{aligned}
a^T\Sigma u &= a^T \mathrm{Cov}(Y,Y)u \\
&= \mathrm{Cov}(a^TY, u^TY) \\
&= E\left[a^TYu^TY\right] - E\left[a^TY\right]E\left[u^TY\right] \\
&\quad \text{Note: } E\left[u^TY\right] = u^TX\beta = 0 \\
&= E\left[a^TYu^TY\right] \\
&\quad \text{Note: } a^TY = a^T\left[MY + (I-M)Y\right] = a^TMY \\
&= E\left[a^TMYu^TY\right] \\
&= E\left[a^TX\widehat{\beta}u^TY\right] \\
&= 0
\end{aligned}$$

  Hence $X^T\Sigma U = 0$.
- Prove (iii) $\Rightarrow$ (iv):
  We're told that Σ is positive semidefinite and we know Σ is symmetric. By the Spectral Theorem, there exists the decomposition $\Sigma = QQ^T$ where $\mathrm{rank}(\Sigma) = \mathrm{rank}(Q)$. Define $Q = XA + UB$ for some matrices A and B because $C(X) \cup C(U) = \mathbb{R}^n$. Therefore

$$\begin{aligned}
\Sigma &= QQ^T \\
&= (XA+UB)(XA+UB)^T \\
&= (XA+UB)\left(A^TX^T + B^TU^T\right) \\
&= XAA^TX^T + UBA^TX^T + XAB^TU^T + UBB^TU^T \\
\Rightarrow X\left(X^TAB^TU^T\right)X &= XX^TAB^TU^TX = 0 \\
\Rightarrow X^T\Sigma U &= X^TXAB^TU^TU = 0
\end{aligned}$$

  So we see that $X^TXAB^TU^TU = X^TXAB^TU^TX = 0 \Rightarrow X^TXAB^TU^T = 0$. This last equation means that $XAB^TU^T \in N(X^T) = C(X)^\perp$. But notice that the columns of XAB^TU^T are contained in $C(X)$. These two statements are valid if and only if $XAB^TU^T = 0$ and hence $UBA^TX^T = \left(XAB^TU^T\right)^T = (0)^T = 0$. We can finally conclude that

$$\Sigma = XAA^TX^T + UBB^TU^T \equiv XV_1X^T + UV_2U^T$$

- Prove (iv) $\Rightarrow$ (v):
  Notice that

$$\begin{aligned}
X(X^TX)^-X^T\Sigma &= X(X^TX)^-X^T\left(XV_1X^T + UV_2U^T\right) \\
&= XV_1X^T = XAA^TX^T \\
&= XA(XA)^T
\end{aligned}$$

  which is symmetric.
- Prove (v) $\Rightarrow$ (ii):
  Note that for any $\eta \in C(X)^\perp$,

$$\begin{aligned}
E\left[\lambda^T\widehat{\beta}\eta^TY\right] &= \mathrm{Cov}\left(\lambda^T\widehat{\beta}, \eta^TY\right) + E\left[\lambda^T\widehat{\beta}\right]E\left[\eta^TY\right] \\
&= \mathrm{Cov}\left(\lambda^T\widehat{\beta}, \eta^TY\right) \\
&= \mathrm{Cov}\left(\rho^TMY, \eta^TY\right) \\
&= \rho^TM \cdot \mathrm{Cov}(Y,Y) \cdot \eta \\
&= \rho^T\left[X(X^TX)^-X^T\Sigma\right]\eta \\
&= \rho^T\left[\Sigma X(X^TX)^-X^T\right]\eta \\
&= 0
\end{aligned}$$

(b) Consider the model in (0.1) and (0.2) (above) where $\epsilon \sim N_n(0, \sigma^2 I_n)$, X is of full rank, and β and σ^2 are unknown. We wish to estimate β under the loss function $L(\beta, a) = (\beta - a)^T(\beta - a)$. Consider the estimator

$$\tilde{\beta} = \widehat{\beta} - \frac{(p-2)\widehat{\sigma}^2}{\left\|(X^TX)\left(\widehat{\beta} - c\right)\right\|^2}\left(X^TX\right)\left(\widehat{\beta} - c\right),$$

where $\widehat{\sigma}^2 = SSE/(n-p+2)$, SSE denotes the error sum of squares, and $c \in \mathbb{R}^p$ is a column vector of fixed constants. Show that the frequentist risk of $\tilde{\beta}$ is smaller than the frequentist risk of $\widehat{\beta}$.

Solution

Before proceeding, let's redefine the notation. Let $X \sim \mathcal{N}(\theta, \sigma^2 D)$, where $\sigma^2 > 0$ is known, and D is a known positive definite matrix. Let $\delta_{c,r}$ be an estimator characterized as

$$\delta_{c,r} = X - \frac{(p-2)r\sigma^2}{\|D^{-1}(X-c)\|^2}(X - c)$$

Hence $X \equiv \widehat{\beta}$, $D \equiv (X^TX)^{-1}$, $\theta \equiv \beta$, and r is a constant when σ^2 is known. There is an interesting property regarding the frequentist risk to make note of.

$$\begin{aligned}
R(\theta, \delta_{c,r}) &= E\left[L(\theta, \delta_{c,r})\right] = E\left[(\delta_{c,r} - \theta)^T(\delta_{c,r} - \theta)\right] \\
&= E\left[\|\delta_{c,r} - \theta\|^2\right] \\
&= E\left[\left\|X - \frac{(p-2)r\sigma^2}{\|D^{-1}(X-c)\|^2}(X-c) - \theta\right\|^2\right] \\
&= E\left[\left\|(X-c) - (\theta - c) - \frac{(p-2)r\sigma^2}{\|D^{-1}(X-c)\|^2}(X-c)\right\|^2\right] \\
&= R(\theta - c, \delta_{0,r})
\end{aligned}$$

Therefore we only need to consider the case with $c = 0$. Now let's transform the random variables. Let $Z \equiv D^{-1/2}X$, $Z \sim \mathcal{N}(b, \sigma^2 I)$ where $b \equiv D^{-1/2}\theta$. So now

$$\begin{aligned}
R(\theta, \delta_{0,r}) &= E\left[\left\|X - \theta - \frac{(p-2)r\sigma^2}{\|D^{-1}X\|^2}X\right\|^2\right] \\
&= R(\theta, X) - 2(p-2)r\sigma^2 E\left[(Z-b)^T\frac{Z}{\left\|D^{-1/2}Z\right\|^2}\right] + (p-2)^2r^2\sigma^4 E\left[\frac{1}{\left\|D^{-1/2}Z\right\|^2}\right]
\end{aligned}$$

Looking at the middle term,

$$\begin{aligned}
E\left[(Z-b)^T\frac{Z}{\left\|D^{-1/2}Z\right\|^2}\right] &= E\left[\sum_{j=1}^p (Z_j - b_j)\frac{Z_j}{\left\|D^{-1/2}Z\right\|^2}\right] \\
&= \sum_{j=1}^p E\left[(Z_j - b_j)\frac{Z_j}{\left\|D^{-1/2}Z\right\|^2}\right]
\end{aligned}$$

