STAT 30100 : Mathematical Statistics-1

Winter 2020

•

Homework 7

•

Solutions by

Jinhong Du

12243476

STAT 30100, Homework 7

1. (Casella and Berger Problem 6.30) Let $X_1, \ldots, X_n$ be a random sample from the pdf $f(x|\mu) = e^{-(x-\mu)}$, where $-\infty < \mu < x < \infty$.

(a) Show that $X_{(1)} = \min_i X_i$ is a complete sufficient statistic.

Proof. The joint density of $\boldsymbol{X}$ is $f_\mu(\boldsymbol{x}) = e^{-\sum_{i=1}^n (x_i-\mu)} \mathbb{1}_{x_{(1)}>\mu}$. Let $\Theta_{\boldsymbol{x}} = \{\mu : f_\mu(x) > 0\} = \{\mu : x_{(1)} > \mu\}$ for $\boldsymbol{x} \in \mathcal{X}$. Since for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, $\Theta_{\boldsymbol{x}} = \Theta_{\boldsymbol{y}}$ and $\frac{f_\mu(\boldsymbol{x})}{f_\mu(\boldsymbol{y})} = e^{-\sum_{i=1}^n (x_i - y_i)} \frac{\mathbb{1}_{x_{(1)}>\mu}}{\mathbb{1}_{y_{(1)}>\mu}}$ is constant as a function of μ if and only if $x_{(1)} = y_{(1)}$. Then by Lehmann and Scheffé Theorem, $X_{(1)}$ is a sufficient statistic.

The density for $X_{(1)}$ is given by

$$
\begin{aligned}
F_{X_{(1)}}(t) &= 1 - [1 - F(x)]^n = [1 - e^{n\mu} e^{-nx}] \\
f_{X_{(1)}}(t) &= n e^{n\mu} e^{-nx} \mathbb{1}_{\{x>\mu\}}
\end{aligned}
$$

Suppose that g is a function such that $\mathbb{E}_\mu[g(X_{(1)})] = 0 \ \forall\ \mu \in \mathbb{R}$, then

$$
\begin{aligned}
0 = \frac{\mathrm{d}}{\mathrm{d}\mu}\mathbb{E}_\mu[g(X_{(1)})] &= \frac{\mathrm{d}}{\mathrm{d}\mu} \int_\mu^\infty g(t) n e^{n(\mu-t)} \mathrm{d}t \\
&= e^{n\mu} \frac{\mathrm{d}}{\mathrm{d}\mu} \int_\mu^\infty g(t) n e^{-nt} \mathrm{d}t + \frac{\mathrm{d}e^{n\mu}}{\mathrm{d}\mu} \cdot \int_\mu^\infty g(t) n e^{n(\mu-t)} \mathrm{d}t \\
&= -g(\mu),
\end{aligned}
$$

which implies that $\mathbb{P}(g(X_{(1)}) = 0) = 1$, $\forall\ \mu \in \mathbb{R}$. So $X_{(1)}$ is complete.
Therefore, $X_{(1)}$ is a complete sufficient statistic. □

(b) Use Basu's Theorem to show that $X_{(1)}$ and S^2 are independent.

Since X_i's are in the location family with standard density $f_0(x) = e^{-x}\mathbb{1}_{(0,\infty)}$. There exists an iid sample $Z_1, \ldots, Z_n$ from $f_0(x)$ such that $X_i = Z_i + \mu$. So the distribution of S^2 is given by

$$
\begin{aligned}
\mathbb{P}(S^2 \le t) &= \mathbb{P}\left(\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 \le t\right) \\
&= \mathbb{P}\left(\frac{1}{n-1}\sum_{i=1}^n (Z_i - \overline{Z})^2 \le t\right),
\end{aligned}
$$

which is free of μ. So S^2 is an ancillary statistic. By Basu's Theorem, $X_{(1)}$ and S^2 are independent.

2. (Casella and Berger Problem 7.6) Let $X_1, \dots, X_n$ be a random sample from the pdf

$$f(x|\theta) = \theta x^{-2}, \quad 0 < \theta \le x < \infty.$$

(a) What is a sufficient statistic for θ?

Since the joint density of $\boldsymbol{X}$ is $f_\theta(\boldsymbol{x}) = \theta^n \mathbb{1}_{x_{(1)} \ge \theta} \cdot \prod_{i=1}^n x_i^{-2n}$, by Fisher-Neymann Factorization Theorem, we have $X_{(1)}$ is a sufficient statistic for θ.

(b) Find the MLE of θ.

