Statistics GIDP
Ph.D. Qualifying Exam
Theory

May 25, 2017, 9:00am-1:00pm

Instructions: Provide your answers on the supplied pads of paper; write on only one side of each sheet. Please complete exactly 5 of the 6 problems. Each problem is EQUALLY weighted. Turn in only those sheets you wish to grade. Please put your code ONLY instead of your name on the answer sheets. Stay calm and do your best. Good luck with the exam.

1. A standard 6-sided die has been modified so that 3 of its faces are painted green, 2 faces are painted blue, and the final face is painted red. Painting has not affected the weighting of the die, so that on any single roll, the probability of rolling green is 1/2, the probability of rolling blue is 1/3, and the probability of rolling red is 1/6. Suppose Frank, Ginny, and Harry each roll the die once. All rolls are independent.
   (a) What is the probability that Frank, Ginny and Harry all roll the same color?
   (b) What is the probability that Frank, Ginny and Harry all roll different colors?
   (c) What is the probability that exactly two of the three people roll the same color, and the third person rolls a different color?
   (d) What is the probability that at least one of the three people roll red?

2. Let $U_1, U_2, \cdots, U_n$ be independent random variables having the uniform distribution on $[0,1]$. Define $Y_n = (\prod_{i=1}^{n} U_i)^{-1/n}$.
   (a) Define $X_1 = -\ln U_1$. Find the distribution of X_1.
   (b) Establish the asymptotic distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
   (c) Show the asymptotic distribution of Y_n.
   (d) Find the exact distribution of $\bar{X}_n$.

3. Let X_1, X_2, X_3 be three random variables from Unif$(0,\theta)$, where $\theta > 0$.
   (a) Find the distribution of $X_{(1)}/\theta$.
   (b) Compute $\mathrm{E}[X_{(1)}]$.
   (c) Find the conditional pdf of $X_{(1)}$ given $X_{(3)}$. (Recall the joint pdf of $X_{(i)}$ and $X_{(j)}$ for a random sample of $X_1, X_2, \cdots, X_n$ is

$$\begin{aligned} f_{X_{(i)},X_{(j)}}(u,v) &= \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(u)f(v) \\ &\times \; [F(u)]^{i-1}[F(v)-F(u)]^{j-1-i}[1-F(v)]^{n-j}, \end{aligned}$$

   where $f(\cdot)$ and $F(\cdot)$ are pdf and cdf of X_i's respectively.)

(d) Compute $\mathrm{E}[X_{(1)}|X_{(3)}]$.

4. Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) observations from $\mathrm{Unif}[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ with probability density function

$$f(x|\theta) = \begin{cases} 1 & \text{if } \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

where $-\infty < \theta < \infty$ is an unknown parameter. Consider the following estimator of θ

$$\hat{\theta}_1 = \frac{X_{(1)} + X_{(n)}}{2},$$

where $X_{(1)}$ and $X_{(n)}$ denote the minimum and maximum order statistics.

(a) Is $\hat{\theta}_1$ unbiased for θ? Show your answer.

(b) Find the method of moment estimator of θ, and call it $\hat{\theta}_2$. Also, identify the limiting distribution of $\hat{\theta}_2$ as $n \to \infty$.

(c) Which of the two estimators do you prefer, $\hat{\theta}_1$ or $\hat{\theta}_2$? Justify your answer.

(d) Assume θ follows a prior distribution $\mathrm{Unif}[-A, A]$, where $A > 0$. Find the Bayes estimator of θ.

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

$$f(x|\theta) = \frac{1}{x\sqrt{2\pi\theta}} \exp\left\{-\frac{(\log x - \theta)^2}{2\theta}\right\}, \quad x > 0,$$

where $\theta > 0$ is an unknown parameter.

(a) Find the maximum likelihood estimator (MLE) for θ. Call it $\hat{\theta}$.

(b) Find a complete and sufficient statistic for θ.

(c) Find the uniform minimum variance unbiased estimator (UMVUE) for $\theta^2 + \theta$.

(d) Show how to construct a UMVUE of $a\theta + b\theta^2$, for any given constants a and b. You do not need to give the explicitly formula for the UMVUE.

6. Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) observations from a distribution with the probability density function

$$f(x|\mu, \sigma) = \begin{cases} \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \text{if } x > \mu, \\ 0 & \text{otherwise,} \end{cases}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$ are unknown parameters.

(a) Show that $\frac{X_1 - \mu}{\sigma}$ has the same distribution as $|Z|$, where Z is a standard normal variable.

(b) Identify a set of sufficient statistics for the unknown parameters (μ, σ^2).

(c) Derive the maximum likelihood estimator (MLE) of (μ, σ^2).

(d) Suppose that $\mu = 0$ is known. Derive the UMP test for $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma = \sigma_1$ for two fixed parameter values $\sigma_1 > \sigma_0 > 0$ at the level $\alpha > 0$. Specify the rejection region.

2017 May - Theory

1. A standard 6-sided die has been modified so that 3 of its faces are painted green, 2 faces are painted blue, and the final face is painted red. Painting has not affected the weighting of the die, so that on any single roll, the probability of rolling green is 1/2, the probability of rolling blue is 1/3, and the probability of rolling red is 1/6. Suppose Frank, Ginny, and Harry each roll the die once. All rolls are independent.

   (a) What is the probability that Frank, Ginny and Harry all roll the same color?

   (b) What is the probability that Frank, Ginny and Harry all roll different colors?

   (c) What is the probability that exactly two of the three people roll the same color, and the third person rolls a different color?

   (d) What is the probability that at least one of the three people roll red?

2. Let $U_1, U_2, \cdots, U_n$ be independent random variables having the uniform distribution on $[0,1]$. Define $Y_n = (\prod_{i=1}^{n} U_i)^{-1/n}$.

   (a) Define $X_1 = -\ln U_1$. Find the distribution of X_1.

   (b) Establish the asymptotic distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

   (c) Show the asymptotic distribution of Y_n.

   (d) Find the exact distribution of $\bar{X}_n$.

3. Let X_1, X_2, X_3 be three random variables from Unif$(0, \theta)$, where $\theta > 0$.

   (a) Find the distribution of $X_{(1)}/\theta$.

   (b) Compute $\mathrm{E}[X_{(1)}]$.

   (c) Find the conditional pdf of $X_{(1)}$ given $X_{(3)}$. (Recall the joint pdf of $X_{(i)}$ and $X_{(j)}$ for a random sample of $X_1, X_2, \cdots, X_n$ is

   $$\begin{aligned} f_{X_{(i)},X_{(j)}}(u,v) &= \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(u)f(v) \\ &\times [F(u)]^{i-1}[F(v)-F(u)]^{j-1-i}[1-F(v)]^{n-j}, \end{aligned}$$

   where $f(\cdot)$ and $F(\cdot)$ are pdf and cdf of X_i's respectively.)

   (d) Compute $\mathrm{E}[X_{(1)}|X_{(3)}]$.

4. Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) observations from $\text{Unif}[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ with probability density function

$$f(x|\theta) = \begin{cases} 1 & \text{if } \theta - \frac{1}{2} \le x \le \theta + \frac{1}{2}, \\ 0 & \text{otherwise}, \end{cases}$$

where $-\infty < \theta < \infty$ is an unknown parameter. Consider the estimator

$$\hat{\theta}_1 = \frac{X_{(1)} + X_{(n)}}{2},$$

where $X_{(1)}$ and $X_{(n)}$ denote the minimum and maximum order statistics.

(a) Is $\hat{\theta}_1$ unbiased for θ? Explain your answer.

(b) Find the method of moment estimator of θ, say $\hat{\theta}_2$. Also, identify the limiting distribution of $\hat{\theta}_2$, as $n \to \infty$.

(c) Which of the two estimators do you prefer, $\hat{\theta}_1$ or $\hat{\theta}_2$? Explain your answer.

(d) Assume θ follows a prior distribution $\text{Unif}[-A, A]$, where $A > 0$. Find the Bayes estimator of θ.

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

$$f(x|\theta) = \frac{1}{x\sqrt{2\pi\theta}} \exp\left\{-\frac{(\log x - \theta)^2}{2\theta}\right\}, \quad x > 0,$$

where $\theta > 0$ is an unknown parameter.