At this point, we can apply **James Stein's Lemma** b/c of our setup. The lemma states that if $X \sim \mathcal{N}(\theta, \sigma^2 I_p)$ and $g : \mathbb{R}^n \to \mathbb{R}^n$ differentiable, then

$$E\left[(X_i - \theta_i)g_i(X)\right] = \sigma^2 E\left[\frac{\partial}{\partial x_i}g(X)\right] \text{ for } i = 1, \ldots, p$$

And so, the middle term is

$$\begin{aligned}
\sum_{j=1}^p E\left[(Z_j - b_j)\frac{Z_j}{\left\|D^{-1/2}Z\right\|^2}\right] &= \sum_{j=1}^p \sigma^2 E\left[\frac{\partial}{\partial z_j}\frac{Z_j}{\left\|D^{-1/2}Z\right\|^2}\right] \\
&= \sigma^2 E\left[\frac{pZ^TD^{-1}Z - 2Z^TD^{-1}Z}{\left\|D^{-1/2}Z\right\|^4}\right] \\
&= \sigma^2(p-2)E\left[\frac{1}{\left\|D^{-1/2}Z\right\|^2}\right]
\end{aligned}$$

Looking back at the risk, it further simplifies to

$$R(\theta, \delta_{0,r}) = R(\theta, X) - (2r - r^2)(p-2)^2\sigma^4 E\left[\frac{1}{\left\|D^{-1/2}Z\right\|^2}\right]$$

Now substitute the original terms into the frequentist risk and treating σ^2 as unknown, hence $r = \dfrac{\widehat{\sigma}^2}{\sigma^2}$, we'll need to use conditional expectation. Recall that $\widehat{\beta} \perp \widehat{\sigma}^2$.

$$\begin{aligned}
R(\beta, \tilde{\beta}) &= E\left[L(\beta, \tilde{\beta})\right] \\
&= E_{\sigma^2}\left[E\left[L(\beta, \tilde{\beta})\middle|\widehat{\sigma}^2\right]\right] \\
&= E_{\sigma^2}\left[R(\beta, \widehat{\beta}) - \left(2\frac{\widehat{\sigma}^2}{\sigma^2} - \frac{\widehat{\sigma}^4}{\sigma^4}\right)(p-2)^2\sigma^4 E\left[(Z-b)^T \frac{Z}{\left\|D^{-1/2}Z\right\|^2}\right]\right] \\
&= R(\beta, \widehat{\beta}) - \left(2\frac{E\left[\widehat{\sigma}^2\right]}{\sigma^2} - \frac{E\left[\widehat{\sigma}^4\right]}{\sigma^4}\right)(p-2)^2\sigma^4 E\left[\frac{1}{\left\|D^{-1/2}Z\right\|^2}\right] \\
&\quad \text{Note: } \frac{\widehat{\sigma}^2(n-p+2)}{\sigma^2} \sim \chi^2_{n-p} \Rightarrow E\left[\widehat{\sigma}^2\right] = \frac{n-p}{n-p+2}\sigma^2 \text{ and } E\left[\widehat{\sigma}^4\right] = \frac{n-p}{n-p+2}\sigma^4 \\
&= R(\beta, \widehat{\beta}) - \frac{n-p}{n-p+2}(p-2)^2\sigma^4 E\left[\frac{1}{\left\|D^{-1/2}Z\right\|^2}\right]
\end{aligned}$$

Finally, we've shown that the frequentist risk of $\tilde{\beta}$ is smaller than that of $\widehat{\beta}$.

(c) Consider the model in (0.1) and (0.2) (above) with $\epsilon \sim N_n(0, \sigma^2 I_n)$, where β and σ^2 are unknown. We wish to conduct a hypothesis test of

$$H_0 : E[Y] \in C(X_0) \text{ vs. } H_1 : E[Y] \in C(X)$$

where $C(X_0) \subset C(X)$ for a known matrix X_0. In developing a test statistic, it is conjectured that a better estimator of σ^2 is $\|P_2 Y\|^2/(n-q)$ since it is based on more degrees of freedom, where $q = \text{rank}(X_0) \leq r$ and P_2 is the orthogonal projection operator onto the orthogonal complement of $C(X_0)$. Consider the statistic

$$G = \left(\frac{n-q}{r-q}\right)\frac{\|P_1 Y\|^2}{\|P_2 Y\|^2},$$

where P_1 denotes the orthogonal projection operator onto $C(X) \cap C(X_0)^c$, and $C(X_0)^c$ denotes the complement of $C(X_0)$.

(i) Derive the distribution of G under H_0 and under H_1.

Solution

The orthogonal projection operator onto $C(X_0)$ shall be defined as $M_0 = X_0\left(X_0^T X_0\right)^- X_0^T$. So the orthogonal projection onto the orthogonal complement of $C(X_0)$ is $P_2 \equiv I - M_0$. If we let $M = X\left(X^T X\right)^- X^T$ be the orthogonal projection onto $C(X)$, then $P_1 \equiv M - M_0$, the orthogonal projection operator onto $C(X) \cap C(X_0)^c$. Looking at the statistic,

$$\begin{aligned}
G &= \left(\frac{n-q}{r-q}\right)\frac{\|P_1 Y\|^2}{\|P_2 Y\|^2} = \left(\frac{n-q}{r-q}\right)\frac{Y^T(M-M_0)Y}{Y^T(I-M_0)Y} \\
&= \left(\frac{n-q}{r-q}\right)\frac{Y^T(M-M_0)Y}{Y^T(I-M+M-M_0)Y} \\
&= \left(\frac{n-q}{r-q}\right)\frac{Y^T(M-M_0)Y}{Y^T(I-M)Y + Y^T(M-M_0)Y} \\
&= \left(\frac{n-q}{r-q}\right)\frac{\frac{Y^T(M-M_0)Y}{Y^T(I-M)Y}}{1+\frac{Y^T(M-M_0)Y}{Y^T(I-M)Y}} \\
&= \left(\frac{n-q}{r-q}\right)\frac{\frac{Y^T(M-M_0)Y}{Y^T(I-M)Y}\left(\frac{1/(r-q)}{1/(n-r)}\right)\left(\frac{r-q}{n-r}\right)}{1+\frac{Y^T(M-M_0)Y}{Y^T(I-M)Y}\left(\frac{1/(r-q)}{1/(n-r)}\right)\left(\frac{r-q}{n-r}\right)}
\end{aligned}$$

Note: $F^* \equiv \dfrac{Y^T(M-M_0)Y/(r-q)}{Y^T(I-M)Y/(n-r)} \sim F(r-q, n-r, \gamma)$ where $\gamma = \dfrac{E[Y]^T(M-M_0)E[Y]}{2\sigma^2} = \dfrac{\|(M-M_0)X\beta\|^2}{2\sigma^2}$

$$= \left(\frac{n-q}{r-q}\right)\frac{\left(\frac{r-q}{n-r}\right)F^*}{1+\left(\frac{r-q}{n-r}\right)F^*}$$

Note: If $X \sim F(a, b, \gamma)$, then $\dfrac{\frac{a}{b}X}{1+\frac{a}{b}X} \sim Beta\left(\frac{a}{2}, \frac{b}{2}, \gamma\right)$

$$\equiv \left(\frac{n-q}{r-q}\right)B^*$$

So under H_0, $G \sim \left(\frac{n-q}{r-q}\right)\cdot Beta\left(\frac{r-q}{2}, \frac{n-r}{2}\right)$ and under H_1, $G \sim \left(\frac{n-q}{r-q}\right)\cdot Beta\left(\frac{r-q}{2}, \frac{n-r}{2}, \gamma\right)$.

