Statistics GIDP
Ph.D. Qualifying Exam
Theory Solution

Jan, 2017, 9:00am-1:00pm

1. The joint density of (X, Y, Z) is given by

$$f(x,y,z) = \begin{cases} \frac{1-\sin x \sin y \sin z}{8\pi^3}, & 0 \le x, y, z \le 2\pi; \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Find the density for (X, Y).

   (b) Find the density for X.

   (c) Prove X, Y, Z are pairwise independent.

   (d) Are X, Y, Z independent?

2. Let X_1,...,X_n, Y_1,...,Y_n be IID from Bernoulli($\frac{1}{2}$). Define $Z_i = X_i - Y_i$ for $i = 1, ..., n$.

   (a) Find the probability mass function of Z_1.

   (b) Find $\mathrm{E}[(\sum_{i=1}^n Z_i)^2]$.

   (c) Find $\mathrm{E}[(\sum_{i=1}^n Z_i)^4]$.

3. Suppose that X_1,..., X_n, Y_1,..., Y_m are independent random variables. $X_i \sim$Uniform$[0, \theta_1]$, $i = 1, ..., n$, and $Y_j \sim$Uniform$[0, \theta_2]$, $j = 1, ..., m$.

   (a) Find the joint density of order statistics $f_{X_{(n)},Y_{(m)}}(u, v)$, where $X_{(n)}$ and $Y_{(m)}$ are the maximum observations for each sample respectively.

   (b) Assume $\theta_1 = \theta_2 = \theta$. Find the density of Z where $Z = \max\{X_{(n)}, Y_{(m)}\}$.

   (c) If $\theta_1 = \theta_2 = \theta$, define $T = \frac{X_{(n)}^n Y_{(m)}^m}{Z^{m+n}}$. Find $P(T < c)$ for $0 < c < 1$. What is the distribution of T?

4. Let X_1 and X_2 be independent and identically distributed (iid) Bernoulli observations with $p = P(X_1 = 1) = 1 - P(X_1 = 0)$, where $0 \le p \le 1$ is unknown. It is desired to estimate $\theta = P(X_1 = X_2)$.

   (a) Specify the range of possible values for θ.

   (b) Find an unbiased estimator of θ.

   (c) Find the uniformly minimum-variance unbiased estimator (UMVUE) for θ. Justify the answer.

   (d) Is the UMVUE obtained in (c) a reasonable estimator? Justify the answer.

   (e) Find the maximum likelihood estimator (MLE) of θ.

5. Let $X_1, \ldots, X_n$ be a random sample with probability density function (pdf)

$$f(x;\theta) = \frac{2}{\sqrt{\pi\theta}} \exp\{-\frac{x^2}{\pi}\}, \quad x > 0,$$

where $\theta > 0$. (**Hint: You may use the following facts:** $\frac{2X_1^2}{\theta} \sim \chi_1^2$**, a chi-square variable with** 1 **degree of freedom.**).

(a) Find the Cramér-Rao Lower bound for estimating θ.

(b) Using the Cramér-Rao Lower bound, find the UMVUE of θ.

(c) Given $\theta_0 > 0$ and test size $\alpha \in (0,1)$, show that the likelihood ratio test for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ is given by

$$\text{Reject } H_0 \text{ if } W_n < \theta_0 c_1 \text{ or } W_n > \theta_0 c_2,$$

where $0 < c_1 < c_2 < \infty$ are appropriate constants and $W_n = 2\sum_{i=1}^n X_i^2$. Use chi-square quantiles to describe the required form of c_1 and c_2.

(d) Use the likelihood ratio test in (c) to find a confidence interval for θ with confidence coefficient $1 - \alpha$. Justify your solution.