Since for all $\theta \in \mathbb{R}$,

$$L(\theta; \boldsymbol{x}) = f_\theta(\boldsymbol{x}) = \theta^n \mathbb{1}_{x_{(1)} \ge \theta} \cdot \prod_{i=1}^n x_i^{-2n} \le x_{(1)}^n \cdot 1 \cdot \prod_{i=1}^n x_i^{-2n} = x_{(1)}^n \prod_{i=1}^n x_i^{-2n}$$

with equality if and only if $\theta = x_{(1)}$, the MLE of θ is $\hat{\theta} = X_{(1)}$.

(c) Find the method of moments estimator of θ.

Since

$$\mathbb{E}(X_i^{-1}) = \int_\theta^\infty \theta x^{-3} \mathrm{d}x = -\frac{1}{2}\theta x^{-2}\Big|_\theta^\infty = \frac{1}{2}\theta^2,$$

the moments estimator of θ is $\sqrt{\frac{2}{n}\sum_{i=1}^n X_i^{-1}}$.

3. (Casella and Berger Problem 7.14) Let X and Y be independent exponential random variables, with

$$f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}, x > 0, \quad f(y|\mu) = \frac{1}{\mu} e^{-y/\mu}, y > 0.$$

We observe Z and W with

$$Z = \min(X, Y) \quad \text{and} \quad W = \begin{cases} 1 & \text{if } Z = X \\ 0 & \text{if } Z = Y \end{cases}.$$

In Exercise 4.26 the joint distribution of Z and W was obtained. Now assume that (Z_i, W_i), $i = 1, \dots, n$, are n iid observations. Find the MLEs of λ and μ.

The cdf of X and Y are $F(x|\lambda) = 1 - e^{-\frac{x}{\lambda}}$ and $F(y|\mu) = 1 - e^{-\frac{y}{\mu}}$ respectively.

$$\begin{aligned}
\mathbb{P}(Z \le z, W = 0) &= \mathbb{P}(Y \le z, Y \le X) \\
&= \int_0^z \mathbb{P}(X \ge y) f(y|\mu) \mathrm{d}y \\
&= \frac{1}{\mu}\int_0^z e^{-\frac{y}{\lambda}} e^{-\frac{y}{\mu}} \mathrm{d}y \\
&= \frac{\lambda}{\lambda + \mu}\left(1 - e^{-\frac{\lambda+\mu}{\lambda\mu} z}\right),
\end{aligned}$$

Solution (cont.)

$$f_{(Z,W)}(z,0) = \frac{1}{\mu} e^{-\frac{\lambda+\mu}{\lambda\mu} z}.$$

Similarly, we have $f_{(Z,W)}(z,1) = \frac{1}{\lambda} e^{-\frac{\lambda+\mu}{\lambda\mu} z}$. So the joint density of (Z,W) is given by

$$f_{(Z,W)}(z,w) = \begin{cases} \frac{1}{\lambda} e^{-\frac{\lambda+\mu}{\lambda\mu} z} & , \ w = 1 \\ \frac{1}{\mu} e^{-\frac{\lambda+\mu}{\lambda\mu} z} & , \ w = 0 \\ 0 & , \ \text{otherwise} \end{cases}.$$

The log-likelihood function of (Z_i, W_i) $i = 1, \ldots, n$ is

$$\begin{aligned} l(\lambda, \mu; \boldsymbol{z}, \boldsymbol{w}) &= \sum_{i=1}^{n} \ln \left[\left(\frac{1}{\lambda} e^{-\frac{\lambda+\mu}{\lambda\mu} z} \right)^{w_i} \left(\frac{1}{\mu} e^{-\frac{\lambda+\mu}{\lambda\mu} z} \right)^{1-w_i} \right] \\ &= -\sum_{i=1}^{n} \left\{ w_i \ln \lambda + w_i \left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z_i + (1-w_i) \ln \mu + (1-w_i) \left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z_i \right\} \\ &= -\sum_{i=1}^{n} \left[w_i \ln \lambda + (1-w_i) \ln \mu + \left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z_i \right] \end{aligned}$$

when $w_i = 0$ or 1, and $l(\lambda, \mu; \boldsymbol{z}, \boldsymbol{w}) = -\infty$ otherwise. By setting

$$\begin{aligned} \frac{\partial l(\lambda, \mu; \boldsymbol{z}, \boldsymbol{w})}{\partial \lambda} &= \sum_{i=1}^{n} \left[\frac{z_i}{\lambda^2} - \frac{w_i}{\lambda} \right] = 0 \\ \frac{\partial l(\lambda, \mu; \boldsymbol{z}, \boldsymbol{w})}{\partial \mu} &= \sum_{i=1}^{n} \left[\frac{z_i}{\mu^2} - \frac{1-w_i}{\mu} \right] = 0, \end{aligned}$$

we get

$$\begin{aligned} \lambda &= \frac{\sum_{i=1}^{n} z_i}{\sum_{i=1}^{n} w_i} \\ \mu &= \frac{\sum_{i=1}^{n} z_i}{n - \sum_{i=1}^{n} w_i}. \end{aligned}$$

