Midterm exam, Statistical Inference I: Date 11/6 (2015 Fall): [+30points]

Name: **Shih Jia-Han**

Q1 [+4]. Let $X_1,..,X_n \overset{iid}{\sim} U(0,\theta)$

1) [+1] Find a sufficient statistic for θ [with proof].

Ans:

$$\begin{aligned} f(x_1,\cdots,x_n) &= \prod_{i=1}^{n} f_i(x_i) = \prod_{i=1}^{n} \frac{1}{\theta} I(0<x_i<\theta) \\ &= \frac{1}{\theta^n} I(0<x_1,\cdots,x_n<\theta) \\ &= \frac{1}{\theta^n} I(0<x_{(1)}) I(x_{(n)}<\theta) \\ &= g_\theta(T)h(x), \end{aligned}$$

where $g_\theta(x) = \frac{1}{\theta^n} I(x<\theta)$, $h(x) = I(0<x_{(1)})$ and $T = x_{(n)}$.

By the factorization criterion, $T = X_{(n)}$ is sufficient statistics for θ.

2) [+3] Prove that the statistic is complete and minimal for θ.

Ans:

By 1), $T = X_{(n)}$ is sufficient statistics for θ.

Since

$$\begin{aligned} F_T(t) &= \Pr(T \le t) = \Pr(X_{(n)} \le t) \\ &= \Pr(X_1,\cdots,X_n \le t) = \Pr(X_1 \le t)^n \\ &= \left(\frac{t}{\theta}\right)^n, \quad 0<t<\theta. \end{aligned}$$

Then

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{nt^{n-1}}{\theta^n}, \quad 0<t<\theta.$$

Therefore, if $E_\theta[f(T)]=0$, for all θ, we have

$$\begin{aligned} &\int_0^\theta f(t)\frac{nt^{n-1}}{\theta^n}dt = 0 \Rightarrow \int_0^\theta f(t)t^{n-1}dt = 0 \\ &\Rightarrow \int_0^\theta \{f^+(t) - f^-(t)\}t^{n-1}dt = 0 \\ &\Rightarrow \int_0^\theta f^+(t)t^{n-1}dt = \int_0^\theta f^-(t)t^{n-1}dt \end{aligned}$$

$$\Rightarrow \frac{d}{d\theta}\int_0^\theta f^+(t)t^{n-1}dt = \frac{d}{d\theta}\int_0^\theta f^-(t)t^{n-1}dt$$

$$\Rightarrow f^+(\theta)\theta^{n-1} = f^-(\theta)\theta^{n-1}$$

$$\Rightarrow f^+(\theta) = f^-(\theta)$$

$$\Rightarrow f(\theta) = 0, \quad \text{for all } \theta$$

$$\Rightarrow f(t) = 0, \quad \text{for all } t.$$

Hence we obtain if $E_\theta[f(T)] = 0, \quad \text{for all } \theta \Rightarrow f(t) = 0, \quad \text{for all } t$.

Therefore, $T = X_{(n)}$ is complete and minimal for θ.

Q2 [+4]. Let $X_1,..,X_n \overset{iid}{\sim} p_\theta(x)$, where the density (w.r.t. the counting measure) is

$$p_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0,1,...,n \quad 1 > \theta > 0.$$

1) Express the density as the canonical form (specify all the components).

Ans:

$$\begin{aligned} f(x_1,\cdots,x_n) &= \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \binom{n}{x_i}\theta^{x_i}(1-\theta)^{n-x_i} \\ &= \prod_{i=1}^n \binom{n}{x_i}\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n^2-\sum_{i=1}^n x_i} \\ &= \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i}(1-\theta)^{n^2}\prod_{i=1}^n \binom{n}{x_i} \\ &= \exp\left\{\sum_{i=1}^n x_i \log\left(\frac{\theta}{1-\theta}\right) + n^2\log(1-\theta)\right\}\prod_{i=1}^n \binom{n}{x_i} \\ &= \exp\{\eta_1(\theta)T_1(x) - A(\eta)\}h(x) \end{aligned}$$

is the canonical form, where

$$\eta_1(\theta) = \log\left(\frac{\theta}{1-\theta}\right), \quad T_1(x) = \sum_{i=1}^n x_i, \quad A(\eta) = -n^2\log(1+e^{\eta_1}), \quad h(x) = \prod_{i=1}^n \binom{n}{x_i}$$

$$\chi = \{0,1,\cdots,n\}\times\cdots\times\{0,1,\cdots,n\} \text{ and } \Theta = \{\eta_1 \mid \eta_1 > 0\}.$$

2) Find a complete and sufficient statistics (with proof).

