36-705 Intermediate Statistics Homework #4
Solutions

October 5, 2017

Problem 1 (15 points)
We have that \( C = \{ A \cup B : A \in \mathcal{A}, B \in \mathcal{B} \} \). Let \( F \in \mathcal{F}_n \), then for \( C \in \mathcal{C} \),
\[
C \cap F = (A \cup B) \cap F = (A \cap F) \cup (B \cap F)
\]
Let \( m_A, m_B \) be the number of subsets of \( F \) that \( \mathcal{A} \) and \( \mathcal{B} \) can pick out. Then the number of distinct sets of the form \((A \cup B) \cap F\) is the total number of distinct unions of the form \((A \cap F) \cup (B \cap F)\) which is bounded by \( m_A m_B \). Thus:
\[
S(C, F) \leq S(A, F) \times S(B, F)
\]
Again taking the supremum over all \( F \in \mathcal{F}_n \) on both sides:
\[
s_n(C) = \sup_{F \in \mathcal{F}_n} S(C, F) \leq \sup_{F \in \mathcal{F}_n} (S(A, F) \times S(B, F))
\]
\[
\leq \sup_{F \in \mathcal{F}_n} S(A, F) \sup_{F \in \mathcal{F}_n} S(B, F) = s(A, n) \times s(B, n).
\]

Problem 2 (15 points)
Let \( F \) be a finite set of \( n \) elements. We have that \( \mathcal{C} = \{ A : A \in \mathcal{A} \text{ or } A \in \mathcal{B} \} \), so if \( \mathcal{C} \) picks out \( G \subseteq F \), then either \( \mathcal{A} \) picks out \( G \) or \( \mathcal{B} \) picks out \( G \). Thus, for any \( F \in \mathcal{F}_n \), the total number of subsets picked out by \( \mathcal{C} \) is the total number of distinct \( G \subseteq F \) picked out by either \( \mathcal{A} \) or \( \mathcal{B} \). Therefore, for any finite set \( F \):
\[
S(C, F) \leq S(A, F) + S(B, F)
\]
Taking the supremum over all \( F \in \mathcal{F}_n \) on both sides:
\[
s(C, n) = \sup_{F \in \mathcal{F}_n} S(C, F) \leq \sup_{F \in \mathcal{F}_n} (S(A, F) + S(B, F))
\]
\[
\leq \sup_{F \in \mathcal{F}_n} S(A, F) + \sup_{F \in \mathcal{F}_n} S(B, F) = s(A, n) + s(B, n)
\]

\(^1\) Formal proof: Let \( \mathcal{A}_F = \{ A \cap F : A \in \mathcal{A} \}, \mathcal{B}_F = \{ B \cap F : B \in \mathcal{B} \}, \) and \( \mathcal{C}_F = \{ (A \cap F) \cup (B \cap F) : A \in \mathcal{A}, B \in \mathcal{B} \}. \)
Define a map \( \Phi : \mathcal{A}_F \times \mathcal{B}_F \to \mathcal{C}_F \) by \( \Phi((A \cap F) \times (B \cap F)) = (A \cap F) \cup (B \cap F) \). Then \( \Phi \) is surjective, so \( |\mathcal{C}_F| \leq |\mathcal{A}_F \times \mathcal{B}_F| = |\mathcal{A}_F| \times |\mathcal{B}_F| = m_A m_B \), i.e. \( |((A \cap F) \cup (B \cap F) : A \in \mathcal{A}, B \in \mathcal{B})| \leq m_A m_B \), where \( |X| \) means the number of elements of \( X \) for any set \( X \).
Problem 3 (15 points)