So the MLE of λ and μ is $\hat{\lambda} = \frac{\sum_{i=1}^{n} Z_i}{\sum_{i=1}^{n} W_i}$ and $\hat{\mu} = \frac{\sum_{i=1}^{n} Z_i}{n - \sum_{i=1}^{n} W_i}$.

4. (Casella and Berger Problem 7.19) Suppose that the random variables $Y_1, \ldots, Y_n$ satisfy

$$Y_i = \beta x_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

where $x_1, \ldots, x_n$ are fixed constants, and $\epsilon_1, \ldots, \epsilon_n$ are iid $\mathcal{N}(0, \sigma^2)$, σ^2 unknown.

(a) Find a two-dimensional sufficient statistic for (β, σ^2).

Proof. We have $Y_i \sim \mathcal{N}(\beta x_i, \sigma)$ for $i = 1, \ldots, n$ and they are independent. The joint density of $\boldsymbol{Y}$ is given by

$$\begin{aligned} f_{(\beta,\sigma^2)}(\boldsymbol{y}) &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta x_i)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\beta^2}{2\sigma^2} \sum_{i=1}^{n} x_i^2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 + \frac{\beta}{\sigma^2} \sum_{i=1}^{n} x_i y_i)}. \end{aligned}$$

Solution (cont.)

Then by Fisher-Neymann Factorization Theorem, we have $(\sum_{i=1}^n Y_i^2, \sum_{i=1}^n x_i Y_i)$ is a two-dimensional sufficient statistic for (β, σ). $\square$

(b) Find the MLE of β, and show that it is an unbiased estimator of β.

The log-likelihood is

$$l(\beta, \sigma^2) = \frac{n}{2}\ln(2\pi\sigma^2) - \frac{\beta^2}{2\sigma^2}\sum_{i=1}^n x_i^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2 + \frac{\beta}{\sigma^2}\sum_{i=1}^n x_i y_i.$$

Setting

$$\frac{\partial l(\beta, \sigma^2)}{\partial \beta} = -\frac{\beta}{\sigma^2}\sum_{i=1}^n x_i^2 + \frac{1}{\sigma^2}\sum_{i=1}^n x_i y_i = 0,$$

we get $\beta = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$. Thus, the MLE of β is $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$.

(c) Find the distribution of the MLE of β.

Since $Y_i \sim \mathcal{N}(\beta x_i, \sigma)$ for $i = 1, \ldots, n$ and they are independent, we have $\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} \sim \mathcal{N}\left(\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2}\beta, \frac{\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}\sigma^2\right)$, i.e., $\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{1}{\sum_{i=1}^n x_i^2}\sigma^2\right)$

5. (Relation of Kullback-Leibler information to Fisher information.) Let $f_\theta(x)$ be the density of a one-parameter exponential family of distributions with the natural parameterization,

$$f_\theta(x) = h(x)c(\theta)\exp\{\theta T(x)\}.$$

Let θ_0 be an interior point of the natural parameter space. Find the Kullback-Leibler information number, $K(\theta_0 : \theta)$, for an arbitrary θ and show that as $\theta \to \theta_0$, $K(\theta_0 : \theta) \sim (\theta - \theta_0)^2 \mathcal{I}(\theta_0)/2$, where $\mathcal{I}(\theta)$ is Fisher information. (Note: $a \sim b$ means $\frac{a}{b} \to 1$.)

Proof.

$$\begin{aligned}
K(\theta_0 : \theta) &= \mathbb{E}_{\theta_0}\left[\log\left(\frac{f_{\theta_0}(X)}{f_\theta(X)}\right)\right] \\
&= \mathbb{E}_{\theta_0}\left[\log\left(\frac{c(\theta_0)}{c(\theta)}e^{(\theta_0-\theta)T(X)}\right)\right] \\
&= \mathbb{E}_{\theta_0}\left[\log c(\theta_0) - \log c(\theta) + (\theta_0 - \theta)T(X)\right] \\
&= \log c(\theta_0) - \log c(\theta) + (\theta_0 - \theta)\mathbb{E}_{\theta_0}\left[T(X)\right]
\end{aligned}$$