(a) Find the maximum likelihood estimator (MLE) for θ. Call it $\hat{\theta}$.

(b) Find a complete and sufficient statistic for θ.

(c) Find the uniform minimum variance unbiased estimator (UMVUE) for $\theta^2 + \theta$.

(d) Show how to construct a UMVUE of $a\theta + b\theta^2$, for any given constants a and b. You do not need to give the explicitly formula for the UMVUE.

6. Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) observations from a distribution with the probability density function

$$f(x|\mu, \sigma) = \begin{cases} \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \text{if } x > \mu, \\ 0 & \text{otherwise}, \end{cases}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$ are unknown parameters.

(a) Show that $\frac{X_1-\mu}{\sigma}$ has the same distribution as $|Z|$, where Z is a standard normal variable.

(b) Identify a set of sufficient statistics for the unknown parameters (μ, σ^2).

(c) Derive the maximum likelihood estimator (MLE) of (μ, σ^2).

(d) Suppose that $\mu = 0$ is known. Derive the UMP test for $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma = \sigma_1$ for two fixed parameter values $\sigma_1 > \sigma_0 > 0$ at the level $\alpha > 0$. Specify the rejection region.

Solutions:

1. (a) prob=$(1/2)^3 + (1/3)^3 + (1/6)^3 = 1/6$.

   (b) prob=$3! * \frac{1}{2} * \frac{1}{3} * \frac{1}{6} = 1/6$.

   (c) prob=$3 * (\frac{1}{2} * \frac{1}{2} * \frac{1}{2} + \frac{1}{3} * \frac{1}{3} * \frac{2}{3} + \frac{1}{6} * \frac{1}{6} * \frac{5}{6}) = 2/3$ or use $1 - 1/6 - 1/6 = 2/3$.

   (d) prob=$1 - (5/6)^3$.

2. (a) $X_i \sim Exp(1)$.

   (b) By central limit theorem, $\sqrt{n}(\bar{X}_n - 1) \xrightarrow{d} N(0, 1)$.

   (c) By delta method, $\sqrt{n}(Y_n - e) \xrightarrow{d} N(0, e^2)$

3. (a) $P(X_{(1)}/\theta \leq x) = 1 - P(X_{(1)} > \theta x) = 1 - (1-x)^3$, thus the pdf is $f(x) = 3(1-x)^2$.

   (b) From Beta(1,3) E$[X_{(1)}/\theta] = 1/4$. Hence E$[X_{(1)}] = \theta/4$.

   (c) The joint pdf of $X_{(1)}$ and $X_{(3)}$ is

   $$f_{X_{(1)},X_{(3)}}(u, v) = \frac{6}{\theta^3}(v - u).$$

   And the marginal pdf of $X_{(3)}$ is $f_{X_{(3)}}(v) = 3v^2/\theta^3$. Thus the conditional pdf is $f_{X_{(1)}|X_{(3)}}(u|v) = 2/v - 2u/v^2$ where $0 < u < v$.

   (d) From the above density we can easily get the conditional expectation is $X_{(3)}/3$.

4. (a) Define $Y_i = X_i - \theta$ for $i = 1, \ldots, n$. Then Y_i is iid Unif$[-\frac{1}{2}, \frac{1}{2}]$. By symmetry, $Y_{(1)}$ and $-Y_{(n)}$ have the same distribution. So

   $$E(\hat{\theta}_1 - \theta) = E\frac{X_{(1)} - \theta}{2} + E\frac{X_{(n)} - \theta}{2} = \frac{1}{2}EY_{(1)} + \frac{1}{2}Y_{(n)} = 0,$$

   so $\hat{\theta}_1$ is unbiased.

   (b) $E(X_1) = \theta$, so $\hat{\theta}_2 = \bar{X}$. By CLT, we have

   $$\sqrt{n}(\bar{X} - \theta) \to N(0, \frac{1}{12}).$$

   (c) Compare the variance. Var$(\hat{\theta}_2) = \frac{1}{12n}$.

   $$Var(\hat{\theta}_1) = Var(\frac{Y_{(1)} + Y_{(n)}}{2}) = \frac{1}{2}Var(Y_{(1)}) = \frac{1}{2}\frac{n}{(n+1)^2(n+2)},$$

   prefer $\hat{\theta}_1$ when n is large.

