36-705 Intermediate Statistics Homework #4 Solutions

October 6, 2016

Problem 1 [30 pts.]

(a) **[20 pts.]** Find a minimal sufficient statistic.

A statistic T is minimal sufficient if $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$ does not depend on μ,Σ iff $T(y^n) = T(x^n)$. Suppose $X_1,\dots,X_n \sim N(\mu,\Sigma)$ and $Y_1,\dots,Y_n \sim N(\mu,\Sigma)$. We will show that $(\bar{X}, S_X)$ is minimal sufficient, where $S_X = \dfrac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$.

Let us start by deriving an expression for $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$. We see that

$$\begin{aligned}
\frac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)} &= \frac{\prod_{i=1}^n \Big[(2\pi)^{d/2}|\Sigma|^{1/2}\Big]^{-1}\exp\Big\{-\frac{1}{2}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\Big\}}{\prod_{i=1}^n \Big[(2\pi)^{d/2}|\Sigma|^{1/2}\Big]^{-1}\exp\Big\{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\Big\}} \\
&= \frac{\exp\Big\{\sum_{i=1}^n\Big[-\frac{1}{2}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\Big]\Big\}}{\exp\Big\{\sum_{i=1}^n\Big[-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\Big]\Big\}} \\
&= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n (y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right] + \frac{1}{2}\left[\sum_{i=1}^n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right]\right\}
\end{aligned}$$

Note that $(y_i-\mu)^T$ is $1\times d$, Σ^{-1} is $d\times d$, and $(y_i-\mu)$ is $d\times 1$. So $\sum_{i=1}^n (y_i-\mu)^T\Sigma^{-1}(y_i-\mu)$ is a scalar and will equal its own trace. Let us rewrite $\sum_{i=1}^n (y_i-\mu)^T\Sigma^{-1}(y_i-\mu)$.

$$\begin{aligned}
\sum_{i=1}^n (y_i-\mu)^T\Sigma^{-1}(y_i-\mu) &= \sum_{i=1}^n (y_i-\bar{y}+\bar{y}-\mu)^T\Sigma^{-1}(y_i-\bar{y}+\bar{y}-\mu) \\
&= \sum_{i=1}^n \mathrm{tr}\left[(y_i-\bar{y}+\bar{y}-\mu)^T\Sigma^{-1}(y_i-\bar{y}+\bar{y}-\mu)\right] \\
&= \sum_{i=1}^n \mathrm{tr}\left\{\Sigma^{-1}(y_i-\bar{y}+\bar{y}-\mu)(y_i-\bar{y}+\bar{y}-\mu)^T\right\} \\
&= \sum_{i=1}^n \mathrm{tr}\left\{\Sigma^{-1}\left[(y_i-\bar{y})(y_i-\bar{y})^T + 2(y_i-\bar{y})(\bar{y}-\mu)^T + (\bar{y}-\mu)(\bar{y}-\mu)^T\right]\right\} \\
&= \sum_{i=1}^n \mathrm{tr}\left\{\Sigma^{-1}(y_i-\bar{y})(y_i-\bar{y})^T\right\} + \sum_{i=1}^n \mathrm{tr}\left\{2\Sigma^{-1}(y_i-\bar{y})(\bar{y}-\mu)^T\right\} \\
&\quad + \sum_{i=1}^n \mathrm{tr}\left\{\Sigma^{-1}(\bar{y}-\mu)(\bar{y}-\mu)^T\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{tr}\left\{\Sigma^{-1}\sum_{i=1}^{n}(y_i-\bar{y})(y_i-\bar{y})^T\right\} + \mathrm{tr}\left\{(\bar{y}-\mu)^T 2\Sigma^{-1}\sum_{i=1}^{n}(y_i-\bar{y})\right\} \\
&\quad + \sum_{i=1}^{n}\mathrm{tr}\left\{(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu)\right\} \\
&= \mathrm{tr}\left\{\Sigma^{-1}nS_y\right\} + 0 + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu) \\
&= n\mathrm{tr}\left\{\Sigma^{-1}S_y\right\} + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu)
\end{aligned}
$$