From Problem 8 of Homework 1, we know that $\frac{\partial \log[f_{\theta_0}(x)]}{\partial \theta_0} = T(x) - \mathbb{E}_{\theta_0}[T(X)]$, $\mathbb{E}_{\theta_0}[T(X)] = -\frac{\partial \log c(\theta_0)}{\partial \theta_0}$, and also,

$$\mathcal{I}(\theta_0) = \mathrm{Var}_{\theta_0}\left[\frac{\partial \log[f_{\theta_0}(X)]}{\partial \theta_0}\right] = \mathrm{Var}_{\theta_0}[T(X)] = -\frac{\partial^2 \log[f_{\theta_0}(x)]}{\partial \theta_0^2} = -\frac{\partial^2 \log c(\theta_0)}{\partial \theta_0^2}.$$

Solution (cont.)

Then we have

$$\begin{aligned}\lim_{\theta\to\theta_0}\frac{K(\theta_0:\theta)}{(\theta-\theta_0)^2\mathcal{I}(\theta_0)/2} &= \lim_{\theta\to\theta_0}\frac{-\frac{\partial\log c(\theta)}{\partial\theta}-\mathbb{E}_{\theta_0}[T(X)]}{(\theta-\theta_0)\mathcal{I}(\theta_0)}\\ &= \lim_{\theta\to\theta_0}\frac{-\frac{\partial^2\log c(\theta)}{\partial\theta^2}}{\mathcal{I}(\theta_0)}\\ &= \frac{\mathcal{I}(\theta_0)}{\mathcal{I}(\theta_0)}\\ &= 1,\end{aligned}$$

i.e., $K(\theta_0:\theta)\sim(\theta-\theta_0)^2\mathcal{I}(\theta_0)/2$. □

6. Let $X_1,\ldots,X_n$ be a sample from a distribution with density

$$f(x|\theta_1,\theta_2)=\begin{cases}(\theta_1/\theta_2)\,e^{-x/\theta_2} & \text{for} \quad x>0\\ ((1-\theta_1)/\theta_2)\,e^{x/\theta_2} & \text{for} \quad x<0\end{cases}$$

where $0\le\theta_1\le1$ and $\theta_2>0$.

(a) Show that (S,K) is sufficient for (θ_1,θ_2), where K is the number of positive X_i's, $K=\sum_{i=1}^n\mathbb{1}(X_i>0)$ and $S=\sum_{i=1}^n|X_i|$.

Proof. The joint density of $\boldsymbol{X}$ is given by

$$\begin{aligned}f_{(\theta_1,\theta_2)}(\boldsymbol{x}) &= \prod_{i=1}^n\left(\frac{\theta_1}{\theta_2}e^{-\frac{x}{\theta_2}}\mathbb{1}_{\{x_i>0\}}+\frac{1-\theta_1}{\theta_2}e^{-\frac{x}{\theta_2}}\mathbb{1}_{\{x_i<0\}}\right)\\ &= \left(\frac{\theta_1}{\theta_2}\right)^k e^{-\frac{1}{\theta_2}\sum_{i=1}^n x_i\mathbb{1}_{x_i>0}}\left(\frac{1-\theta_1}{\theta_2}\right)^{n-k}e^{-\frac{1}{\theta_2}\sum_{i=1}^n x_i\mathbb{1}_{x_i<0}}\\ &= \left(\frac{\theta_1}{\theta_2}\right)^k\left(\frac{1-\theta_1}{\theta_2}\right)^{n-k}e^{-\frac{1}{\theta_2}s}.\end{aligned}$$

where $k=\sum_{i=1}^n\mathbb{1}_{\{x_i>0\}}$ and $s=\sum_{i=1}^n|x_i|$. Then by Fisher-Neymann Factorization Theorem, we have (S,K) is sufficient for (θ_1,θ_2). □

(b) Find the MLE $(\hat\theta_1,\hat\theta_2)$ of (θ_1,θ_2).

The log-likelihood is given by

$$l(\theta_1,\theta_2;\boldsymbol{x})=k\ln\theta_1-n\ln\theta_2+(n-k)\ln(1-\theta_1)-\frac{1}{\theta_2}S.$$

Setting

$$\begin{aligned}\frac{\partial l(\theta_1,\theta_2;\boldsymbol{x})}{\partial\theta_1}&=\frac{k}{\theta_1}-\frac{n-k}{1-\theta_1}=0\\ \frac{\partial l(\theta_1,\theta_2;\boldsymbol{x})}{\partial\theta_2}&=-\frac{n}{\theta_2}+\frac{s}{\theta_2^2}=0,\end{aligned}$$

Solution (cont.)