Ans:

By 1),

$$f(x_1,\cdots,x_n) = \exp\{\eta_1(\theta)T_1(x) - A(\eta)\}h(x).$$

Since the natural parameter space $\Theta = \{\eta_1 \mid \eta_1 > 0\}$ contains an one-dimensional open rectangle (e.g., $(1,2) \in \Theta$). Hence it is full rank.

Since it is full rank, $T_1(x) = \sum_{i=1}^n x_i$ is a complete and sufficient statistics.

Q3 [+12] Let $X_1,..,X_n \overset{iid}{\sim} N(\xi,\sigma^2)$, where $\theta=(\xi,\sigma^2)$.

1) **[+3]** Let $\delta=\frac{n-1}{n}S^2$, where S^2 is the sample variance, be an estimator of σ^2. Derive the risk $R(\theta,\delta)$ under the squared error loss.

<u>Ans:</u>

Since

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{\mathrm{df}=1}.$$

Therefore, we have

$$E(S^2)=\sigma^2 \quad \text{and} \quad \mathrm{var}(S^2)=\frac{2\sigma^4}{n-1}.$$

Thus, the risk $R(\theta,\delta)$ under the squared error loss is

$$\begin{aligned}
R(\theta,\delta) &= E\left\{\left(\frac{n-1}{n}\right)S^2-\sigma^2\right\}^2 = E\left\{\left(\frac{n-1}{n}\right)S^2-\left(\frac{n-1}{n}\right)\sigma^2+\left(\frac{n-1}{n}\right)\sigma^2-\sigma^2\right\}^2 \\
&= E\left\{\left(\frac{n-1}{n}\right)(S^2-\sigma^2)-\frac{\sigma^2}{n}\right\}^2 = \left(\frac{n-1}{n}\right)^2 E(S^2-\sigma^2)^2+\frac{\sigma^4}{n^2} \\
&= \left(\frac{n-1}{n}\right)^2 \mathrm{var}(S^2)+\frac{\sigma^4}{n^2} = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1}+\frac{\sigma^4}{n^2} \\
&= \frac{2\sigma^4(n-1)+\sigma^4}{n^2} = \frac{2n\sigma^4-\sigma^4}{n^2}.
\end{aligned}$$

2) **[+3]** Verify that the above risk is smaller or larger than the risk of $\delta' = S^2$.

<u>Ans:</u>

The risk $R(\theta,\delta')$ under the squared error loss is

$$R(\theta,\delta')=E(S^2-\sigma^2)^2=\mathrm{var}(S^2)=\frac{2\sigma^4}{n-1}.$$

Compare $R(\theta,\delta)$ and $R(\theta,\delta')$

$$\begin{aligned}
R(\theta,\delta')-R(\theta,\delta) &= \frac{2\sigma^4}{n-1}-\frac{2n\sigma^4-\sigma^4}{n^2} \\
&= \frac{2n^2\sigma^4-2n^2\sigma^4+2n\sigma^4+n\sigma^4-\sigma^4}{(n-1)n^2} \\
&= \frac{3n\sigma^4-\sigma^4}{(n-1)n^2} > 0.
\end{aligned}$$

Hence

$$R(\theta,\delta') > R(\theta,\delta).$$

3) **[+6]** Let $\bar{X} = \sum_{i=1}^{n} X_i / n$ and $\delta = (\bar{X})^2$ be an estimator of ξ^2. Derive the risk $R(\theta,\delta)$ under the squared error loss.