(ii) Derive the relationship of G to the usual F statistic for conducting the hypothesis test above.

Solution

From (i), we can see that the test statistic G is a function of a F distributed F statistic. Let's solve the test statistic for F in terms of G.

$$\begin{aligned}
G &= \left(\frac{n-q}{r-q}\right)\frac{\left(\frac{n-r}{r-q}\right)F^*}{1+\left(\frac{n-r}{r-q}\right)F^*} \\
&\Rightarrow \quad F^* = \left(\frac{n-r}{r-q}\right)\frac{\frac{r-q}{n-q}G}{1-\frac{r-q}{n-q}G} \sim F(r-q, n-r, \gamma)
\end{aligned}$$

(iii) Is G better than the usual F statistic in terms of statistical power? Justify your answer.
Solution
We know that power is defined as

$$\begin{aligned}\text{Power} &= P\left(F^* > f_{1-\alpha} | H_1 \text{ true}\right) \\ &= P\left(\left(\frac{n-r}{r-q}\right)\frac{\frac{r-q}{n-q}G}{1-\frac{r-q}{n-q}G} > f_{1-\alpha}\middle| H_1 \text{ true}\right)\end{aligned}$$

NOT DONE!!

4.2.3 Question 3, incomplete, refer to Byron's Solution

3. Consider independent observations $y_1, \ldots, y_n$, where $y_i = (y_{i1}, y_{i2})^T$ is a bivariate binary random vector such that y_{ij} takes values 0 and 1 for $j = 1, 2$. Suppose that $y_i \sim QE(\theta, \lambda)$, where $QE(\theta, \lambda)$ is a bivariate binary distribution of quadratic exponential form

$$p\left(y_i|\theta,\lambda\right) = \Delta(\theta,\lambda)^{-1}\exp\left\{y_{i1}\theta_1 + y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1},y_{i2})\right\},$$

where $\Delta(\theta, \lambda)$ is a normalizing constant and $C(y_{i1}, y_{i2})$ is a 'shape' function independent of $\theta = (\theta_1, \theta_2)^T$ and λ.

(a) Derive both the marginal distribution of y_{i1} and the conditional distribution of y_{i2} given y_{i1}. Specify a sufficient and necessary condition such that y_{i1} and y_{i2} are independent.

Solution

For the marginal distribution of y_{i1},

$$\begin{aligned}
P\left(y_{i1}|\theta,\lambda\right) &= \sum_{y_{i2}=0}^{1} P\left(y_{i1},y_{i2}|\theta,\lambda\right) \\
&= \sum_{y_{i2}=0}^{1} \Delta(\theta,\lambda)^{-1}\exp\left\{y_{i1}\theta_1 + y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1},y_{i2})\right\} \\
&= \Delta(\theta,\lambda)^{-1}\exp\left\{y_{i1}\theta_1\right\}\sum_{y_{i2}=0}^{1}\exp\left\{y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1},y_{i2})\right\} \\
&= \Delta(\theta,\lambda)^{-1}e^{y_{i1}\theta_1}\left[e^{-C(y_{i1},0)} + e^{\theta_2 + y_{i1}\lambda - C(y_{i1},1)}\right]
\end{aligned}$$

Given the joint distribution and the marginal, we can easily obtain the conditional.

$$\begin{aligned}
P\left(y_{i2}|y_{i1},\theta,\lambda\right) &= \frac{P\left(y_{i1},y_{i2}|\theta,\lambda\right)}{P\left(y_{i1}|\theta,\lambda\right)} \\
&= \frac{\Delta(\theta,\lambda)^{-1}e^{y_{i1}\theta_1 + y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1},y_{i2})}}{\Delta(\theta,\lambda)^{-1}e^{y_{i1}\theta_1}\left[e^{-C(y_{i1},0)} + e^{\theta_2 + y_{i1}\lambda - C(y_{i1},1)}\right]} \\
&= \frac{e^{y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1},y_{i2})}}{e^{-C(y_{i1},0)} + e^{\theta_2 + y_{i1}\lambda - C(y_{i1},1)}} \\
&= \frac{e^{y_{i2}\theta_2 + y_{i1}y_{i2}\lambda - C(y_{i1},y_{i2})}}{e^{-C(y_{i1},0)} + e^{\theta_2 + y_{i1}\lambda - C(y_{i1},1)}}
\end{aligned}$$

To obtain the sufficient and necessary conditions for independence, the goal is to express the joint as a product of the marginals. The marginal for y_{i2} is

$$P\left(y_{i2}|\theta,\lambda\right) = \sum_{y_{i1}=0}^{1} P\left(y_{i1},y_{i2}|\theta,\lambda\right)$$

(b) Calculate the marginal mean of y_i, denoted by $\mu = (\mu_1, \mu_2)^T = E[y_i]$, the marginal product moment of $y_{i1}y_{i2}$, denoted by $\eta_{12} = E[y_{i1}y_{i2}]$, and the marginal product centered moment of $(y_{i1} - \mu_1)(y_{i2} - \mu_2)$, denoted by $\sigma_{12} = E[(y_{i1} - \mu_1)(y_{i2} - \mu_2)]$.

Solution

(c) Calculate the Jacobian of the transformation from the canonical parameters θ and λ to the marginal parameters μ and η_{12}, denoted by $V = \partial(\theta, \lambda)/\partial(\mu, \eta_{12})$. Use V^{-1} to characterize the covariance matrix of $(y_i^T, y_{i1}y_{i2})^T$ and specify a sufficient and necessary condition such that this transformation is one-to-one.

Solution

(d) Suppose that we also observe $p \times 1$ column vector x_i for each i and that conditionally on $x_i, y_i \sim QE(\theta_i, \lambda_i)$, where $\theta_i = (\theta_{i1}, \theta_{i2})^T$ and λ_i may be depend on x_i, for $i = 1, \ldots, n$. Consider the model

$$E[y_i|x_i] = \mu_i = (\mu_{i1}, \mu_{i2})^T = \mu(x_i, \beta), E[(y_{i1} - \mu_{i1})(y_{i2} - \mu_{i2})|x_i] = \sigma_{i12} = \sigma_{12}(x_i, \beta, \alpha),$$

where β is an unknown $p \times 1$ regression parameter and α is an unknown scalar parameter. Derive the likelihood score equations for $(\alpha, \beta^T)^T$ and simplify them using the result obtained in part (c). Please clarify whether such estimating equations explicitly involve $C(y_{i1}, y_{i2})$.

