Statistical Inference II, 2013 Spring

Final exam

Q1
Q2
Q3
Q4
Q5

YOUR NAME_______________________

NOTE1: Please write down the derivation of your answer very clearly for all questions. The score will be reduced when you only write answer. Also, the score will be reduced if the derivation is not clear. The score will be added even when your answer is incorrect but the derivation is correct.

1. Bayes interval

Let $X = (X_1, ..., X_n) \overset{iid}{\sim} N(\theta, \sigma^2)$, where σ^2 is known, and $\theta \sim N(0, \tau^2)$.

1. Find $1-\alpha$ HPD confidence interval for θ.

2. Compare the length of the HPD confidence interval with that of the usual $1-\alpha$ confidence interval.

3. Calculate the coverage probability $P_\theta(\theta \in C_B(X))$, where $C_B(X)$ is the HPD confidence interval.

2. Goodness-of-fit test

Consider $H_0: F = F_0$ vs. $H_1: F \neq F_0$ based on iid data $(X_1,\ldots,X_n) \overset{iid}{\sim} F$ for a continuous c.d.f. F.

(a) Show that the distribution of the Cramér-von Mises statistc

$$C_n(F_0) = \int \{ F_n(x) - F_0(x) \}^2 \, dF_0(x)$$

does not depend on F_0.

(b) Express the Cramér-von Mises statistc in terms of V-statistic.

(c) Show that the distribution of the Kolmogorov-Smirnov statistic

$$D_n(F_0) = \sup_x | F_n(x) - F_0(x) |$$

does not depend on F_0.

(d) Derive the c.d.f. of $D_n(F_0)$.

3. Data $(X_1,\ldots,X_n)$ follows independently and identically a Weibull distribution with $P_\theta(X_i > x) = \exp\{-(x/\theta)^c\} I_{(0,\infty)}(x) + I_{(-\infty,0]}(x)$, where $\theta > 0$ is unknown and $c > 0$ is known. Derive a level (1-α) Θ'-UMA confidence set for some set Θ'.

4. Likelihood based interval

Let $X = (X_1,..., X_n) \overset{iid}{\sim} Pois(\lambda)$, where λ is the mean parameter. Find $1-\alpha$ confidence sets for λ by inverting the LR, Wald and Score tests. (explicit solution whenever possible)

5. Confidence band

Let $X = (X_1,...,X_n) \overset{iid}{\sim} U(0,\theta)$. Find the $(1-\alpha)$ confidence band for the c.d.f.

Answer 1:

1. Note that $\theta \mid X \sim N(\frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x}, \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2})$. Then, the HPD confidence interval is

$$C_B(X) = \left[\frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} - z_{1-\alpha/2}\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}, \quad \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} + z_{1-\alpha/2}\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}\right].$$

2. For the usual $1-\alpha$ confidence interval,

$$\text{length}\{C(X)\} = 2z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}.$$

On the other hand,

$$\text{length}\{C_B(X)\} = 2z_{1-\alpha/2}\sqrt{\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}} = 2z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n+\sigma^2/\tau^2}} < 2z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}.$$

Hence, the HPD is shorter than the usual confidence set.

3. $C_B(X) = \left[a\bar{x} - z_{1-\alpha/2}\sqrt{a\frac{\sigma^2}{n}}, \quad a\bar{x} + z_{1-\alpha/2}\sqrt{a\frac{\sigma^2}{n}}\right]$, where $a = \frac{n\tau^2}{n\tau^2+\sigma^2}$.

Then,

$$P_\theta(\theta \in C_B(X)) = P_\theta\left(a\bar{X} - z_{1-\alpha/2}\sqrt{a\frac{\sigma^2}{n}} \le \theta \le a\bar{X} + z_{1-\alpha/2}\sqrt{a\frac{\sigma^2}{n}}\right)$$

$$= P_\theta\left(\theta - z_{1-\alpha/2}\sqrt{a\frac{\sigma^2}{n}} \le a\bar{X} \le \theta + z_{1-\alpha/2}\sqrt{a\frac{\sigma^2}{n}}\right)$$

$$= P_\theta\left(\frac{\theta}{a} - z_{1-\alpha/2}\frac{1}{\sqrt{a}}\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \theta + z_{1-\alpha/2}\frac{1}{\sqrt{a}}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P_\theta\left(\frac{\sqrt{n}}{\sigma}\left(\frac{\theta}{a}-\theta\right) - z_{1-\alpha/2}\frac{1}{\sqrt{a}} \le Z \le \frac{\sqrt{n}}{\sigma}\left(\frac{\theta}{a}-\theta\right) + z_{1-\alpha/2}\frac{1}{\sqrt{a}}\right)$$

$$= \Phi\left\{\frac{\sqrt{n}}{\sigma}\left(\frac{\theta}{a}-\theta\right) + z_{1-\alpha/2}\frac{1}{\sqrt{a}}\right\} - \Phi\left\{\frac{\sqrt{n}}{\sigma}\left(\frac{\theta}{a}-\theta\right) - z_{1-\alpha/2}\frac{1}{\sqrt{a}}\right\}$$

Especially, when $\tau^2 \to \infty$, then $a \to 1$. Therefore,

$$\lim_{\tau^2\to\infty} P_\theta(\theta \in C_B(X)) = \Phi\{z_{1-\alpha/2}\} - \Phi\{-z_{1-\alpha/2}\} = 1-\alpha.$$

Answer 2

(a) Notice that $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x) = \frac{1}{n}\sum_{i=1}^{n} I(U_i \le F_0(x))$ where $U_i = F_0(X_i)$.