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

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

$$= \frac{1}{(2\theta)^n} \prod_{i=1}^{n} I\{|X_i| \leq \theta\}$$

$$= \frac{1}{(2\theta)^n} I\{\max_{1 \leq i \leq n} |X_i| \leq \theta\}.$$ 

(b) $T(x_1, ..., x_n) = \max_{1 \leq i \leq n} |x_i|$ is minimal sufficient, because given two samples $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, the ratio

$$R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \frac{p(x_1, ..., x_n; \theta)}{p(y_1, ..., y_n; \theta)} = \frac{I\{T(x_1, ..., x_n) \leq \theta\}}{I\{T(y_1, ..., y_n) \leq \theta\}} = \frac{I(T(x); \infty)(\theta)}{I(T(y); \infty)(\theta)}$$

does not depend on $\theta$ (or it is constant with respect to $\theta$; or $p(x; \theta) \propto p(y; \theta)$) if and only if $T(x) = T(y)$. In fact\(^2\):

- **Case 1:** $T(x) < T(y)$

$$R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} 0 &= 1, \quad \theta < T(x) \\ \frac{0}{0} &= \infty, \quad \theta \in [T(x), T(y)) \\ \frac{0}{0} &= 1, \quad \theta \geq T(y) \end{cases} \quad (1)$$

- **Case 2:** $T(x) > T(y)$

$$R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} 0 &= 1, \quad \theta < T(y) \\ \frac{0}{0} &= 0, \quad \theta \in [T(y), T(x)) \\ \frac{0}{0} &= 1, \quad \theta \geq T(x) \end{cases} \quad (2)$$

- **Case 3:** $T(x) = T(y)$

$$R(x_1, ..., x_n, y_1, ..., y_n, \theta) = \begin{cases} 0 &= 1, \quad \theta < T(x) \\ \frac{0}{0} &= 1, \quad \theta \geq T(x) \end{cases} \quad (3)$$

Thus, only in Case 3 we have $R(x_1, ..., x_n, y_1, ..., y_n, \theta) = 1$ for any $\theta$. Therefore, statistic $T(x) = \max_{1 \leq i \leq n} |x_i|$ is minimal sufficient for $\theta$. Note that writing down Case 2 could be avoided by adding some magic words in Case 1: “without loss of generality assume $T(x) < T(y)$...”. In fact the ratio $R(x_1, ..., x_n, y_1, ..., y_n, \theta)$ is used to check for proportionality of the two likelihood functions $p(x_1, ..., x_n; \theta) \propto p(y_1, ..., y_n; \theta)$ and we could just switch numerator and denominator of the ratio, i.e. using the ratio $R(y_1, ..., y_n, x_1, ..., x_n, \theta) = \frac{p(y_1, ..., y_n; \theta)}{p(x_1, ..., x_n; \theta)}$.

(c) 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.

\(^2\)Notice that since we are actually just checking for proportionality of $p(x_1, ..., x_n; \theta) \propto p(y_1, ..., y_n; \theta)$, we will have $\frac{0}{0} = 1$ and $\frac{0}{0} = \infty$. 

2
Alternate proofs

- Alternatively note that

\[
L(\theta; X_1, ..., X_n) = \frac{1}{(2\theta)^n} I_{[0, \theta]} \left( \max_{1 \leq i \leq n} |X_i| \right)
\]

\[
= \frac{1}{(2\theta)^n} I_{[0, \theta]} \left( \max_{2 \leq i \leq n} |X_i| \right) I_{[0, \theta]} (|X_1|)
\]

showing that the likelihood function cannot be factorized into \(g(\theta; X_1) \times h(X)\) since the rest of the data is still required to define the likelihood function.

- A more direct way to prove that \(X_1\) is not sufficient is to show that the joint distribution of \(X_1, ..., X_n| X_1\) still depends on \(\theta\): for \(x_1, ..., x_n, s \in [0, \theta]\)

\[
P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n| X_1 \leq s) = \frac{P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n, X_1 \leq s)}{P(X_1 \leq s)}
\]

\[
= \frac{P(X_1 \leq \min\{x_1, s\}, X_2 \leq x_2, ..., X_n \leq x_n)}{P(X_1 \leq s)}
\]

\[
= \frac{P(X_1 \leq \min\{x_1, s\}) \times \prod_{i=2}^{n} P(X_i \leq x_i)}{P(X_1 \leq s)}
\]

\[
= \frac{\min\{x_1, s\}}{s} \prod_{i=2}^{n} \frac{x_i}{\theta}
\]