we have $\theta_1 = \frac{k}{n}$ and $\theta_2 = \frac{s}{n}$. The second derivatives are

$$\begin{aligned}
\frac{\partial^2 l(\theta_1, \theta_2; \boldsymbol{x})}{\partial \theta_1 \partial \theta_2} &= \frac{\partial^2 l(\theta_1, \theta_2; \boldsymbol{x})}{\partial \theta_2 \partial \theta_1} = 0 \\
\frac{\partial^2 l(\theta_1, \theta_2; \boldsymbol{x})}{\partial \theta_1^2} &= -\frac{k}{\theta_1^2} - \frac{n-k}{(1-\theta_1)^2} = \frac{k(2\theta_1 - 1) - n\theta_1^2}{\theta_1^2 (1-\theta_1)^2} \\
\frac{\partial^2 l(\theta_1, \theta_2; \boldsymbol{x})}{\partial \theta_2^2} &= \frac{n}{\theta_2^2} - \frac{2s}{\theta_2^3}.
\end{aligned}$$

If $2\theta_1 - 1 \geq 0$, then $k(2\theta_1 - 1) - n\theta_1^2 \leq n(2\theta_1 - 1) - n\theta_1^2 = -k(\theta_1 - 1)^2 < 0$. If $2\theta_1 - 1 < 0$, then $k(2\theta_1 - 1) - n\theta_1^2 < 0$ since $k \geq 0$. So, $\frac{\partial^2 l(\theta_1, \theta_2; \boldsymbol{x})}{\partial \theta_1^2} < 0$ for all θ_1.
Also,

$$\det \left(\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{x})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) \Bigg|_{\boldsymbol{\theta} = \left(\frac{k}{n}, \frac{s}{n}\right)} = \frac{k(2\theta_1 - 1) - n\theta_1^2}{\theta_1^2 (1 - \theta_1^2)} \cdot \left(\frac{n}{\theta_2^2} - \frac{2s}{\theta_2^3} \right) \Bigg|_{\boldsymbol{\theta} = \left(\frac{k}{n}, \frac{s}{n}\right)} < 0.$$

So $\boldsymbol{\theta} = \left(\frac{k}{n}, \frac{s}{n}\right)$ is a local maximum of $l(\boldsymbol{\theta}; \boldsymbol{x})$.
For the boundary, if $\theta_1 = 0$, then $f(x|\boldsymbol{\theta}) = \frac{1}{\theta_2} e^{\frac{x}{\theta_2}} \mathbb{1}_{\{x<0\}}$. Then we will observe all $X_i < 0$, the MLE for $\boldsymbol{\theta}$ will be $\hat{\theta}_1 = 0 = \frac{K}{n}$ and $\hat{\theta}_2 = \frac{S}{n}$. Similarly, $\hat{\theta}_1 = 1$ and $\hat{\theta}_2 = \frac{S}{n}$ if $\theta_1 = 1 = \frac{K}{n}$.
Therefore, the MLE of (θ_1, θ_2) is $\hat{\theta}_1 = \frac{K}{n}$ and $\hat{\theta}_2 = \frac{S}{n}$.

(c) Find the Fisher information matrix $\mathcal{I}(\theta_1, \theta_2)$.

Let $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ such that $\theta_1 \in (0, 1)$ and $\theta_2 > 0$.

$$\frac{\partial^2 l(\boldsymbol{\theta}; \boldsymbol{X})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{bmatrix} -\frac{K}{\theta_1^2} - \frac{n-K}{(1-\theta_1)^2} & 0 \\ 0 & \frac{n}{\theta_2^2} - \frac{2S}{\theta_2^3} \end{bmatrix}$$

Since

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(X_i > 0) &= \int_0^\infty \frac{\theta_1}{\theta_2} e^{-\frac{x}{\theta_2}} \mathrm{d}x = -\theta_1 e^{-\frac{x}{\theta_2}} \Big|_0^\infty = \theta_1 \\
\mathbb{E}_{\boldsymbol{\theta}}(|X_i|) &= \int_0^\infty \frac{\theta_1}{\theta_2} e^{-\frac{x}{\theta_2}} x \mathrm{d}x - \int_{-\infty}^0 \frac{1-\theta_1}{\theta_2} e^{\frac{x}{\theta_2}} x \mathrm{d}x \\
&= -\theta_1 x e^{-\frac{x}{\theta_2}} \Big|_0^\infty + \int_0^\infty \theta_1 e^{-\frac{x}{\theta_2}} \mathrm{d}x \\
&\quad - (1-\theta_1) x e^{\frac{x}{\theta_2}} \Big|_{-\infty}^0 + \int_{-\infty}^0 (1-\theta_1) e^{\frac{x}{\theta_2}} \mathrm{d}x \\
&= -\theta_1 \theta_2 e^{-\frac{x}{\theta_2}} \Big|_0^\infty + (1-\theta_1)\theta_2 e^{\frac{x}{\theta_2}} \Big|_{-\infty}^0 \\
&= \theta_1 \theta_2 + (1-\theta_1)\theta_2 \\
&= \theta_2,
\end{aligned}$$

