HW 4 - Solutions

Intermediate Statistics - 10-705 / 36-705

Fall 2015

Problem 1 [25 points]

a) The likelihood function of $X_1, ..., X_n \sim \text{Uniform}(-\theta, \theta)$ can be manipulated as follows:

$$L(\theta; X_1, ..., X_n) = \prod_{i=1}^{n} I_{[-\theta, \theta]}(X_i) \frac{1}{2\theta}$$

$$= \frac{1}{(2\theta)^n} \prod_{i=1}^{n} I_{[0, \theta]}(|X_i|)$$

$$= \frac{1}{(2\theta)^n} I_{[0, \theta]} \left( \max_{1 \leq i \leq n} |X_i| \right)$$

$T(X_1, ..., X_n) = \max_{1 \leq i \leq n} |X_i|$ is minimal sufficient, because given two samples $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$, the ratio

$$R(X, Y, \theta) = \frac{L(\theta; X_1, ..., X_n)}{L(\theta; Y_1, ..., Y_n)} = \frac{I_{[0, \theta]}(T(X))}{I_{[0, \theta]}(T(Y))} = \frac{I_{[T(X), \infty)}(\theta)}{I_{[T(Y), \infty)}(\theta)}$$

does not depend on $\theta$ (or it is constant with respect to $\theta$; or $L(\theta; X) \propto L(\theta; Y)$ if and only if $T(X) = T(Y)$. In fact:

- Case 1: $T(X) < T(Y)$

$$R(X, Y, \theta) = \begin{cases} 
0 & = 1, \quad \theta < T(X) \\
\frac{0}{\theta} & = \infty, \quad \theta \in [T(X), T(Y)) \\
\frac{1}{\theta} & = 1, \quad \theta \geq T(Y) 
\end{cases}$$

(1)

- Case 2: $T(X) > T(Y)$

$$R(X, Y, \theta) = \begin{cases} 
0 & = 1, \quad \theta < T(Y) \\
\frac{0}{\theta} & = 0, \quad \theta \in [T(Y), T(X)) \\
\frac{1}{\theta} & = 1, \quad \theta \geq T(X) 
\end{cases}$$

(2)

\[1\text{Notice that since we are actually just checking for proportionality of } L(\theta; X_1, ..., X_n) \propto L(\theta; Y_1, ..., Y_n), \text{ we will have } 0 \frac{0}{\theta} = 1 \text{ and } \frac{1}{\theta} = \infty.\]
• Case 3: \( T(X) = T(Y) \)

\[
R(x, y, \theta) = \begin{cases} 
0 & , \theta < T(X) \\
\frac{1}{\theta} & , \theta \geq T(X)
\end{cases}
\] (3)

Thus, if and only if Case 3 holds then we have \( R(X, Y, \theta) = 1 \) for any \( \theta \). Therefore, statistic \( T(X) = \max_{1 \leq i \leq n} |X_i| \) is minimal sufficient for \( \theta \).

b) There is no function \( g(t) \) such that \( T(X_1, ..., X_n) = h(X_1) \) for any sample \( X_1, ..., X_n \), i.e. the minimal sufficient statistic cannot be derived by \( X_1 \) alone and therefore \( X_1 \) cannot be sufficient. This implies that the likelihood function cannot be factorized into \( g(\theta; X_1) \times h(X) \) (try by yourself!).

Alternate proof

We can show that the joint distribution of \( X_1, ..., X_n | X_1 \) still depends on \( \theta \): for \( x_1, ..., x_n, s \in [0, \theta] \)

\[
P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n | X_1 \leq s) = \frac{P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n, X_1 \leq s)}{P(X_1 \leq s)}
= \frac{P(X_1 \leq \min\{x_1, s\}, X_2 \leq x_2, ..., X_n \leq x_n)}{P(X_1 \leq s)}
= \frac{P(X_1 \leq \min\{x_1, s\}) \times \prod_{i=2}^{n} P(X_i \leq x_i)}{P(X_1 \leq s)}
= \frac{\min\{x_1, s\} / \theta \prod_{i=2}^{n} x_i}{s / \theta \prod_{i=2}^{n} \theta}
= \frac{\min\{x_1, s\} \prod_{i=2}^{n} x_i}{s \prod_{i=2}^{n} \theta}
\]

Problem 2 [0 points]

Let \( Y = \frac{X_1 - X_2}{2} \) so that \( X_1 = \bar{X} + Y \) and \( X_2 = \bar{X} - Y \). We have that the random vector \( (Y, \bar{X}) \) is a linear transformation of \( (X_1, X_2) \):

\[
\begin{pmatrix} Y \\ \bar{X} \end{pmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \times \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}
\]