Now we see that

$$
\begin{aligned}
\frac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)} = \exp\Bigg[-\frac{1}{2}\Bigg\{& n\mathrm{tr}\left\{\Sigma^{-1}S_y\right\} + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu) \\
&- n\mathrm{tr}\left\{\Sigma^{-1}S_x\right\} - n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\Bigg\}\Bigg].
\end{aligned}
$$

Now we can show that $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$ does not depend on μ,Σ iff $\bar{x}=\bar{y}$ and $S_x = S_y$.

($\Rightarrow$) Suppose that $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$ does not depend on μ,Σ. Define

$$
A = -\frac{1}{2}\left\{n\mathrm{tr}\left\{\Sigma^{-1}S_y\right\} + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu) - n\mathrm{tr}\left\{\Sigma^{-1}S_x\right\} - n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\right\}.
$$

Since $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$ does not depend on μ, we know that

$$
0 = \frac{\partial}{\partial\mu}\frac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)} = \exp[A]\frac{\partial}{\partial\mu}A.
$$

That means that

$$
\begin{aligned}
0 &= \frac{\partial}{\partial\mu}A \\
&= \frac{\partial}{\partial\mu}\left[-\frac{1}{2}n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu) + \frac{1}{2}n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\right] \\
&= \frac{\partial}{\partial\mu}\left[-\frac{1}{2}n(\mu-\bar{y})^T\Sigma^{-1}(\mu-\bar{y}) + \frac{1}{2}n(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x})\right] \\
&= -\frac{n}{2}\left[\Sigma^{-1}(\mu-\bar{y}) + \left(\Sigma^{-1}\right)^T(\mu-\bar{y})\right] + \frac{n}{2}\left[\Sigma^{-1}(\mu-\bar{x}) + \left(\Sigma^{-1}\right)^T(\mu-\bar{x})\right] \\
&= -\frac{n}{2}\left[\Sigma^{-1}(\mu-\bar{y}) + \left(\Sigma^{T}\right)^{-1}(\mu-\bar{y})\right] + \frac{n}{2}\left[\Sigma^{-1}(\mu-\bar{x}) + \left(\Sigma^{T}\right)^{-1}(\mu-\bar{x})\right]
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{n}{2}\left[\Sigma^{-1}(\mu-\bar{y})+\Sigma^{-1}(\mu-\bar{y})\right]+\frac{n}{2}\left[\Sigma^{-1}(\mu-\bar{x})+\Sigma^{-1}(\mu-\bar{x})\right] \\
&= -n\left[\Sigma^{-1}(\mu-\bar{y})\right]+n\left[\Sigma^{-1}(\mu-\bar{x})\right]
\end{aligned}
$$

Thus, we see that

$$
\begin{aligned}
n\left[\Sigma^{-1}(\mu-\bar{y})\right] &= n\left[\Sigma^{-1}(\mu-\bar{x})\right] \\
\Sigma^{-1}(\mu-\bar{y}) &= \Sigma^{-1}(\mu-\bar{x}) \\
\Sigma\Sigma^{-1}(\mu-\bar{y}) &= \Sigma\Sigma^{-1}(\mu-\bar{x}) \\
\mu-\bar{y} &= \mu-\bar{x} \\
\bar{y} &= \bar{x}.
\end{aligned}
$$

Also, since $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$ does not depend on Σ, we know that

$$
0 = \frac{\partial}{\partial\Sigma}\frac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)} = \exp[A]\frac{\partial}{\partial\Sigma}A.
$$