6. Let $(X_i, Y_i), i = 1, \ldots, n$ be a random sample from a (bivariate) uniform distribution defined on a circle centered at $(0,0)$ with radius $\theta > 0$. That is, the joint probability density function of (X_1, Y_1) is given by

$$f(x,y;\theta) = \begin{cases} \frac{1}{\pi\theta^2} & \text{if } \sqrt{x^2+y^2} < \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Define $R_n = \max_{i=1,\ldots,n} \sqrt{X_i^2 + Y_i^2}$. (**Hint: if needed, you may use the following fact:** $P(X_1^2 + Y_1^2 \leq s^2) = (s/\theta)^2$ **for** $s \in [0, \theta]$.)

(a) For $0 \leq r \leq \theta$, show that $P_\theta(R_n \leq r) = (r/\theta)^{2n}$.

(b) Show that R_n is a consistent estimator of θ.

(c) Suppose a confidence interval for θ of the form $I_n = (0, R_n 5^{1/(2n+\sqrt{n})}]$. For a given $\theta > 0$, determine the coverage probability of I_n as $n \to \infty$.

Solutions:

1. (a) Density of (X,Y) is $\int_0^{2\pi} f(x,y,z)dz = \frac{1}{4\pi^2}$.

   (b) Density of X is $\int_0^{2\pi}\int_0^{2\pi} f(x,y,z)dydz = \frac{1}{2\pi}$.

   (c) By last two questions, X and Y are independent. Same for other pairs.

   (d) If X, Y, Z are independent, by (2), the joint density would be $\frac{1}{8\pi^3}$. So they are not independent.

2. (a) $P(Z_i = 0) = \frac{1}{2}$, $P(Z_i = 1) = P(Z_i = -1) = \frac{1}{4}$.

   (b) $E(Z_i^2) = \frac{1}{2}$, $E(Z_iZ_j) = E(Z_i)E(Z_j) = 0$, $i \neq j$. So $E(\sum Z_i)^2 = E\sum Z_i^2 = \frac{n}{2}$.

   (c) Write $(\sum Z_i)^4$ as a sum of degree 4 monomials. Only the monomials of type Z_i^4 and $Z_i^2Z_j^2$ have nonzero expectation. $E(Z_i^4) = \frac{1}{2}$ and $E(Z_i^2Z_j^2) = \frac{1}{4}$, $i \neq j$. There are n terms of type Z_i^4 and $\binom{n}{2}\binom{4}{2} = 3n(n-1)$ terms of type $Z_i^2Z_j^2$. So $E(\sum Z_i)^4 = \frac{n}{2} + \frac{3n(n-1)}{4}$.

3. (a) $P(X_{(n)} \leq u) = P(X_i \leq u, i = 1,..,n) = (u/\theta_1)^n$, so the density of $X_{(n)}$ is nu^{n-1}/θ_1^n. Similarly, $Y_{(m)}$ has density mv^{m-1}/θ_2^m. As they are independent, the joint density is $\frac{nu^{n-1}mv^{m-1}}{\theta_1^n\theta_2^m}$.

   (b) Given $\theta_1 = \theta_2 = \theta$, X_i's and Y_j's are IID from Uniform$[0,\theta]$, so Z has density $(n+m)z^{n+m-1}/\theta^{n+m}$.

   (c) Note that $T = \begin{cases} X_{(n)}^n/Y_{(m)}^n, & \text{when } X_{(n)} < Y_{(m)}; \\ Y_{(m)}^m/X_{(n)}^m, & \text{when } X_{(n)} > Y_{(m)}. \end{cases}$ We have

$$\begin{aligned} P(T < c) &= P(T < c, X_{(n)} < Y_{(m)}) + P(T < c, X_{(n)} > Y_{(m)}) \\ &= P(X_{(n)}/Y_{(m)} < c^{\frac{1}{n}}, X_{(n)} < Y_{(m)}) + P(Y_{(m)}/X_{(n)} < c^{\frac{1}{m}}, X_{(n)} > Y_{(m)}) \\ &= P(X_{(n)}/Y_{(m)} < c^{\frac{1}{n}}) + P(Y_{(m)}/X_{(n)} < c^{\frac{1}{m}}) \end{aligned}$$

   where the probabilities in the last line above are independent of the value of θ. So we can assume $\theta = 1$, and

$$\begin{aligned} P(X_{(n)}/Y_{(m)} < c^{\frac{1}{n}}) &= \int_0^1\int_0^{vc^{\frac{1}{n}}} nu^{n-1}mv^{m-1}dudv \\ &= \int_0^1 (vc^{\frac{1}{n}})^n mv^{m-1}dv \\ &= \int_0^1 cmv^{n+m-1}dv \\ &= c\frac{m}{m+n} \end{aligned}$$