Problem 4 (15 points)

\(X_1, ..., X_n \sim N(\mu, \mu^2)\) implies,

\[
f(x_i) = \frac{1}{\sqrt{2\pi\mu}} \exp \left\{ -\frac{1}{2\mu^2} (x_i - \mu)^2 \right\}.
\]

Therefore the likelihood function is given by,

\[
L(\mu; X_1, ..., X_n) = p(X_1, ..., X_n; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\mu}} \exp \left\{ -\frac{1}{2\mu^2} (x_i - \mu)^2 \right\} = \left( \frac{1}{\sqrt{2\pi\mu}} \right)^n \exp \left\{ -\frac{1}{2\mu^2} \sum_{i=1}^{n} (X_i - \mu)^2 \right\}.
\]

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

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

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

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

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

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, the minimal sufficient statistic is \(T(X_1, ..., X_n) = \left( \sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)\).
**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 (20 points)**

We have $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\theta = X_1$ and sufficient statistic $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$.

(a) The Rao-Blackwellized estimator is then given by,

$$\bar{\theta} = \mathbb{E}[X_1 | T].$$

Now notice that $X_1 \overset{d}{=} X_i$ for any $2 \leq i \leq n$ and interchanging $X_1$ with $X_i$ does not change $T$, that is, $T(X_1, \ldots, X_i, \ldots, X_n) = T(X_i, X_1, \ldots, X_n)$. Therefore $\mathbb{E}[X_1 | T] = \mathbb{E}[X_i | T]$ for any $2 \leq i \leq n$. Therefore,

$$n\mathbb{E}[X_1 | T] = \sum_{i=1}^{n} \mathbb{E}[X_i | T] = \mathbb{E}[\sum_{i=1}^{n} X_i | T] = \mathbb{E}[T | T] = T.$$ 

So,

$$\bar{\theta} = \mathbb{E}[X_1 | T] = \frac{T}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$ 

Alternate Solution: $(X_1, \sum_{i=1}^{n} X_i)$ has a jointly Gaussian distribution with mean,

$$\mu = \mathbb{E} \left[ \frac{X_1}{\sum_{i=1}^{n} X_i} \right] = \left( \begin{array}{c} \mu \\ n\mu \end{array} \right).$$

Now $\text{Var}(X_1) = \sigma^2$, $\text{Cov}(X_1, \sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Cov}(X_1, X_i) = \text{Var}(X_1) = \sigma^2$ and $\text{Var}(\sum_{i=1}^{n} X_i) = n\sigma^2$. Therefore, covariance matrix is given by,

$$\Sigma = \left( \begin{array}{cc} \sigma^2 & \sigma^2 \\ \sigma^2 & n\sigma^2 \end{array} \right).$$

Then $X_1 | T \sim N \left( \mu + \frac{\sigma}{\sqrt{n\sigma^2}} \left( t - n\mu \right), \left( 1 - \frac{1}{n} \right) \sigma^2 \right) = N \left( \frac{T}{n}, \left( 1 - \frac{1}{n} \right) \sigma^2 \right)$.

Therefore, the Rao-Blackwellized estimator is given by

$$\bar{\theta} = \mathbb{E}[\hat{\theta} | T] = \mathbb{E}[X_1 | T] = \frac{T}{n} = \overline{X}.$$ 

(b) The risk of $\bar{\theta} = X_1$ is given by,

$$R(\bar{\theta}, \mu) = \mathbb{E} \left[ (\bar{\theta} - \mu)^2 \right] = \mathbb{E} \left[ (X_1 - \mu)^2 \right] = \text{Var}(X_1) = \sigma^2.$$ 

The risk of the Rao-Blackwellized estimator is given by,

$$R(\bar{\theta}, \mu) = \mathbb{E} \left[ (\bar{\theta} - \mu)^2 \right] = \mathbb{E} \left[ (\overline{X} - \mu)^2 \right] = \text{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$ 

Therefore, $R(\bar{\theta}, \mu) \leq R(\hat{\theta}, \mu)$.