That means that

$$
\begin{aligned}
0 &= \frac{\partial}{\partial\Sigma}A \\
&= \frac{\partial}{\partial\Sigma}\left[-\frac{1}{2}n\mathrm{tr}\{\Sigma^{-1}S_y\}+\frac{1}{2}n\mathrm{tr}\{\Sigma^{-1}S_x\}\right] \\
&= -\frac{n}{2}\{-\Sigma^{-1}S_y\Sigma^{-1}\}+\frac{n}{2}\{-\Sigma^{-1}S_x\Sigma^{-1}\}.
\end{aligned}
$$

So

$$
\begin{aligned}
\frac{n}{2}\{\Sigma^{-1}S_x\Sigma^{-1}\} &= \frac{n}{2}\{\Sigma^{-1}S_y\Sigma^{-1}\} \\
\Sigma^{-1}S_x\Sigma^{-1} &= \Sigma^{-1}S_y\Sigma^{-1} \\
\Sigma\Sigma^{-1}S_x\Sigma^{-1}\Sigma &= \Sigma\Sigma^{-1}S_y\Sigma^{-1}\Sigma \\
S_x &= S_y.
\end{aligned}
$$

Thus, we have shown that $\bar{x} = \bar{y}$ and $S_x = S_y$.

($\Leftarrow$) Suppose that $\bar{x} = \bar{y}$ and $S_x = S_y$. Then

$$\begin{aligned}
\frac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)} &= \exp\Bigg[-\frac{1}{2}\Bigg\{n\mathrm{tr}\big\{\Sigma^{-1}S_y\big\} + n(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu) \\
&\quad - n\mathrm{tr}\big\{\Sigma^{-1}S_x\big\} - n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\Bigg\}\Bigg] \\
&= \exp\Bigg[-\frac{1}{2}\Bigg\{n\mathrm{tr}\big\{\Sigma^{-1}S_x\big\} + n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu) \\
&\quad - n\mathrm{tr}\big\{\Sigma^{-1}S_x\big\} - n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)\Bigg\}\Bigg] \\
&= \exp(0) \\
&= 1.
\end{aligned}$$

So $\dfrac{p(y^n;\mu,\Sigma)}{p(x^n;\mu,\Sigma)}$ does not depend on μ, Σ.

We conclude that $(\bar{X}, S_X)$ is a minimal sufficient statistic.

(b) **[10 pts.]** Show that $X_1 + X_2$ is not a sufficient statistic.

Since $(\bar{X}, S_X)$ is minimal sufficient for (μ, Σ), if $X_1 + X_2$ were sufficient there would be a function f such that $(\bar{X}, S_X) = f(X_1 + X_2)$, which is clearly impossible.

Problem 2 [30 pts.]

(a) **[20 pts.]** Find the distribution of (X_1, X_2) given T, where $T = \max\{X_1, X_2\}$.

Since X_1 and X_2 are independent, we see that

$$\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \\
&= \left(\frac{1}{\theta}\right)\cdot I(X_1 \le \theta)\cdot\left(\frac{1}{\theta}\right)\cdot I(X_2 \le \theta) \\
&= \left(\frac{1}{\theta}\right)^2 \cdot I(\max\{X_1, X_2\} \le \theta) \\
&= \left(\frac{1}{\theta}\right)^2 \cdot I(T \le \theta).
\end{aligned}$$

Also,

$$\begin{aligned}
f_{X_1,X_2,T}(x_1,x_2,t) &= \begin{cases} 0 & : T = \max\{x_1,x_2\} \ne t \\ f_{X_1,X_2}(x_1,x_2) & : T = \max\{x_1,x_2\} = t \end{cases} \\
&= \begin{cases} 0 & : T = \max\{x_1,x_2\} \ne t \\ \left(\frac{1}{\theta}\right)^2 \cdot I(T \le \theta) & : T = \max\{x_1,x_2\} = t \end{cases}
\end{aligned}$$

Next, let us solve for $f_T(t)$. We determine

$$F_T(t) = \mathbb{P}(T \le t) = \mathbb{P}(X_1 \le t \cap X_2 \le t) = \mathbb{P}(X_1 \le t)\mathbb{P}(X_2 \le t).$$