Hence, $C_n(F_0) = \int \{ \frac{1}{n}\sum_{i=1}^{n} I(U_i \le u) - u \}^2 du$.

(b)

$$
\begin{aligned}
C_n(F_0) &= \int \{ \frac{1}{n}\sum_{i=1}^{n} I(U_i \le u) - u \}^2 du \\
&= \int \left[\frac{1}{n}\sum_{i=1}^{n} \{ I(U_i \le u) - u \} \right] \left[\frac{1}{n}\sum_{j=1}^{n} \{ I(U_j \le u) - u \} \right] du \\
&= \frac{1}{n^2}\sum_i \sum_j \int \{ I(U_i \le u) - u \}\{ I(U_j \le u) - u \} du \\
&= \frac{1}{n^2}\sum_i \sum_j h(U_i, U_j)
\end{aligned}
$$

(c)

$$
\begin{aligned}
D_n(F_0) &= \sup_{-\infty < x < \infty} | F_n(x) - F_0(x) | \\
&= \sup_{-\infty < x < \infty} \left| \frac{1}{n}\sum_{i=1}^{n} \{ I(X_i \le x) - F_0(x) \} \right| \\
&= \sup_{-\infty < x < \infty} \left| \frac{1}{n}\sum_{i=1}^{n} \{ I(F_0(X_i) \le F_0(x)) - F_0(x) \} \right| \\
&= \sup_{0 < u < 1} \left| \frac{1}{n}\sum_{i=1}^{n} \{ I(U_i \le u) - u \} \right|
\end{aligned}
$$

(d) As discussed in class,

$$D_n(F) = \max\{ D_n^+(F), D_n^+(F) \},$$

where

$$D_n^+(F) = \sup \{ F_n(x) - F(x) \} = \max_{0 \le i \le n} \left(\frac{i}{n} - U_{(i)} \right),$$

$$D_n^-(F) = \sup \{ F(x) - F_n(x) \} = \max_{0 \le i \le n} \left(U_{(i+1)} - \frac{i}{n} \right),$$

where $0 \equiv U_{(0)} < U_{(1)} < \cdots U_{(n)} < U_{(n+1)} \equiv 1$ are order statistics for the uniform random variables. Then,

$$\begin{aligned}\Pr\{D_n(F)\le t\} &= \Pr\{D_n^+(F)\le t, D_n^1(F)\le t\}\\ &= \Pr\left(\frac{i}{n}-U_{(i)}\le t,\quad U_{(i+1)}-\frac{i}{n}\le t,\quad \forall i=0,\dots,n\right) \qquad (*)\\ &= \Pr\left(\frac{i}{n}-t\le U_{(i)}\le t+\frac{i-1}{n},\quad \forall i=1,\dots,n\right)\end{aligned}$$

Now we find $\Pr\{D_n(F)\le t\}$ in three cases:

Case i) $t\ge 1$: $\Pr\{D_n(F)\le t\}=1$ since $D_n(F)\le 1$ by definition.

Case ii) $t\le \frac{1}{2n}$: Then, $\frac{i}{n}-t<t+\frac{i-1}{n}$. Hence, the last equation in (*) is zero.

Case iii) $\frac{1}{2n}<t<1$:

The p.d.f. of $(U_{(1)},\cdots,U_{(n)})$ is $n!I(0<u_1<\cdots<u_n<1)$. It follows that

$$\Pr\left(\frac{i}{n}-t\le U_{(i)}\le t+\frac{i-1}{n},\quad \forall i=1,\dots,n\right)$$
$$= n!\prod_{i=1}^{n}\int_{\max\{0,\frac{i-1}{n}-t\}}^{\min\{u_{i+1},t+\frac{i-1}{n}\}} du_1\dots du_n = n!\prod_{j=1}^{n}\int_{\max\{0,\frac{n-j+1}{n}-t\}}^{\min\{u_{n-j+2},t+\frac{n-j}{n}\}} du_1\dots du_n,$$

where $j=n-i+1$.