Solution

(e) Consider generalized estimation equations for α and β given by

$$\sum_{i=1}^{n} \frac{\partial(\mu_i, \sigma_{i12})}{\partial(\alpha, \beta^T)} \frac{\partial l\,(y_i|\theta_i, \lambda_i)}{\partial(\theta_i, \lambda_i)} = 0$$

Compare the estimate of $(\alpha, \beta^T)^T$ in part (d) with that in part (e) in terms of the statistical efficiency. To do so, provide an explicit comparison of the asymptotic variances of these estimators.

Solution

(f) Will the results in parts (a)-(e) be changed if y_{i1} and y_{i2} are continuous variables instead of binary variables? Please explain. If so, then derive the corresponding results and compare with those obtained above.

Solution

5 Theory 2013

5 Part 1

5.1.1 Question 1

1. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d sample of n pairs of random variables, each pair having joint density

$$f(x, y; \alpha) = \alpha(\alpha+1)(1+x+y)^{-(\alpha+2)}, x, y > 0,$$

for parameter $0 < \alpha < \infty$. Do the following:

(a) Show that the maximum likelihood estimator for α, $\widehat{\alpha}_n$, has the following properties:

(i) $\widehat{\alpha}_n$ exists and is unique and has the form $g^{-1}(\widehat{\mu}_n)$, where $\widehat{\mu}_n = n^{-1}\sum_{i=1}^n \log(1+X_i+Y_i)$ and g^{-1} is the inverse of some function g, i.e., if $b = g(a)$, then $g^{-1}(b) = a$. Give the form of g and show g^{-1} exists.

Solution

The joint likelihood and log likelihood are

$$\begin{aligned}
P(\boldsymbol{x}, \boldsymbol{y}|\alpha) &= \prod_{i=1}^n P(x_i, y_i|\alpha) = \prod_{i=1}^n \alpha(\alpha+1)(1+x_i+y_i)^{-(\alpha+2)} \\
&= (\alpha(\alpha+1))^n \prod_{i=1}^n (1+x_i+y_i)^{-(\alpha+2)} \\
l_n(\alpha) = \log(P(\boldsymbol{x}, \boldsymbol{y}|\alpha)) &= n[\log(\alpha) + \log(\alpha+1)] - (\alpha+2)\sum_{i=1}^n \log(1+x_i+y_i)
\end{aligned}$$

When maximizing with respect to α,

$$\begin{aligned}
\frac{\partial l_n(\alpha)}{\partial \alpha} &= n\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right) - \sum_{i=1}^n \log(1+x_i+y_i) = 0 \\
&\Leftrightarrow \frac{1}{\alpha} + \frac{1}{\alpha+1} = \widehat{\mu}_n \Leftrightarrow 2\alpha + 1 = \alpha(\alpha+1)\widehat{\mu}_n \\
&\Leftrightarrow 0 = \widehat{\mu}_n \alpha^2 + (\widehat{\mu}_n - 2)\alpha - 1 \\
&\Leftrightarrow \alpha = \frac{-(\widehat{\mu}_n - 2) \pm \sqrt{(\widehat{\mu}_n - 2)^2 - 4\widehat{\mu}_n(-1)}}{2\widehat{\mu}_n} \\
&\quad \text{Note: We're told that } \alpha > 0. \\
&\Leftrightarrow \alpha = \frac{-(\widehat{\mu}_n - 2) + \sqrt{(\widehat{\mu}_n - 2)^2 + 4\widehat{\mu}_n}}{2\widehat{\mu}_n} \\
&\Rightarrow \widehat{\alpha}_n = g^{-1}(\widehat{\mu}_n) \Leftrightarrow \widehat{\mu}_n = g(\widehat{\alpha}_n)
\end{aligned}$$

The form of g is $g(x) = \frac{1}{x} + \frac{1}{x+1}$.

(ii) $\widehat{\alpha}_n \to \alpha_0$, almost surely, where α_0 is the true value of α.

Solution

Using the **SLLN** and **CMT**, we know that $\widehat{\mu}_n \to_{a.s.} \mu$ and hence, for $h \equiv g^{-1}$, we have $\widehat{\alpha}_n = h(\widehat{\mu}_n) \to_{a.s.} \alpha_0 = h(\mu)$.

(iii) $\sqrt{n}(\widehat{\alpha}_n - \alpha_0)$ is asymptotically normal with mean zero and variance

$$\sigma_1^2 = \frac{\alpha_0^2(\alpha_0+1)^2}{\alpha_0^2 + (\alpha_0+1)^2}$$

Solution

The asymptotic variance can be obtain through (1) **CLT** and **Delta Method** or (2) applying MLE theory. I'll use the second approach.

$$\begin{aligned}
\frac{\partial^2 l_n(\alpha)}{\partial \alpha^2} &= n\left(-\frac{1}{\alpha^2} - \frac{1}{(\alpha+1)^2}\right) \\
I_n(\alpha) = E\left[-\frac{\partial^2 l_n(\alpha)}{\partial \alpha^2}\right] &= n\left(\frac{1}{\alpha^2} + \frac{1}{(\alpha+1)^2}\right) = n\frac{\alpha^2 + (\alpha+1)^2}{\alpha^2(\alpha+1)^2} \\
I(\alpha) &= \frac{1}{n} I_n(\alpha)
\end{aligned}$$

And so by MLE theory, we know that

$$\sqrt{n}(\widehat{\alpha}_n - \alpha_0) \to_d \mathcal{N}\left(0, \frac{\alpha_0^2(\alpha_0+1)^2}{\alpha_0^2 + (\alpha_0+1)^2}\right) = \mathcal{N}(0, \sigma_1^2)$$

(b) Suppose now that $X_1, \dots, X_n$ are fixed and known, i.e., $(X_1, \dots, X_n) = (x_1, \dots, x_n)$, and we observe the sample of independent observations $Y_1, \dots, Y_n$, where, for $i = 1, \dots, n$, Y_i is drawn from the conditional distribution of Y_i given $X_i = x_i$, and where the unconditional joint density of (X_i, Y_i) is given above. Show that for $i = 1, \dots, n$, the density of Y_i given $X_i = x_i$ is

$$\tilde{f}_i(y; \alpha) = (\alpha + 1)(1 + x_i)^{-1} \left(1 + \frac{y}{1 + x_i}\right)^{-(\alpha+2)}, y > 0.$$

Solution

To derive the conditional distribution, we have

$$\begin{aligned} \tilde{f}_i(y; \alpha) \equiv P(y|x; \alpha) &= \frac{P(x, y|\alpha)}{P(x|\alpha)} \\ &= \frac{P(x, y|\alpha)}{\int_y P(x, y|\alpha)\, dy} \end{aligned}$$