we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(K) &= \sum_{i=1}^n \mathbb{P}_{\boldsymbol{\theta}}(X_i > 0) = n\theta_1 \\
\mathbb{E}_{\boldsymbol{\theta}}(S) &= \sum_{i=1}^n \mathbb{E}_{\boldsymbol{\theta}}(|X_i|) = n\theta_2.
\end{aligned}$$

Solution (cont.)

Thus,

$$\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta};\boldsymbol{X})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right] = \begin{bmatrix} -\frac{n}{\theta_1(1-\theta_1)} & 0 \\ 0 & -\frac{n}{\theta_2^2} \end{bmatrix}.$$

Since $ae^{-ax} \le \frac{1}{x}e^{-1}$ for $a > 0$ and $x > 0$, $K(x) = \frac{1}{|x|}e^{-1}$ is the dominating function for $f(x|\theta_1,\theta_2)$ and $\frac{\partial f(x|\theta_1,\theta_2)}{\partial\theta_1}$. So we can pass the derivatives under the integral sign to get the Fisher information matrix

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta};\boldsymbol{X})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right] = \begin{bmatrix} \frac{n}{\theta_1(1-\theta_1)} & 0 \\ 0 & \frac{n}{\theta_2^2} \end{bmatrix}.$$

(d) Find the asymptotic joint distribution of $(\hat\theta_1, \hat\theta_2)$.

Proof. First consider the case when $\theta_1 \in (0,1)$. Since $\Theta = \{(\theta_1,\theta_2) : 0 < \theta_1 < 1, \theta_2 > 0\}$ is open, $\frac{\partial^2 l(\boldsymbol{\theta};\boldsymbol{X})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ exists and the derivatives and integral sign can be changed near (θ_1,θ_2), $\mathcal{I}(\boldsymbol{\theta})$ is positive definite, then

$$\sqrt{n}\begin{pmatrix} \hat\theta_1 - \theta_1 \\ \hat\theta_2 - \theta_2 \end{pmatrix} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathcal{I}_1(\boldsymbol{\theta})^{-1})$$

where $\mathcal{I}_1(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \theta_1(1-\theta_1) & 0 \\ 0 & \theta_2^2 \end{bmatrix}$. For $\theta_1 = 0$ or 1, similar result holds expect the asymptotic variance of $\hat\theta_1$ is 0. ☐

7. Suppose (X_i, Y_i), $i = 1,\dots,n$ are i.i.d. bivariate normal $\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}\right)$. Let r_n be the sample correlation of (X_i, Y_i), $i = 1,\dots,n$, that is,

$$r_n = \frac{\sum_i \left(X_i - \bar X\right)\left(Y_i - \bar Y\right)}{\sqrt{\Sigma_i \left(X_i - \bar X\right)^2}\sqrt{\sum_i \left(Y_i - \bar Y\right)^2}}.$$

(a) Find a minimum sufficient statistic for ρ. Is it complete? Prove your conclusion.

Proof. The density of (X_i, Y_i) is given

$$f(x_i,y_i) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x_i^2 - 2\rho x_i y_i + y_i^2}{2(1-\rho^2)}} = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x_i^2+y_i^2}{2(1-\rho^2)} + \frac{\rho}{1-\rho^2}x_i y_i}.$$

Let $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$. The joint density is

$$f_\rho(\boldsymbol{z}) = \frac{1}{(2\pi\sqrt{1-\rho^2})^n} e^{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^n (x_i^2+y_i^2) + \frac{\rho}{1-\rho^2}\sum_{i=1}^n x_i y_i}.$$