(d) The posterior density of θ is

$$\begin{aligned}\pi(\theta|X) &\propto I(|\theta| \leq A)I(\theta \in [X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}])\\ &\propto I\left[\max(X_{(n)} - \frac{1}{2}, -A) \leq \theta \leq \min(X_{(1)} + \frac{1}{2}, A)\right].\end{aligned}$$

Therefore,

$$\hat{\theta}_{Bayes} = \frac{1}{2}\left(\max(X_{(n)} - \frac{1}{2}, -A) + \min(X_{(1)} + \frac{1}{2}, A)\right)$$

5. (a) Let $Y_i = \log X_i$, $i = 1, \cdots, n$, and $T = \frac{1}{n}\sum_{i=1}^n Y_i^2$. The log-likelihood function is

$$\begin{aligned}L(\theta) &= \sum_{i=1}^n \log f_\theta(X_i) = -\frac{n}{2}\log(2\pi) - \sum_{i=1}^n Y_i - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^n (Y_i - \theta)^2\\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}(\log\theta + \frac{T}{\theta} + \theta).\end{aligned}$$

Therefore, $\frac{d}{d\theta}L_n(\theta) = -\frac{n}{2}(1 + \frac{1}{\theta} - \frac{T}{\theta^2})$. Solving the likelihood equation, we have $\hat{\theta} = \frac{-1+\sqrt{1+4T}}{2}$.

(b) This is one-dimensional full rank exponential family

$$f_\theta(x) = \exp\left[-\frac{1}{2}\log(x) - \frac{\theta}{2} - \frac{1}{2\theta}(\log x)^2\right]],$$

with the set $\{-\frac{1}{2\theta} : \theta > 0\} = (-\infty, 0)$ being an open set. Therefore, $T = \sum_{i=1}^n Y_i^2 = \sum_{i=1}^n (\log X_i)^2$ is a complete and sufficient statistic.

(c) $Y_i \sim N(\theta, \theta)$. Since $E(Y_i^2) = \theta + \theta^2$, So $E(T/n) = \theta + \theta^2$ and T/n is UMVUE by Rao-Blackwell.

(d) Note that

$$E(\alpha Y_1 + \beta Y_1^2) = \alpha\theta + \beta(\theta^2 + \theta) = (\alpha + \beta)\theta + \beta\theta^2.$$

Choose $\beta = b, \alpha = a - b$, then $g(Y_1) = (a - b)Y_1 + bY_1^2$ satisfies that $E(g(Y_1)) = a\theta + b\theta^2$. Then UMVUE is $E(g(Y_1)|T)$.

6. (a) For $z > 0$, we have

$$P(|Z| \leq z) = P(-z \leq Z \leq z) = \Phi(z) - \Phi(-z),$$

so the density of $|Z|$ is $2\phi(z), z > 0$, where ϕ is the standard normal density. Using the transformation $Y = \frac{X_1 - \mu}{\sigma}$, we get the same density $2\phi(y)$ for $y > 0$.

(b) The density function is proportional to

$$(\sigma^2)^{-n/2}\exp\{-\frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2}\}I(X_{(1)}\geq \mu).$$

By the factorization theorem, the sufficient statistics are $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, X_{(1)})$.

(c) The likelihood function is increasing at $(-\infty, X_{(1)})$, so the maximum likelihood estimator (MLE) is

$$\hat{\mu} = X_{(1)}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - X_{(1)})^2.$$

(d) Suppose that $\mu = 0$ is known. By Neymann-Pearson, the UMP test reject H_0 if $\frac{L(\sigma_1)}{L(\sigma_0)} > K$, which is equivalent to $\sum_{i=1}^n X_i^2 \geq C$, where C is chosen such that $P(\sum_{i=1}^n X_i^2 \geq C|H_0) = \alpha$. Note that

$$P(\sum_{i=1}^n X_i^2 \geq C|H_0) = P(\sum_{i=1}^n X_i^2/\sigma_0^2 \geq C/\sigma_0^2) = P(\chi_n^2 \geq C) = \alpha,$$

so $C = \sigma_0^2\chi^2_{n,\alpha}$.