To obtain the marginal distribution of X,

$$\begin{aligned} P(x|\alpha) &= \int_y P(x, y|\alpha)\, dy \\ &= \int_0^\infty \alpha(\alpha + 1)(1 + x + y)^{-(\alpha+2)} dy \\ &= \alpha(\alpha + 1) \int_0^\infty (1 + x + y)^{-(\alpha+2)} dy \\ &= \alpha(\alpha + 1) \left(\frac{(1 + x + y)^{-(\alpha+2)+1}}{-(\alpha + 2) + 1} \right) \Bigg|_0^\infty \\ &= -\alpha \left((1 + x + y)^{-(\alpha+2)+1} \right) \Big|_0^\infty \\ &= -\alpha \left(0 - (1 + x)^{-(\alpha+2)+1} \right) \\ &= \alpha (1 + x)^{-(\alpha+2)+1} \end{aligned}$$

Plugging this result into the conditional distribution derivation,

$$\begin{aligned} \tilde{f}_i(y; \alpha) &= \frac{\alpha(\alpha + 1)(1 + x + y)^{-(\alpha+2)}}{\alpha (1 + x)^{-(\alpha+2)+1}} \\ &= (\alpha + 1)(1 + x)^{-1} \left(1 + \frac{y}{1 + x}\right)^{-(\alpha+2)} \end{aligned}$$

(c) In the setting of (b), verify that the maximum likelihood estimator, $\tilde{\alpha}_n$, has the following properties:

(i) $\tilde{\alpha}_n$ exists, is unique, and can be expressed in explicit closed form.

Solution

The joint likelihood and log likelihood are

$$\begin{aligned}
P(\boldsymbol{y}|\boldsymbol{x};\alpha) &= \prod_{i=1}^{n} P(y_i|x_i;\alpha) \\
&= \prod_{i=1}^{n}(\alpha+1)(1+x_i)^{-1}\left(1+\frac{y_i}{1+x_i}\right)^{-(\alpha+2)} \\
&= (\alpha+1)^n \prod_{i=1}^{n}(1+x_i)^{-1}\left(1+\frac{y_i}{1+x_i}\right)^{-(\alpha+2)} \\
l_n(\alpha) \equiv \log\left(P(\boldsymbol{y}|\boldsymbol{x};\alpha)\right) &= n\log(\alpha+1) - \sum_{i=1}^{n}\left[\log(1+x_i) + (\alpha+2)\log\left(1+\frac{y_i}{1+x_i}\right)\right]
\end{aligned}$$

To obtain the MLE, maximize the log likelihood.

$$\begin{aligned}
\frac{\partial l_n(\alpha)}{\partial \alpha} &= \frac{n}{\alpha+1} - \sum_{i=1}^{n}\log\left(1+\frac{y_i}{1+x_i}\right) = 0 \\
&\Leftrightarrow \quad \tilde{\alpha}_n = \frac{n}{\sum_{i=1}^{n}\log\left(1+\frac{y_i}{1+x_i}\right)} - 1 \equiv g(\tilde{\mu}_n)
\end{aligned}$$

(ii) $\tilde{\alpha}_n \to \alpha_0$, almost surely. Hint: Consider $U_i = Y_i/(1+x_i)$.

Solution

First, by the **SLLN** we know that $\frac{1}{n}\sum_{i=1}^{n}\log\left(1+\frac{y_i}{1+x_i}\right) \to_{a.s.} E\left[\log\left(1+\frac{y_i}{1+x_i}\right)\right]$. If we're interested in knowing the expectation, notice that the density belongs to the exponential family. All we need is the first derivative of the cumulant function obtained from the density in canonical form.

$$\begin{aligned}
P(y|x;\alpha) &= \exp\left\{\log(\alpha+1) - \log(1+x) - (\alpha+2)\log\left(1+\frac{y}{1+x}\right)\right\} \\
&= \exp\left\{-\alpha\log\left(1+\frac{y}{1+x}\right) + \log(\alpha+1) - 2\log\left(1+\frac{y}{1+x}\right) - \log(1+x)\right\} \\
&\quad \text{Note: Let } \theta \equiv -\alpha \Rightarrow b(\theta) = -\log(1-\theta) \Rightarrow \dot{b}(\theta) = \frac{1}{1-\theta} = \frac{1}{1+\alpha} = E\left[\log\left(1+\frac{y}{1+x}\right)\right]
\end{aligned}$$

And so $\tilde{\mu}_n \equiv \frac{1}{n}\sum_{i=1}^{n}\log\left(1+\frac{y_i}{1+x_i}\right) \to_{a.s.} \frac{1}{1+\alpha}$. And second, by **CMT**, $\tilde{\alpha}_n = g(\tilde{\mu}_n) \to_{a.s.} g(\mu) = \frac{1}{\frac{1}{1+\alpha_0}} - 1 = \alpha_0$.

(iii) $\sqrt{n}(\tilde{\alpha}_n - \alpha_0)$ is asymptotically normal with mean zero and variance $\sigma_2^2 = h(\alpha_0)$, and give the form of h.

Solution

Using the MLE approach,

$$\begin{aligned}
\frac{\partial^2 l_n(\alpha)}{\partial \alpha^2} &= -\frac{n}{(\alpha+1)^2} \\
I_n(\alpha) \equiv E\left[-\frac{\partial^2 l_n(\alpha)}{\partial \alpha^2}\right] &= \frac{n}{(\alpha+1)^2} \\
I(\alpha) &= \frac{1}{n}I_n(\alpha) = \frac{1}{(\alpha+1)^2}
\end{aligned}$$

And so we see that $\sqrt{n}(\tilde{\alpha}_n - \alpha_0) \to_d \mathcal{N}(0, (\alpha_0+1)^2)$. And so the form of h is $h(x) = (x+1)^2$.

(d) What is the asymptotic relative efficiency of $\tilde{\alpha}_n$ to $\widehat{\alpha}_n$?

Solution

When comparing the asymptotic variances from the joint likelihood and conditional likelihood, we have

$$\begin{aligned}
\frac{\sigma_2^2}{\sigma_1^2} &= \frac{(\alpha_0+1)^2}{\left(\frac{\alpha_0^2(\alpha_0+1)^2}{\alpha_0^2+(\alpha_0+1)^2}\right)} = \frac{\alpha_0^2+(\alpha_0+1)^2}{\alpha_0^2} \\
&= 1 + \left(1+\frac{1}{\alpha_0}\right)^2
\end{aligned}$$

The MLE derived from the joint likelihood has the smaller asymptotic variance compared to the conditional. This makes sense intuitively because conditioning on X results in a loss of information.

5.1.2 Question 2, incomplete

2. Consider the following:

   (a) For each $\theta_0 \in \Theta$, let T_{θ_0} be a test of $H_0 : \theta = \theta_0$ (versus some H_1) with significance level α and acceptance region $A(\theta_0)$. For each y in the range of the random variable Y, define

   $$C(y) = \{\theta : y \in A(\theta)\}.$$

   Show that $C(Y)$ is a level $1 - \alpha$ confidence set for θ.