Let $\Theta_{\boldsymbol{z}} = \{\rho : f_\rho(\boldsymbol{z}) > 0\} = \mathbb{R} \setminus \{\pm 1\}$. Then for $\boldsymbol{z} \ne \boldsymbol{z}'$, $\Theta_{\boldsymbol{z}} = \Theta_{\boldsymbol{z}'}$ and $\frac{f_\rho(\boldsymbol{z})}{f_\rho(\boldsymbol{z}')} = e^{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^n (x_i^2+y_i^2-x_i'^2-y_i'^2) + \frac{\rho}{1-\rho^2}\sum_{i=1}^n (x_i y_i - x_i' y_i')}$ is constant as a function of ρ if and only if $\left(\sum_{i=1}^n (x_i^2+y_i^2), \sum_{i=1}^n x_i y_i\right) = \left(\sum_{i=1}^n (x_i'^2+y_i'^2), \sum_{i=1}^n x_i' y_i'\right)$. Thus by Lehmann-Scheffé Theorem, $T((\boldsymbol{X},\boldsymbol{Y})^\top) = \left(\sum_{i=1}^n (X_i^2+Y_i^2), \sum_{i=1}^n X_i Y_i\right)^\top$ is a minimum sufficient statistic for ρ.
Notice that $\mathbb{E}(X_i) = \mathbb{E}(Y_i) = 0$ and $\mathbb{E}(X_i^2) = \mathbb{E}(Y_i^2) = 1$. Let $g(x,y) = x + 2y - 2n$, then $\mathbb{E}_\theta\left[g\left(\sum_{i=1}^n (X_i^2+Y_i^2), \sum_{i=1}^n X_i Y_i\right)\right] = \mathbb{E}[\sum_{i=1}^n (X_i^2 + Y_i^2 + 2X_iY_i - 2)] = 0$. However, $\mathbb{P}(\sum_{i=1}^n (X_i +$

Solution (cont.)

$Y_i)^2 = 2n) < 1$ since the probability of $\sum_{i=1}^n (X_i + Y_i)^2 \le n$ must be positive. Therefore, it is not complete. $\square$

(b) Using the result from the first part, show, without actually maximizing the likelihood function, that r_n cannot be a MLE of ρ.

Proof. This density is in the exponential family. So the MLE must be a function of a sufficient statistic. However, we cannot express r_n as a function of $\left(\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_i Y_i\right)^\top$. $\square$

(c) Discuss how many modes the likelihood function has. Find the asymptotic variance of the MLE.

Proof. The log-likelihood is

$$l(\rho; \boldsymbol{z}) = -n\ln(2\pi) - \frac{n}{2}\ln(1-\rho^2) - \frac{1}{2(1-\rho^2)}\sum_{i=1}^n (x_i^2 + y_i^2) + \frac{\rho}{1-\rho^2}\sum_{i=1}^n x_i y_i$$

The score equation is

$$\frac{\partial l}{\partial \rho} = \frac{n\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}\sum_{i=1}^n (x_i^2 + y_i^2) + \frac{1+\rho^2}{(1-\rho^2)^2}\sum_{i=1}^n x_i y_i = 0,$$

i.e.

$$\rho(1-\rho^2) + (1+\rho^2)\frac{\sum_{i=1}^n x_i y_i}{n} - \rho\frac{1}{n}\sum_{i=1}^n (x_i^2 + y_i^2) = 0,$$

$(\rho \neq \pm 1)$ which is a cubic polynomial in ρ. Since the equation changes sign between $\rho = -1$ and $\rho = 1$, there is at least one root in that interval but there can be as many as three real ones both inside and outside the interval $[-1, 1]$. So there may be one or two modes of the likelihood function. Since the density is the exponential family, we can pass the derivative under the integral sign to get the Fisher information

$$\mathcal{I}_1(\theta) = -\mathbb{E}_\rho\left(\frac{\partial^2 l(; X_1, Y_1)}{\partial \rho^2}\right) = \frac{\rho^2 + 1}{(1-\rho^2)^2}$$

so the asymptotic variance of the MLE is $\frac{\rho^2+1}{(1-\rho^2)^2}$. $\square$

8. (Casella and Berger Problem 7.51) Gleser and Healy (1976) give a detailed treatment of the estimation problem in the $\mathcal{N}(\theta, a\theta^2)$ family, where a is a known constant (of which Exercise 7.50 is a special case). We explore a small part of their results here. Again let $X_1, \ldots, X_n$ be iid $\mathcal{N}(\theta, \theta^2)$, $\theta > 0$, and let $\overline{X}$ and cS be as in Exercise 7.50. For this model both $\overline{X}$ and cS are unbiased estimators of θ, where $c = \frac{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}$. Define the class of estimators

$$\mathcal{T} = \left\{T : T = a_1\bar{X} + a_2(cS)\right\},$$

where we do not assume that $a_1 + a_2 = 1$.

(a) Find the estimator $T \in \mathcal{T}$ that minimizes $\mathbb{E}_\theta(\theta - T)^2$; call it T^*.