Thus \( (Y, \bar{X}) \sim N \left( \begin{pmatrix} 0 \\ \mu \end{pmatrix}, \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \right) \)
where $\text{Cov}(Y, \bar{X}) = 0 \Rightarrow Y, \bar{X}$ independent since $(Y, \bar{X})$ are bivariate Normal distributed. Thus $X_1 = \bar{X} + Y$ and $X_2 = \bar{X} - Y$, which have: $\mathbb{E}[X_1|\bar{X}] = \mathbb{E}[\bar{X} + Y|\bar{X}] = \mathbb{E}[\bar{X}|\bar{X}] = \bar{X}$, $\text{Var}[X_1|\bar{X}] = \text{Var}[\bar{X} + Y|\bar{X}] = \text{Var}[Y] = \frac{1}{2}$; similarly $\mathbb{E}[X_2|\bar{X}] = \bar{X}$, $\text{Var}[X_2|\bar{X}] = \frac{1}{2}$; and $\text{Cov}(X_1, X_2|\bar{X}) = \text{Cov}(\bar{X} + Y, \bar{X} - Y|\bar{X}) = \text{Cov}(Y, -Y|\bar{X}) = -\text{Var}[Y] = -\frac{1}{2}$. Therefore conditionally on $\bar{X}$ we have

$$\left( \begin{array}{c} X_1 \\ X_2 \end{array} \right) | \bar{X} \sim N \left( \left( \begin{array}{c} \bar{X} \\ \bar{X} \end{array} \right), \left( \begin{array}{cc} 1/2 & -1/2 \\ -1/2 & 1/2 \end{array} \right) \right)$$

which does not depend on $\mu$ and is a degenerate distribution on the line $g(x) = 2\bar{X} - x$ since the covariance matrix is singular (determinant equal to zero). The result is not surprising because conditioning to a sufficient statistic makes the likelihood independent of the parameter.

**Problem 3 [25 points]**

Let $g(t) = \log \left( \frac{p(t; \theta_1)}{p(t; \theta_0)} \right)$, with $\theta_0, \theta_1$ fixed, so that $U_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$.

a) Note that $g(X_1), \ldots, g(X_n)$ are i.i.d. because $X_1, \ldots, X_n$ are i.i.d.. Thus, by linearity of expectations and identical distributions of $g(X_1), \ldots, g(X_n)$

$$\mu = \mathbb{E}[U_n] = \mathbb{E}[g(X_1)] = \mathbb{E} \left[ \log \left( \frac{p(X_1; \theta_1)}{p(X_1; \theta_0)} \right) \right] = -\mathbb{E} \left[ \log \left( \frac{p(X_1; \theta_0)}{p(X_1; \theta_1)} \right) \right] = -D (p(t; \theta_0)||p(t; \theta_1))$$

where $D$ is the Kullback-Leibler divergence, which we assume to be finite.

b) Since $U_n$ is the sample mean of $g(X_1), \ldots, g(X_n)$, which are i.i.d., then by Central Limit Theorem

$$\sqrt{n}(U_n - \mu) \sim N(0, \tau^2) \tag{4}$$

where $\tau^2 = \text{Var}(g(X_1)) = \mathbb{E} \left[ \left( \log \left( \frac{p(X_1; \theta_1)}{p(X_1; \theta_0)} \right) \right)^2 \right] - D (p(t; \theta_0)||p(t; \theta_1))^2$, which we need to assume to be finite ($\tau^2 < \infty$).

**Remark**

Note that all the expectations are with respect to $p(x; \theta_0)$, where $\theta_0$ is the true value of $\theta$. $X_1, \ldots, X_n$ are the only random quantities; $\theta_0, \theta_1$ are fixed.
Problem 4 [25 points]

For two samples \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_n)\), the ratio

\[
\frac{L(\mu; X_1, \ldots, X_n)}{L(\mu; Y_1, \ldots, Y_n)} = \frac{\exp \left\{ -\frac{1}{2\mu^2} \sum_{i=1}^{n} (X_i - \mu)^2 \right\}}{\exp \left\{ -\frac{1}{2\mu^2} \sum_{i=1}^{n} (Y_i - \mu)^2 \right\}}
\]

\[
= \frac{\exp \left\{ -\frac{1}{2\mu^2} \left( \sum_{i=1}^{n} X_i^2 + n\mu^2 - 2\mu \sum_{i=1}^{n} X_i \right) \right\}}{\exp \left\{ -\frac{1}{2\mu^2} \left( \sum_{i=1}^{n} Y_i^2 + n\mu^2 - 2\mu \sum_{i=1}^{n} Y_i \right) \right\}}
\]

\[
= \exp \left\{ -\frac{1}{2\mu^2} \left( \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} Y_i^2 \right) + \frac{1}{\mu} \left( \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i \right) \right\}
\]

does not depend on \(\mu\) if and only if both \(\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i\) and \(\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i^2\). Therefore \(T(X_1, \ldots, X_n) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)\) is minimal sufficient for \(\mu\).

Remark

This is a case where the minimal sufficient statistic is bivariate and the parameter is univariate.

Try to find the MSS of \(X_1, \ldots, X_n \sim \text{Uniform}(\theta, \theta + 1)\).