   Solution

   The goal is to show that $P(\theta \in C(Y)) = 1 - \alpha$.

   Notice that

   $$\begin{aligned} P(\theta \in C(Y)) &= P(y \in A(\theta)) \\ &= 1 - P(y \notin A(\theta)) \\ &= 1 - P(\text{Reject } H_0 | H_0 \text{ true}) \\ &= 1 - \alpha \end{aligned}$$

(b) Suppose $X_1, \ldots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$, where $\sigma^2 = \gamma\mu^2$ and $-\infty < \mu < \infty$ and $\gamma > 0$ are both unknown scalar parameters, and $\mu \neq 0$. Using part (a), derive a confidence set for γ with confidence coefficient $1-\alpha$ by inverting the acceptance region of the likelihood ratio test for testing $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \neq \gamma_0$.

Solution

The joint likelihood and log likelihood are

$$
\begin{aligned}
P\left(\boldsymbol{x}|\mu,\gamma\right) &= \prod_{i=1}^{n} P\left(x_i|\mu,\gamma\right) \\
&= \prod_{i=1}^{n} \left(2\pi\gamma\mu^2\right)^{-1/2} \exp\left\{-\frac{(x_i-\mu)^2}{2\gamma\mu^2}\right\} \\
&= \left(2\pi\gamma\mu^2\right)^{-n/2} \exp\left\{-\frac{1}{2\gamma\mu^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
l_n(\mu,\gamma) \equiv \log\left(P\left(\boldsymbol{x}|\mu,\gamma\right)\right) &= -\frac{n}{2}\log\left(2\pi\gamma\mu^2\right) - \frac{1}{2\gamma\mu^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&\propto -\frac{n}{2}\left[\log\left(\gamma\right) + \log\left(\mu^2\right)\right] - \frac{1}{2\gamma\mu^2}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}
$$

We need to maximize this under Θ_0 and $\Theta \equiv \Theta_0 \cup \Theta_1$. Under Θ_0,

$$
\begin{aligned}
l_n(\mu,\gamma_0) &\propto -\frac{n}{2}\left[\log\left(\gamma_0\right) + \log\left(\mu^2\right)\right] - \frac{1}{2\gamma_0\mu^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
\frac{\partial l_n(\mu,\gamma_0)}{\partial\mu} &= -\frac{n}{2}\frac{2\mu}{\mu^2} - \frac{1}{2\gamma_0}\cdot\frac{\mu^2\cdot -2\sum_i(x_i-\mu) - 2\mu\cdot\sum_i(x_i-\mu)^2}{\mu^4} = 0 \\
&\Leftrightarrow\ \frac{n}{\mu} = \frac{\mu\sum_i(x_i-\mu) + \sum_i(x_i-\mu)^2}{\gamma_0\mu^3} \\
&\Leftrightarrow\ 0 = n\gamma_0\mu^2 + \left(\sum_i x_i\right)\mu - \sum_i x_i^2 \\
&\Leftrightarrow\ \tilde{\mu}_n = \frac{-\sum_i x_i \pm \sqrt{\left(\sum_i x_i\right)^2 + 4n\gamma_0\sum_i x_i^2}}{2n\gamma_0} \\
&\Rightarrow\ \tilde{\mu}_n = \left\{\begin{array}{ll} \tilde{\mu}_{n+}(\gamma_0) & l_n(\tilde{\mu}_{n+}(\gamma_0),\gamma_0) > l_n(\tilde{\mu}_{n-}(\gamma_0),\gamma_0) \\ \tilde{\mu}_{n-}(\gamma_0) & l_n(\tilde{\mu}_{n-}(\gamma_0),\gamma_0) > l_n(\tilde{\mu}_{n+}(\gamma_0),\gamma_0) \end{array}\right.
\end{aligned}
$$

Under the full parameter space,

$$
\begin{aligned}
\frac{\partial l_n(\mu,\gamma)}{\partial\mu} &= -\frac{n}{\mu} - \frac{1}{\gamma}\left[-\frac{\sum_i x_i^2}{\mu^3} + \frac{\sum_i x_i}{\mu^2}\right] = 0 \\
&\Leftrightarrow\ 0 = n\gamma\mu^2 + \sum_i x_i\mu - \sum_i x_i^2 \\
\frac{\partial l_n(\mu,\gamma)}{\partial\gamma} &= -\frac{n}{2\gamma} + \frac{1}{\gamma^2}\left[\frac{\sum_i x_i^2}{2\mu^2} - \frac{\sum_i x_i}{\mu} + \frac{n}{2}\right] = 0 \\
&\Leftrightarrow\ n\gamma\mu^2 = \sum_i x_i^2 - 2\mu\sum_i x_i + n\mu^2 \\
\Leftrightarrow \sum_i x_i^2 - \mu\sum_i x_i &= \sum_i x_i^2 - 2\mu\sum_i x_i + n\mu^2 \\
&\Rightarrow\ \widehat{\mu} = \frac{1}{n}\sum_i X_i \equiv \bar{X} \\
&\Rightarrow\ \widehat{\gamma} = \frac{\sum_i(X_i-\bar{X})^2}{n\bar{X}^2} = \frac{1}{n\widehat{\mu}^2}\sum_i(X_i-\bar{X})^2
\end{aligned}
$$

The LRT statistic is therefore

$$
\Lambda_n = \frac{(2\pi\gamma_0\tilde{\mu}_n^2)^{-n/2}\exp\left\{-\frac{\sum_i(x_i-\tilde{\mu}_n)^2}{2\gamma_0\tilde{\mu}_n^2}\right\}}{(2\pi\widehat{\gamma}\widehat{\mu}^2)^{-n/2}\exp\left\{-\frac{n}{2}\right\}}
$$

The LRT is

$$
\phi(\boldsymbol{x}) = \left\{\begin{array}{ll} 1 & \Lambda_n < c_\alpha \\ 0 & \Lambda_n \geq c_\alpha \end{array}\right.
$$

where $\alpha = E_{H_0}\left[\phi(\boldsymbol{x})\right] = P_{H_0}\left(\Lambda_n < c_\alpha\right)$. To get the confidence region, invert the test and accept H_0 when $\{\gamma : \Lambda_n(\gamma) \geq c_\alpha\}$.

(c) Under the set-up of part (b), show that a UMP test does not exist for testing $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \geq \gamma_0$.

Solution

We know that a UMP test does not exist because we have a multiparameter distribution. Another way to see this is that through using the Neyman Pearson Lemma, the rejection region depends on μ which is unknown.

(d) Suppose $X_1, \dots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are scalar parameters. Assume that $\theta = (\mu, \phi)$, where $\phi = \sigma^2$, is unknown. Derive $1 - \alpha$ asymptotically correct confidence sets for μ by

(i) inverting acceptance regions for (1) the likelihood ratio test, (2) the Wald test, and (3) the score test.