(c) Since $X_1$ and $X_2$ are independent, $\mathbb{E}[X_1 | X_2] = \mathbb{E}[X_1] = \mu$ which is not an estimator at all.
Problem 6 (20 points)

(a) The gamma distribution with density,

\[
p(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} = \exp\left[-\log(\Gamma(k)) - k \log(\theta) + (k - 1) \log(x) - \frac{x}{\theta}\right]
\]

is a 2-parameter exponential family, with canonical parameters \( (k, -\frac{1}{\theta}) \) and sufficient statistics \((\log(x), x)\).

Here \( A(\theta) = \log(\Gamma(k)) + k \log(\theta) \). Then

\[
E[\log(X)] = \frac{\partial A(\theta)}{\partial k} = \frac{\partial A(\theta)}{\partial (\frac{1}{\theta})} = k/\theta = k\theta.
\]

(b) The central \( \chi^2 \) which has density,

\[
p(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2 - 1} e^{-x/2} = \exp\left[-\log(\Gamma(k/2)) - \frac{k}{2} \log(2) + \left(\frac{k}{2} - 1\right) \log(x) - \frac{x}{2}\right],
\]

is a 1-parameter exponential family, with canonical parameter \( k \) and sufficient statistic \( \frac{\log(x)}{2} \).

Here \( A(\theta) = \log(\Gamma(k/2)) + \frac{k}{2} \log(2) \). Then

\[
E\left[\frac{\log(X)}{2}\right] = \frac{\partial A(\theta)}{\partial k} = \frac{1}{2} \frac{\Gamma'(k/2)}{\Gamma(k/2)} = \frac{\log 2}{2} = \psi(k/2) + \frac{\log 2}{2},
\]

where \( \psi(k) \) is the digamma function.

(c) The multinomial distribution which has density,

\[
p(x; p_1, \ldots, p_k) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}
\]

\[
= \frac{n!}{x_1! \ldots x_k!} \exp\left[\sum_{i=1}^k x_i \log(p_i)\right]
\]

\[
= \frac{n!}{x_1! \ldots x_k!} \exp\left[\sum_{i=1}^{k-1} x_i \log(p_i) + x_k \log(p_k)\right]
\]

\[
= \frac{n!}{x_1! \ldots x_k!} \exp\left[\sum_{i=1}^{k-1} x_i \log(p_i) + \left(n - \sum_{i=1}^{k-1} x_i\right) \log(p_k)\right]
\]

\[
= \frac{n!}{x_1! \ldots x_k!} \exp\left[\sum_{i=1}^{k-1} x_i \log\left(\frac{p_i}{p_k}\right) + \log(p_k)\right]
\]

is a \((k-1)\)-parameter exponential family, with canonical parameters \( \left(\log\left(\frac{p_1}{p_k}\right), \ldots, \log\left(\frac{p_{k-1}}{p_k}\right)\right) \)

and sufficient statistics \((x_1, \ldots, x_{k-1})\).
Let us define $\eta_i = \log \left( \frac{p_i}{p_k} \right)$, then

$$A(\theta) = -n \log p_k = n \log \left( \frac{1}{p_k} \right) = n \log \left( 1 + \sum_{j=1}^{k-1} \frac{p_j}{p_k} \right) = n \log \left( 1 + \sum_{j=1}^{k-1} e^{\eta_j} \right).$$

Then for $i \leq k - 1$,

$$\mathbb{E}[X_i] = \frac{\partial A(\theta)}{\partial \eta_i} = \frac{n}{1 + \sum_{j=1}^{k-1} e^{\eta_j}} e^{\eta_i} = \frac{n}{1 + \sum_{j=1}^{k-1} \frac{p_j}{p_k}} \frac{p_i}{p_k} = np_i \frac{p_i}{p_k} = np_i.$$