Problem 5 [0 points]

Let \(X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)\). Define \(\hat{\theta} = \max\{X_1, \ldots, X_n\}\). Then, since iid r.v.’s:

\[
P(\hat{\theta} \leq t) = P(X_1 \leq t, \ldots, X_n \leq t) = \prod_{i=1}^{n} P(X_i \leq t) = P(X_1 \leq t)^n = \left( \frac{t}{\theta} \right)^n I_{(0, \theta)}(t) I_{(\theta, \infty)}(t)
\]

such that the pdf of \(\hat{\theta}\) is \(f_{\hat{\theta}}(t) = \frac{n}{\theta} \left( \frac{t}{\theta} \right)^{n-1} I_{(0, \theta)}(t)\).

We have:

\[
\text{Bias} [\hat{\theta}] = \mathbb{E} [\hat{\theta}] - \theta
\]

\[
= \int_{0}^{\theta} \frac{n}{\theta} \left( \frac{t}{\theta} \right)^{n-1} dt - \theta
\]

\[
= \frac{n}{n+1} \theta - \theta
\]

\[
= -\frac{1}{n+1} \theta
\]
\[
\text{Var} \left[ \hat{\theta} \right] = \mathbb{E} \left[ \hat{\theta}^2 \right] - \left( \mathbb{E} \left[ \hat{\theta} \right] \right)^2
\]
\[
= \int_0^\theta t^2 \frac{n}{\theta} \left( \frac{t}{\theta} \right)^{n-1} dt - \left( \frac{n}{n+1} \right)^2
\]
\[
= \left[ \frac{n}{n+2} - \left( \frac{n}{n+1} \right)^2 \right] \theta^2
\]
\[
= \frac{n}{(n+1)^2(n+2)} \theta^2
\]

where clearly \( \text{SD} (\hat{\theta}) = \sqrt{\text{Var} \left[ \hat{\theta} \right]} \), and

\[
\text{MSE} \left[ \hat{\theta} \right] = \left( \text{Bias} \left[ \hat{\theta} \right] \right)^2 + \text{Var} \left[ \hat{\theta} \right]
\]
\[
= \frac{1}{(n+1)^2} \theta^2 + \frac{n}{(n+1)^2(n+2)} \theta^2
\]
\[
= \frac{2}{(n+1)(n+2)} \theta^2
\]

**Problem 6 [0 points]**

We need to estimate two parameters, \( \alpha \) and \( \beta \), so we need to use the first two moments. Using the pdf parametrized as \( p(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta} \), then we have \( \mathbb{E} [X_1] = \alpha \beta \) and \( \mathbb{E} [X_1^2] = \alpha \beta^2 + \alpha^2 \beta^2 \). Let \( M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k \). Thus we have to solve the system of equations:

\[
\begin{align*}
M_1 &= \hat{\alpha} \hat{\beta} \\
M_2 &= \hat{\alpha} \hat{\beta}^2 + \hat{\alpha}^2 \hat{\beta}^2
\end{align*}
\]

To get the MOM estimators

\[
\hat{\alpha}_{\text{MOME}} = \frac{M_1^2}{M_2 - M_1^2} \quad \text{and} \quad \hat{\beta}_{\text{MOME}} = \frac{M_2 - M_1^2}{M_1}
\]

**Remark**

Notice that if we use the alternate parametrization of the pdf \( p(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x/\beta} \), then we have \( \mathbb{E}[X_1] = \alpha / \beta \), \( \mathbb{E}[X_1^2] = \alpha / \beta^2 + \alpha^2 / \beta^2 \) and we get \( \hat{\alpha}_{\text{MOME}} = \frac{M_2}{M_2 - M_1} \) and \( \hat{\beta}_{\text{MOME}} = \frac{M_1}{M_2 - M_1} \).
Problem 7 [25 points]

There is only one parameter to estimate, $\lambda$, so we need to use only the first moment, which is $\mathbb{E}[X_1] = \lambda$. Thus the method of moments estimator is given by:

$$\hat{\lambda}_{MOME} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

To find the maximum likelihood estimator $\hat{\lambda}_{MLE}$ calculate the log likelihood function:

$$l_n(\lambda) = -n \lambda + \sum_{i=1}^{n} X_i \log \lambda - \sum_{i=1}^{n} \log X_i!$$

Its first derivative equals to:

$$\frac{\partial}{\partial \lambda} l_n(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} X_i$$

Setting it equal to 0 we get the unique solution $\lambda^* = \frac{1}{n} \sum_{i=1}^{n} X_i$. This is a global maximum because the second derivative of $l_n$ is negative for every $\lambda$

$$\frac{\partial^2}{\partial \lambda^2} l_n(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^{n} X_i < 0$$

i.e. $\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{\lambda}_{MOME}$.

Remark

Note that, in this case MOME and MLE agree. Yet, this is not always the case. You might try to find the MLE of $\Gamma(\alpha, \beta)$, with both $\alpha, \beta$ unknown (proving concavity is not so easy!). Then compare it with the MOME (Problem 6 of this HW).