Solution

Starting with the joint and log likelihood, we have

$$\begin{aligned} P\left(\boldsymbol{x}|\mu, \sigma^2\right) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{(\boldsymbol{x} - \boldsymbol{J}\mu)^T (\boldsymbol{x} - \boldsymbol{J}\mu)}{2\sigma^2}\right\} \\ l_n(\mu, \sigma^2) &\propto -\frac{n}{2}\log\left(\sigma^2\right) - \frac{(\boldsymbol{x} - \boldsymbol{J}\mu)^T (\boldsymbol{x} - \boldsymbol{J}\mu)}{2\sigma^2} \end{aligned}$$

After maximizing under the constrained parameter space and the full, we have

$$\begin{aligned} \tilde{\mu} &= \mu_0 \\ \tilde{\sigma}^2 &= \frac{(\boldsymbol{x} - \boldsymbol{J}\mu_0)^T (\boldsymbol{x} - \boldsymbol{J}\mu_0)}{n} \\ \widehat{\mu} &= \frac{\boldsymbol{J}^T\boldsymbol{x}}{n} \\ \widehat{\sigma}^2 &= \frac{(\boldsymbol{x} - \boldsymbol{J}\widehat{\mu})^T (\boldsymbol{x} - \boldsymbol{J}\widehat{\mu})}{n} = \frac{1}{n}\boldsymbol{x}^T (\boldsymbol{I} - \boldsymbol{M}_J)\, \boldsymbol{x} \\ & \text{Note: } \boldsymbol{M}_J \equiv \frac{1}{n}\boldsymbol{J}\boldsymbol{J}^T \end{aligned}$$

For the LRT,

$$\begin{aligned} \Lambda_n &= \frac{(2\pi\tilde{\sigma}^2)^{-n/2} \exp\left\{-\frac{n}{2}\right\}}{(2\pi\widehat{\sigma}^2)^{-n/2} \exp\left\{-\frac{n}{2}\right\}} < k \\ \Leftrightarrow & \quad F^* \equiv \frac{\frac{n(\bar{x} - \mu_0)^2}{\sigma^2}}{\frac{(n-1)S^2}{\sigma^2}} > k \\ & \quad \text{Note: } F^* \sim^{H_0} F(1, n-1) \\ \phi(\boldsymbol{x}) &= \begin{cases} 1 & F^* > F(1, n-1, 1-\alpha) \\ 0 & F^* \leq F(1, n-1, 1-\alpha) \end{cases} \end{aligned}$$

with confidence set $\{\mu : F^* \leq F(1, n-1, 1-\alpha)\}$.

For the Wald Test, the confidence set is $\left\{\mu : \frac{n(\bar{x} - \mu)^2}{\widehat{\sigma}^2} \leq \chi^2_{1,1-\alpha}\right\}$ when applying asymptotics.

For the Score Test, the confidence set is $\left\{\mu : \frac{n(\bar{x} - \mu)^2}{\tilde{\sigma}^2} \leq \chi^2_{1,1-\alpha}\right\}$ when applying asymptotics.

(ii) Are these sets always intervals? Justify your answer.

Solution

If the null hypothesis is testing $H_0 : \theta = \theta_0$, then the confidence sets would be regions rather than intervals.

Hint: Using the notation from part (a), if $\lim_{n\to\infty} P(\theta \in C(Y)) = 1 - \alpha$ for all θ, then $C(Y)$ is a $1 - \alpha$ asymptotically correct confidence set for θ.

(e) Now, under the set-up in part (d), suppose we do not make any distributional assumptions about $X_i, i = 1, \ldots, n$ but only assume that $X_1, \ldots, X_n$ are i.i.d with mean μ and variance σ^2 with finite fourth moment. Derive an asymptotically correct confidence interval for $\theta = \mu/\sigma$.

Solution

5.1.3 Question 3, incomplete

3. Consider an r-sided coin and suppose that on each flip of the coin exactly one of the sides appears: side i with probability P_i, $\sum_{i=1}^{r} P_i = 1$. For given positive integers $n_1, n_2, \ldots, n_r$, let N_i denote the number of flips required until side i has appeared for the n_i time, $i = 1, \ldots, r$, and let $N = \min_{i=1,\ldots,r} N_i$. Thus, N is the number of flips required until some side i has appeared n_i times, for $i = 1, \ldots, r$.

   (a) Derive the marginal distribution of N_i, for $i = 1, \ldots, r$.
   Solution

   (b) Prove whether or not N_i, $i = 1, \ldots, r$ are independent random variables.
   Solution

   Now, suppose that the flips are performed at random times generated by a Poisson process with rate $\lambda = 1$. Let T_i denote the time until side i has appeared for the n_i time, $i = 1, \ldots, r$, and let $T = \min_{i=1,\ldots,r} T_i$.

   (c) Derive the marginal distribution of T_i, for $i = 1, \ldots, r$.
   Solution

   (d) Prove whether or not $T_i, i = 1, \ldots, r$ are independent random variables.
   Solution

   (e) Derive the density of T.
   Solution

   (f) Obtain an expression for $E[N]$ as a function of $E[T]$.
   Solution

5 Part 2

5.2.1 Question 1, incomplete

1. Suppose that y_{ij} for $i = 1, \ldots, m$ and $j = 1, 2$ follow a Poisson mixed effects model:

$$y_{ij}|u_1, \ldots, u_m \sim \text{Poisson}(\lambda_{ij}), \log(\lambda_{ij}) = x_{ij}^T \beta + u_i,$$

where x_{ij} is a $p \times 1$ covariate vector, β is an unknown $p \times 1$ parameter vector, and $u_1, \cdots, u_m$ are independent and identically distributed. Let $z_i = \exp(u_i)$ for all i and define γ to be the coefficient of variation of z_i, that is,

$$\gamma = \frac{\sqrt{\text{Var}(z_i)}}{E[z_i]}.$$

(a) Show that $\text{Var}(y_{ij}|x_{ij}) = \mu_{ij}(1 + \gamma^2 \mu_{ij})$ and $\text{Cov}(y_{ij}, y_{ik}|x_{ij}, x_{ik}) = \gamma^2 \mu_{ij} \mu_{ik}$ for $j \neq k$, where $\mu_{ij} = E[y_{ij}|x_{ij}]$.
Solution

(b) It is assume that the $z_i \sim \text{Gamma}(\alpha, 1/\alpha)$, for some unknown scalar $\alpha > 0$. Calculate μ_{ij} and γ^2. Write down the likelihood for (β, α) and show that it can be expressed in closed form using the gamma function.
Solution

(c) Suggest an algorithm to calculate the maximum likelihood estimator of $\theta = (\beta, \alpha)$, denoted by $\widehat{\theta}_M = \left(\widehat{\beta}_M, \widehat{\alpha}_M\right)$. Derive the asymptotic distribution of $\widehat{\beta}_M$. Please give the explicit form of the asymptotic covariance of $\widehat{\beta}_M$.
Solution