   Similarly, the second summand equals $c\frac{n}{m+n}$ and $P(T < c) = c$ for all $c \in (0,1)$. That is, $T \sim$Uniform$(0,1)$.

4. Let X_1 and X_2 be independent and identically distributed (iid) Bernoulli observations with $p = P(X_1 = 1) = 1 - P(X_1 = 0)$, where $0 \le p \le 1$ is unknown. It is desired to estimate $\theta = P(X_1 = X_2)$.

   (a) $\theta = P(X_1 = X_2) = P(X_1 = 0, X_2 = 0) + P(X_1 = 1, X_2 = 1) = p^2 + (1-p)^2 \in [\frac{1}{2}, 1]$ for $p \in [0, 1]$. The parameter θ has a minimum value $\frac{1}{2}$ when $p = \frac{1}{2}$.

   (b) The unbiased estimator of θ is $\hat{\theta} = I(X_1 = X_2)$.

   (c) Show $S = X_1 + X_2$ is a complete and sufficient statistic. Then $\hat{\theta} = I(X_1 = X_2) = I(S \neq 1)$, is a function of S. By Lehmann-Scheffe theorem, $\hat{\theta}$ is the UMVUE of θ.

   (d) The $\hat{\theta} = I(X_1 = X_2)$ takes the value out of the range of θ, so it is not reasonable.

   (e) The MLE of p is $\hat{p}_{MLE} = (X_1 + X_2)/2$. So the MLE of θ is $\hat{\theta}_{MLE} = \hat{p}_{MLE}^2 + (1 - \hat{p}_{MLE})^2$.

5. Let $X_1, \dots, X_n$ be a random sample with probability density function (pdf)

$$f(x;\theta) = \frac{2}{\sqrt{\pi\theta}} \exp\{-\frac{x^2}{\pi}\}, \quad x > 0,$$

   where $\theta > 0$.

   (a) This is a regular exponential family. The Fisher information number based on X_1 is

$$I_1(\theta) = -E_\theta \frac{d^2 \log f(X_1|\theta)}{d\theta^2} = E_\theta \left(\frac{2X_1^2}{\theta^3} - \frac{1}{2\theta^2} \right) = \frac{1}{2\theta^2}$$

   The CRLB for estimating θ is $\frac{1}{nI_1(\theta)} = \frac{2\theta^2}{n}$.

   (b) Let $T = \frac{2\sum_{i=1}^n X_i^2}{n}$. Since $E(T) = \theta$, it is unbiased for θ. And $Var(T) = \frac{2\theta^2}{n}$, which is equal to the bound. So T is UMVUE of θ. (Note $\sum_{i=1}^n$ is a complete and sufficient statistic for θ.)

   (c) It is easy to check that the MLE is $\hat{\theta} = \frac{W}{n}$, where $W = 2\sum_{i=1}^n X_i^2$. The likelihood ratio test statistic for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ is given by

$$\lambda(\theta_0) = \frac{L(\theta_0)}{L(\hat{\theta})} = n^{-n/2}(W_{0n})^{n/2} \exp[(n - W_{0n})/2] = g(W_{0n}),$$

   where $W_{0n} = W/\theta_0$ and $g(x) = (x/n)^{n/2} \exp[(n-x)/2], x > 0$. Note that $\lim_{x\to 0^+} = 0 = \lim_{x\to\infty} g(x)$ and $g(\cdot)$ is concave down with a maximum of $g(n) = 1$ at $x = n$. In other words, $g(\cdot)$ is increasing on $(0, n])$ and decreasing on $[n, \infty)$. So for $\lambda \in (0, 1)$, $g(W_{0n}) < \lambda$ if and only if $W_{0n} < c_1$ or $W_{0n} > c_2$, where $0 < c_1 < n < c_2$ are constants satisfying $g(c_1) = g(c_2) = \lambda$ and

$$\alpha = P_{\theta_0}(W_{0n} < c_1 \text{ or } W_{0n} > c_2) = 1 - P_{\theta_0}(c_1 \le W_{0n} \le c_2).$$