Since $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \sim \mathcal{N}(\theta\mathbf{1}, \theta^2\boldsymbol{I})$, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\mathbf{1}^\top\boldsymbol{X} \sim \mathcal{N}(\theta, \frac{\theta^2}{n})$, $\frac{(n-1)S^2}{\theta^2} = \frac{1}{\theta^2}\sum_{i=1}^n(X_i - \overline{X})^2 = \frac{1}{\theta^2}\boldsymbol{X}^\top(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top)\boldsymbol{X} \sim \chi^2_{n-1}$ and $\frac{1}{\theta^2}(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top)\theta^2\boldsymbol{I}\frac{1}{n}\mathbf{1} = 0$, by Theorem 6 in the handout of quadractic forms, we have $\overline{X}$ and $\frac{(n-1)S^2}{\theta^2}$ are independent. So

$$\begin{aligned}
\mathbb{E}_\theta[(\theta - T)^2] &= \theta^2 - 2\theta\mathbb{E}_\theta(T) + \mathbb{E}_\theta(T^2) \\
&= \theta^2 - 2a_1\theta\mathbb{E}_\theta(\overline{X}) - 2a_2\theta\mathbb{E}_\theta(cS) + a_1^2\mathbb{E}_\theta(\overline{X}^2) + 2a_1a_2\mathbb{E}_\theta(\overline{X})\mathbb{E}_\theta(cS) + a_2^2\mathbb{E}_\theta(c^2S^2) \\
&= (1 - 2a_1 - 2a_2 + 2a_1a_2)\theta^2 + a_1^2\mathbb{E}_\theta(\overline{X}^2) + a_2^2\mathbb{E}_\theta(c^2S^2) \\
&= (1 - 2a_1 - 2a_2 + 2a_1a_2)\theta^2 + \frac{n+1}{n}a_1^2\theta^2 + c^2a_2^2\theta^2 \\
&= \left[(a_1 + a_2 - 1)^2 + \frac{1}{n}a_1^2 + (c^2 - 1)a_2^2\right]\theta^2
\end{aligned}$$

To minimize $\mathbb{E}_\theta[(\theta - T)^2]$ is to minimize the first part. Taking derivatives with respect to a_1 and a_2 yields

$$\begin{cases} 2(a_1 + a_2 - 1) + \frac{2}{n}a_1 = 0 \\ 2(a_1 + a_2 - 1) + 2(c^2 - 1)a_2 = 0 \end{cases},$$

i.e.,

$$\begin{cases} a_1 = \frac{nc^2 - n}{(n+1)c^2 - n} \\ a_2 = \frac{1}{(n+1)c^2 - n} \end{cases}.$$

So $T^* = \frac{nc^2 - n}{(n+1)c^2 - n}\overline{X} + \frac{1}{(n+1)c^2 - n}cS$

(b) Show that the MSE of T^* is smaller than the MSE of the estimator derived in Exercise 7.50(b).

Proof. Since $\min\limits_{a_1, a_2 \in \mathbb{R}} \mathbb{E}_\theta[(\theta - T)^2] \leq \min\limits_{\substack{a_1, a_2 \in \mathbb{R} \\ a_1 + a_2 = 1}} \mathbb{E}_\theta[(\theta - T)^2]$, so the MSE of T^* is smaller than the MSE of the estimator derived in Exercise 7.50(b). □

(c) Show that the MSE of ${T^*}^+ = \max\{0, T^*\}$ is smaller than the MSE of T^*.

Proof. Since

$$\begin{aligned}
\mathbb{E}_\theta[(\theta - T^*)^2] &= \theta^2 + \mathbb{E}_\theta(T^{*2}) - 2\theta\mathbb{E}_\theta(T^*) \\
&\geq \theta^2 + \mathbb{E}_\theta(\max\{0, T^*\}^2) - 2\theta\mathbb{E}_\theta(\max\{0, T^*\}) \\
&= \mathbb{E}_\theta[(\theta - \max\{0, T^*\})^2],
\end{aligned}$$

Solution (cont.)

we have that the MSE of ${T^*}^+ = \max\{0, T^*\}$ is smaller than the MSE of T^*. $\square$

(d) Would θ be classified as a location parameter or a scale parameter? Explain.

Sicne $\theta \in \mathbb{R}^+ \neq \mathbb{R}$, so θ is not a location parameter. θ is a scale parameter. Let f_0 be the density of $\mathcal{N}(1, 1)$, then $\frac{1}{\theta} f_0(\frac{x}{\theta})$ is the density of $\mathcal{N}(\theta, \theta^2)$. On the other hand, if $f_\theta(x)$ is a density of $\mathcal{N}(\theta, \theta^2)$, then $f_\theta(x) = \frac{1}{\theta} f_0(\frac{x}{\theta})$.