(d) Suppose that $\widehat{\beta}_E$ is the solution of a set of estimating equations for β given by

$$\sum_{i=1}^{m} \frac{\partial \mu_i}{\partial \beta}^T (y_i - \mu_i) = 0_p,$$

where $\mu_i = (\mu_{i1}, \mu_{i2})^T, y_i = (y_{i1}, y_{i2})^T$, and 0_p is a $p \times 1$ vector of zeros. Derive the asymptotic distribution of $\widehat{\beta}_E$.
Solution

(e) Rigorously compare the asymptotic covariances of $\widehat{\beta}_E$ and $\widehat{\beta}_M$. Which estimator is more efficient? Are there scenarios where the asymptotic covariances are equal?
Solution

5.2.2 Question 2, incomplete

2. Consider the linear model
$$Y = X\beta + Z\gamma + \epsilon,$$
where Y is $n \times 1$, X is $n \times p$ of rank p, Z is $n \times q$ of rank q, β is an unknown $p \times 1$ parameter vector, γ is $q \times 1$, $\epsilon \sim \mathcal{N}_n(0, R)$, $\gamma \sim \mathcal{N}_q(0, D)$, R and D are positive definite matrices, ϵ and γ are independent, and $\mathcal{N}_n(a, b)$ is an n variate normal random variable with mean vector a and covariance matrix b.

   (a) For known R and D, the distribution of $Y|X, \gamma \sim \mathcal{N}(X\beta + Z\gamma, R)$. Derive the marginal distribution of $Y|X$.
   Solution

   (b) In the following, continue to assume that R and D are known and treat γ as an unknown parameter in $Y|X, \gamma$.

      (i) Show that the predictor of γ given by $\widehat{\gamma} = DZ^T V^{-1}(Y - X\widehat{\beta})$ satisfies the conditional likelihood equations for (β, γ), where $\widehat{\beta}$ is the MLE of β and $V = ZDZ^T + R$.
      Solution

      (ii) Derive the exact distribution of $\widehat{\gamma}$.
      Solution

      (iii) Show that $\widehat{\gamma}$ is the best linear unbiased predictor of γ.
      Solution

   (c) Now suppose that R is of the form $R = \sigma^2 I_n$, where I_n is the $n \times n$ identity matrix, and β, σ^2, and D are unknown. Devise a detailed EM algorithm for jointly estimating (β, σ^2, D).
   Solution

   (d) Next, consider the case that D, R, and β are unknown and that R has a general structure. Define $A = I_n - M$, where M is the orthogonal projection operator on the column space of X, and write $W = B^T Y$ where $A = BB^T$ and $B^T B = I_n$. Consider estimation of the unknown parameters using the marginal distribution of $Y|X$ in (a).

      (i) Let $\widehat{\beta}$ denote the MLE of β when (D, R) are fixed. Show that $\mathrm{Cov}(W, \widehat{\beta}) = 0$.
      Solution

      (ii) Use the result in (i) to derive the density of W.
      Solution

      (iii) Devise a joint estimation scheme for (D, R, β) using (i) and (ii).
      Solution

   (e) For $i = 1, \ldots, n$, consider the linear model $y_i = x_i^T \beta + \gamma_i + \epsilon_i$, where y_i is a scalar random variable, x_i is a $p \times 1$ vector of covariates, β is the $p \times 1$ regression parameter, γ_i are i.i.d. $\mathcal{N}_1(0, \tau^2)$ and ϵ_i are independent $\mathcal{N}_1(0, \sigma^2 \exp(\lambda^T x_i^*))$, where λ is a $q \times 1$ unknown parameter vector, $q < p$, and x_i^* is a subset of the covariate vector which does *not* contain a constant (intercept) term, and σ^2 and τ^2 are scalar parameters. Assume further that γ_i is independent ϵ_i, for $i = 1, \ldots, n$. The parameters in $(\beta, \tau^2, \sigma^2, \lambda)$ are unknown. Using the marginal distribution of $y_i|x_i$, derive the score test for the hypothesis $H_0 : \lambda = 0$ and state its asymptotic distribution under H_0.
   Solution

5.2.3 Question 3, incomplete

3. We consider N independent random variables, denoted by $Y_1, \ldots, Y_N$, from a population of N subjects. We assume $Y_i = \beta x_i + \mathcal{N}(0, \sigma^2)$, where $x_1, \ldots, x_N$ are known positive constants and β and σ^2 are unknown scalar parameters. To estimate β and σ^2, we take a random sample of $Y_i, i = 1, \ldots, N$, from these N subjects. For $k = 1, \ldots, N$, we observe x_k and R_k, the indicator variable for whether or not Y_k for the kth subject is selected. We assume $R_1, \ldots, R_N$ are mutually independent and independent of (Y_i, x_i), $i = 1, \ldots, N$. For $k = 1, \ldots, N$, assume $P(R_k = 1) = \pi_k$ for some known constant $\pi_k \in (0, 1)$. Thus, only Y_i from selected subjects with $R_i = 1$ are observable.

   (a) Write the likelihood function for the observed data.
   Solution

   (b) Compute the maximum likelihood estimator for β and σ^2, denoted by $\widehat{\beta}$ and $\widehat{\sigma}^2$ respectively. If the sample size is zero, define $\widehat{\beta} = 0$ and $\widehat{\sigma}^2 = 0$.
   Solution

   (c) Derive the mean and variance of $\widehat{\beta}$.
   Solution

   (d) Calculate the distribution of $\widehat{\beta}$.
   Solution

   (e) Construct a confidence interval of level $(1 - \alpha)$ for β based on the conditional distribution of $\widehat{\beta}$ given $R_1, \ldots, R_N$.
   Solution

   (f) Define $\tilde{\beta} = \left\{\sum_{i=1}^N \frac{R_i}{\pi_i} Y_i\right\} \left\{\sum_{i=1}^N x_i\right\}^{-1}$. Show $\tilde{\beta}$ is an unbiased estimator for β and derive the variance of $\tilde{\beta}$.
   Solution

   (g) Find the optimal π_i to minimize $\operatorname{var}\left(\tilde{\beta}\right)$ under the condition that the expected sample size is fixed at n, i.e., $\sum_{i=1}^N \pi_i = n$.
   Solution

   (h) For any given finite function $g(\cdot)$ and π_i, $i = 1, \ldots, N$, show

   $$\tilde{\beta}(g) \equiv \frac{\sum_{i=1}^N g(x_i) + \sum_{i=1}^N \frac{R_i}{\pi_i} \{Y_i - g(x_i)\}}{\sum_{i=1}^N x_i}$$

   is unbiased for β and calculate its variance.
   Solution

   (i) For given $\pi_i, i = 1, \ldots, N$, determine the optimal $g(\cdot)$ minimizing the variance of $\tilde{\beta}(g)$. Suggest a method for estimating the optimal g using the observed data.
   Solution

6 Theory 2014

6 Part 1

6 Part 2