   Under $H_0 : \theta = \theta_0$, $W_{0n} \sim \chi_n^2$. Let $\chi_{n,\gamma}^2$ denote the γ-quantile of a χ_n^2 variable, i.e., $P(\chi_n^2 \le \chi_{n,\gamma}^2) = \gamma \in (0, 1)$. So we pick $c_1 = \chi_{n,\gamma_1}^2$ and $c_2 = \chi_{n,\gamma_2}^2$, where $\gamma_2 - \gamma_1 = 1 - \alpha$.

(d) The acceptance region for the size-α likelihood ratio test of $H_0 : \theta = \theta_0$ is

$$A(\theta_0) = \{(X_1, \cdots, X_n) : c_1 \le 2\sum_{i=1}^{n} X_i^2/\theta_0 \le c_2\}.$$

By inverting the region, the $100(1-\alpha)\%$ confidence region of θ is given by $\{\theta > 0; (X_1, \cdots, X_n) \in A(\theta)\}$, which can be expressed as

$$\{\theta > 0; c_1 \le 2\sum_{i=1}^{n} X_i^2/\theta_0 \le c_2\} = \left[\frac{2\sum_{i=1}^n X_i^2}{c_2}, \frac{2\sum_{i=1}^n X_i^2}{c_1}\right].$$

6. Let $(X_i, Y_i), i = 1, \dots, n$ be a random sample from a (bivariate) uniform distribution defined on a circle centered at $(0, 0)$ with radius $\theta > 0$. That is, the joint probability density function of (X_1, Y_1) is given by

$$f(x, y; \theta) = \begin{cases} \frac{1}{\pi\theta^2} & \text{if } \sqrt{x^2 + y^2} < \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Define $R_n = \max_{i=1,\dots,n} \sqrt{X_i^2 + Y_i^2}$.

(a) For $0 \le r \le \theta$,

$$P_\theta(R_n \le r) = \prod_{i=1}^{n}(X_i^2 + Y_i^2 \le r^2) = (r/\theta)^{2n}.$$

(b) The variable R_n has a probability density function given by

$$f(r) = 2n\theta^{-2n}r^{2n-1}, \quad 0 <\le r \le \theta.$$

The mean and second moment of R_n are then

$$E_\theta R_n = 2n\theta^{-2n}\int_0^\theta r^{2n}dr = \frac{2n}{2n+1}\theta,$$

and

$$E_\theta R_n^2 = 2n\theta^{-2n}\int_0^\theta r^{2n+1}dr = \frac{2n}{2n+2}\theta^2,$$

So the bias of R_n converges to zero, as $E_\theta R_n - \theta = -\frac{1}{2n+1}\theta \to 0$ as $n \to \infty$. Also, $E_\theta R_n^2 \to \theta^2$ as $n \to \infty$, so that the variance of R_n converges to zero; $E_\theta R_n^2 - [E_\theta R_n]^2 \to \theta^2 - \theta^2 = 0$. Therefore, R_n is consistent for θ.

(c) For any $\theta > 0$, we have

$$\begin{aligned} P(\theta \in I_n) &= P(\theta < R_n 5^{1/(2n+\sqrt{n})}) = P(\theta 5^{-1/(2n+\sqrt{n})} \le R_n) \\ &= 1 - P(R_n < \theta 5^{-1/(2n+\sqrt{n})} = 1 - 5^{-2n/(2n+\sqrt{n})} \\ &\to 1 - 5^{-1} = 4/5, \end{aligned}$$

as $n \to \infty$.