We know that

$$\mathbb{P}(X_1 \le t) = \begin{cases} 0 & : t < 0 \\ \frac{t}{\theta} & : 0 \le t \le \theta \\ 1 & : t > \theta \end{cases}$$

So

$$F_T(t) = \begin{cases} 0 & : t < 0 \\ \left(\frac{t}{\theta}\right)^2 & : 0 \le t \le \theta \\ 1 & : t > \theta \end{cases}$$

Differentiating with respect to t, we see that $f_T(t)$ will be non-zero on the interval $0 \le t \le \theta$. Specifically,

$$f_T(t) = \frac{2t}{\theta^2}, \qquad 0 \le t \le \theta.$$

Hence $f_{X_1,X_2|T}(x_1,x_2|t)$ is defined where $t = \max\{x_1,x_2\}$ (and hence t is in the interval $0 \le t \le \theta$). For $0 \le x_1 \le \theta$, $0 \le x_2 \le \theta$, and $t = \max\{x_1,x_2\}$,

$$\begin{aligned} f_{X_1,X_2|T}(x_1,x_2|t) &= \frac{f_{X_1,X_2,T}(x_1,x_2,t)}{f_T(t)} \\ &= \frac{(1/\theta)^2}{2t/\theta^2} \\ &= \frac{1}{2t} \end{aligned}$$

We conclude that

$$f_{X_1,X_2|T}(x_1,x_2|t) = \begin{cases} \frac{1}{2t} & : 0 \le x_1 \le \theta, 0 \le x_2 \le \theta, t = \max\{x_1,x_2\} \\ 0 & : \text{ else} \end{cases}$$

(b) **[10 pts.]** Show that $X_1 + X_2$ is not sufficient.

Consider the ratio

$$\begin{aligned} R(x^n, y^n; \theta) &= \frac{p(y_1,y_2;\theta)}{p(x_1,x_2;\theta)} \\ &= \frac{\left(\frac{1}{\theta}\right)^2 \mathbb{1}_{\left\{\max\{Y_1,Y_2\}\le\theta\right\}}}{\left(\frac{1}{\theta}\right)^2 \mathbb{1}_{\left\{\max\{X_1,X_2\}\le\theta\right\}}} \\ &= \frac{\mathbb{1}_{\left\{\max\{Y_1,Y_2\}\le\theta\right\}}}{\mathbb{1}_{\left\{\max\{X_1,X_2\}\le\theta\right\}}}. \end{aligned} \tag{1}$$

(1) is independent of θ if and only if $\max\{X_1,X_2\} = \max\{Y_1,Y_2\}$, so $\max\{X_1,X_2\}$ is a minimal sufficient statistic for θ. If $X_1 + X_2$ were sufficient there would be a function f such that $\max\{X_1,X_2\} = f(X_1 + X_2)$, which is clearly impossible.

Problem 3

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \left(\frac{1}{3\theta}\right) \mathbb{1}_{\left\{-\theta \le X_i \le 2\theta\right\}} \\
&= \left(\frac{1}{3\theta}\right)^n \mathbb{1}_{\left\{-\theta \le X_{(1)},\, \theta \ge X_{(n)}/2\right\}} \\
&= \left(\frac{1}{3\theta}\right)^n \mathbb{1}_{\left\{\theta \ge \max\{-X_{(1)},\, X_{(n)}/2\}\right\}}
\end{aligned}
$$

Problem 4

By linearity of the expectation we have that:

$$\mathbb{E}\left[\hat{\lambda}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}n\lambda = \lambda$$

So bias = 0. Moreover, since X_i's are independent:

$$\mathbb{V}\left(\hat{\lambda}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(X_i\right) = \frac{1}{n^2}n\lambda = \frac{\lambda}{n}$$

such that se $= \sqrt{\frac{\lambda}{n}}$ and MSE$(\hat{\lambda}) = 0^2 + \frac{\lambda}{n} = \frac{\lambda}{n}$.