<u>**Ans:**</u>

Since

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow \bar{X} \sim N(\xi, \sigma^2).$$

Therefore,

$$\frac{\bar{X}-\xi}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\Rightarrow \left(\frac{\bar{X}-\xi}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2_{\text{df}=1}$$
$$\Rightarrow \frac{n}{\sigma^2}(\bar{X}-\xi)^2 \sim \chi^2_{\text{df}=1}$$
$$\Rightarrow \frac{n}{\sigma^2}(\bar{X}^2+\xi^2-2\xi\bar{X}) \sim \chi^2_{\text{df}=1}.$$

Hence

$$E\left\{\frac{n}{\sigma^2}(\bar{X}^2+\xi^2-2\xi\bar{X})\right\} = 1$$
$$\Rightarrow E(\bar{X}^2)+\xi^2-2\xi^2 = \frac{\sigma^2}{n}$$
$$\Rightarrow E(\bar{X}^2) = \xi^2+\frac{\sigma^2}{n}$$

and

$$\text{var}\left\{\frac{n}{\sigma^2}(\bar{X}^2+\xi^2-2\xi\bar{X})\right\} = 2$$
$$\Rightarrow \text{var}(\bar{X}^2-2\xi\bar{X}) = \frac{2\sigma^4}{n^2}$$
$$\Rightarrow \text{var}(\bar{X}^2)+4\xi^2\,\text{var}(\bar{X}) = \frac{2\sigma^4}{n^2}$$
$$\Rightarrow \text{var}(\bar{X}^2)+\frac{4\xi^2\sigma^2}{n} = \frac{2\sigma^4}{n^2}$$
$$\Rightarrow \text{var}(\bar{X}^2) = \frac{2\sigma^4}{n^2}-\frac{4\xi^2\sigma^2}{n}.$$

Thus,

$$
\begin{aligned}
R(\theta, \delta) &= E\left\{\overline{X}^2 - \left(\xi^2 + \frac{\sigma^2}{n}\right) + \frac{\sigma^2}{n}\right\}^2 \\
&= E\left\{\overline{X}^2 - \left(\xi^2 + \frac{\sigma^2}{n}\right)\right\}^2 + \frac{\sigma^4}{n^2} \\
&= \operatorname{var}(\overline{X}^2) + + \frac{\sigma^4}{n^2} \\
&= \frac{3\sigma^4}{n^2} - \frac{4\xi^2\sigma^2}{n}.
\end{aligned}
$$

Q4 [+10]

[+7] State and prove Basu's Theorem

Ans:

Basu's Theorem

If T is an complete sufficient statistics for θ and V is ancillary for θ.

Then $T \perp V$.

Proof:

Let

$$P_A = \Pr_\theta(V \in A)$$

Since V is ancillary for θ, we have

$$P_A = \Pr(V \in A).$$

Let

$$\eta_A(t) = \Pr(V \in A | T = t), \quad t \in \{\text{support of } T\}.$$

Then,

$$P_A = \Pr(V \in A) = E_\theta[P(V \in A | T)] = E_\theta[\eta_A(T)].$$

Hence

$$E_\theta[\eta_A(T) - P_A] = 0, \quad \text{for all } \theta.$$

By the definition of completeness,

$$\begin{aligned}
&\eta_A(t) - P_A = 0, \quad \text{for all } t \in \{\text{supprot of } T\} \\
&\Rightarrow \eta_A(t) = P_A \\
&\Rightarrow \Pr(V \in A | T = t) = \Pr(V \in A) \\
&\Rightarrow T \perp V.
\end{aligned}$$

[+3] State how the sufficiency is used in the proof.

Ans:

Since

$$\begin{aligned}
&X \perp Y \\
&\Leftrightarrow \Pr(X \in A, Y \in B) = \Pr(X \in A)\Pr(Y \in B), \quad \text{for all } A, B \\
&\Leftrightarrow \frac{\Pr(X \in A, Y \in B)}{\Pr(Y \in B)} = \Pr(X \in A), \quad \text{for all } A, B \\
&\Leftrightarrow \Pr(X \in A | Y \in B) = \Pr(X \in A), \quad \text{for all } A, B.
\end{aligned}$$

Hence

$$X \perp Y \Leftrightarrow \Pr(X \in A | Y \in B) = \Pr(X \in A), \quad \text{for all } A, B.$$