Therefore by Cases (i)-(iii),

$$\Pr\{D_n(F)\le t\}=\left[\prod_{j=1}^{n}\int_{\max\{0,\frac{n-j+1}{n}-t\}}^{\min\{u_{n-j+2},t+\frac{n-j}{n}\}} du_1\dots du_n\right]I\left(\frac{1}{2n}<t<1\right)+I(t\ge 1).\square$$

Answer 3:

Since $f_\theta(x_i)=cx_i^{c-1}I_{(0,\infty)}(x)\exp(-x_i^c/\theta^c-c\log\theta)$, the p.d.f. of full data is

$$f_\theta(x)=c^n\prod_{i=1}^{n}x_i^{c-1}I_{(0,\infty)}(x_{(1)})\exp(-\sum_{i=1}^{n}x_i^c/\theta^c-nc\log\theta)=\exp\{\eta(\theta)Y(X)-\xi(\theta)\},$$

where $\eta(\theta)=-1/\theta^c$ is strictly increasing and $\Pr(2X_i^c/\theta^c>y)=\exp(-x/2)\sim\chi_2^2$.

Hence, $2Y(X)/\theta^c\sim\chi^2_{df=2n}$. Solving $P_{\theta_0}(Y>d)=P_{\theta_0}(\frac{2Y}{\theta_0^c}>\frac{2d}{\theta_0^c})=\alpha$,

$d=\theta_0^c\chi^2_{df=2n,1-\alpha}/2$. Hence, by Corollary 6.1, the UMP for $H_0:\theta\le\theta_0$ vs. $H_1:\theta>\theta_0$ is $T(X)=I_{(d_\alpha,\infty)}(Y(X))$, where $d_\alpha=\theta_0^c\chi^2_{df=2n,1-\alpha}/2$.

The acceptance region is

$$A(\theta_0) = \left\{ X : \sum_{i=1}^{n} X_i^c \le \theta_0^c \chi^2_{df=2n,1-\alpha} / 2 \right\}.$$

By inverting the test,

$$C(X) = \left\{ \theta : 2\sum_{i=1}^{n} X_i^c / \chi^2_{df=2n,1-\alpha} \le \theta^c \right\} = \left[\left(2\sum_{i=1}^{n} X_i^c / \chi^2_{df=2n,1-\alpha} \right)^{1/c}, \quad \infty \right).$$

Let $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$, where $\Theta_{\theta_0} = (\theta_0, \infty)$. Then,

$$\Theta' = \{\theta' : \theta \in (\theta', \infty)\} = \{\theta' : \theta > \theta'\} = (-\infty, \theta).$$

Answer 4:

$$\log \ell(\lambda) = -\sum_i \log X_i + n\overline{X} \log \lambda - n\lambda ,$$

$$s_n(\lambda) = \frac{\partial}{\partial \lambda} \log \ell(\lambda) = \frac{n\overline{X}}{\lambda} - n ,$$

MLE is $\hat{\lambda} = \overline{X}$.

1) LR test: $-2\{\log \ell(\lambda) - \log \ell(\lambda)\} = 2n\overline{X}(\log \overline{X} - \log \lambda) + n(\overline{X} - \lambda) \to_d \chi_1^2$.

$$C_1(X) = \left[2\overline{X}(\log \overline{X} - \log \lambda) + (\overline{X} - \lambda) \le \chi^2_{\alpha,1} / n \right].$$

2) Wald test: $I(\hat{\lambda})(\hat{\lambda} - \lambda)^2 = n(\overline{X} - \lambda)^2 / \overline{X} \to_d \chi_1^2$.

$$C_2(X) = \left[\overline{X} - \sqrt{(\overline{X}/n)\chi^2_{\alpha,1}},\ \overline{X} - \sqrt{(\overline{X}/n)\chi^2_{\alpha,1}} \right].$$

3) Score test: $s_n(\lambda)^2 / I(\hat{\lambda}) = n(\overline{X} - \lambda)^2 / \lambda \to_d \chi_1^2$:

$$C_2(X) = \left[\overline{X} + \frac{\chi^2_{\alpha,1}}{2n} - \sqrt{\overline{X}\,\frac{\chi^2_{\alpha,1}}{n} + \frac{(\chi^2_{\alpha,1})^2}{4n^2}},\ \overline{X} + \frac{\chi^2_{\alpha,1}}{2n} + \sqrt{\overline{X}\,\frac{\chi^2_{\alpha,1}}{n} + \frac{(\chi^2_{\alpha,1})^2}{4n^2}} \right].$$

Answer 5:

Note that the c.d.f. is $F_\theta(t) = \begin{cases} 0, & t \le 0 \\ t/\theta, & 0 < t < \theta \\ 1, & t \ge \theta \end{cases}$, and thus,

$$\theta_1 \le \theta_2 \Leftrightarrow \frac{t}{\theta_1} \ge \frac{t}{\theta_2} \Leftrightarrow F_{\theta_1}(t) \ge F_{\theta_2}(t).$$

It follows that

$$1-\alpha = \Pr(X_{(n)} \le \theta \le \alpha^{-1/n} X_{(n)}) = \Pr(F_{X_{(n)}}(t) \ge F_\theta(t) \ge F_{\alpha^{-1/n} X_{(n)}}(t)).$$

Therefore, the $(1-\alpha)$ confidence band for the c.d.f. is $\left[F_{\alpha^{-1/n} X_{(n)}}(t), \quad F_{X_{(n)}}(t)\right]$.