Problem 5 [10 pts.]

$$
\begin{aligned}
\mathbb{E}_\theta[\widehat{\theta}] &= \mathbb{E}_\theta[2\overline{X}_n] \\
&= 2\mathbb{E}_\theta[\overline{X}_n] \\
&= 2 \cdot \frac{\theta}{2} \\
&= \theta,
\end{aligned}
$$

so $\text{Bias}_\theta(\widehat{\theta}) = \mathbb{E}_\theta[\widehat{\theta}] - \theta = 0$.

$$
\begin{aligned}
\text{se}_\theta(\widehat{\theta}) &= \sqrt{\mathbb{V}_\theta(2\overline{X}_n)} \\
&= \sqrt{4\mathbb{V}_\theta(\overline{X}_n)} \\
&= \sqrt{4 \cdot \frac{\theta^2}{12n}} \\
&= \frac{\theta}{\sqrt{3n}}
\end{aligned}
$$

Thus,

$$\begin{aligned}\mathbb{E}\left[(\widehat{\theta}-\theta)^2\right] &= \mathrm{Bias}^2_\theta(\widehat{\theta})^2 + \mathrm{se}^2_\theta(\widehat{\theta}) \\ &= \frac{\theta^2}{3n}.\end{aligned}$$

Problem 6 [30 pts.]

(a) **[10 pts.]** The first and second moments of $X_1 \sim \text{Uniform}(a, b)$ are:

$$\mathbb{E}[X_1] = \frac{a+b}{2}$$

$$\mathbb{E}\left[X_1^2\right] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2+ab+b^2}{3}$$

Let $\hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^n X_i$ and $\hat{\alpha}_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$ be the first and second sample moments. By solving the following system of equations

$$\begin{cases} \hat{\alpha}_1 &= \frac{b+a}{2} \\ \hat{\alpha}_2 &= \frac{a^2+ab+b^2}{3} \end{cases}$$

we obtain the following estimators:

$$\hat{a} = \hat{\alpha}_1 - \sqrt{3\left(\hat{\alpha}_2 - \hat{\alpha}_1^2\right)}$$

$$\hat{b} = \hat{\alpha}_1 + \sqrt{3\left(\hat{\alpha}_2 - \hat{\alpha}_1^2\right)}$$

To obtain $\hat{a}$ for example note that the first equation gives us that $2\hat{\alpha}_1 - a = b$, replacing this in the second equation:

$$\frac{a^2 + a(2\hat{\alpha}_1 - a) + (2\hat{\alpha}_1 - a)^2}{3} = \hat{\alpha}_2$$

Expanding the square, cancelling an regrouping we obtain:

$$(a - \hat{\alpha}_1)^2 = 3\left(\hat{\alpha}_2 - \hat{\alpha}_1^2\right)$$

So that $\hat{a} = \hat{\alpha}_1 - \sqrt{3\left(\hat{\alpha}_2 - \hat{\alpha}_1^2\right)}$. $\hat{b}$ is obtained similarly.

(b) **[10 pts.]** The likelihood function of $x = (x_1, ..., x_n)$ is:

$$\mathcal{L}(a,b) = \prod_{i=1}^n (b-a)^{-1} I_{(a,b)}(x_i) = (b-a)^{-n} I_{(-\infty, x_{(1)})}(a) I_{(x_{(n)},\infty)}(b)$$

By inspection this is maximized when $b - a$ is as small as possible while also satisfying $a \le x_{(1)} \le \ldots \le x_{(n)} \le b$. So the MLE estimators are $\hat{a} = X_{(1)}, \hat{b} = X_{(n)}$.

(c) **[10 pts.]** We have $\mathbb{E}[X_1] = \frac{a+b}{2}$. By equivariance property of the MLE we have that the MLE of τ is:

$$\hat{\tau} = \frac{\hat{a}+\hat{b}}{2} = \frac{X_{(1)} + X_{(n)}}{2}$$