19

Probability and Statistics Preliminary Examination: August 2014

Instructions:

- Work all 6 Problems. Neither calculators nor electronic devices of any kind are allowed. Clearly state any theorem or fact that you use. Each of the 6 Problems is equally weighted.
- Abbreviations/Acronyms.
  - pmf (probability mass function); pdf (probability density function); cdf (cumulative distribution function); mgf (moment generating function); iid (independent and identically distributed).
  - MOME (method of moments estimator); MLE (maximum likelihood estimator); PBE (posterior Bayes estimator); UMVUE (uniform minimum variance unbiased estimator); UMP (uniformly most powerful); LRT (likelihood ratio test); MLR (monotone likelihood ratio).
- Notation.
  - $I(x \in A)$ or $I_A(x)$: indicator function for set A; takes on the value 1 if $x \in A$ and 0 otherwise.
  - $\mathbb{E}(X)$: expectation of random variable X.
  - $\mathbb{V}(X)$: variance of random variable X.
  - $X \sim \mathrm{N}(a, b)$: X has a normal distribution with mean a and variance b.
- Common distributions and other results.

Exponential(λ): $\mathbb{E}(X) = \lambda$, $\mathbb{V}(X) = \lambda^2$, and pdf and mgf

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} I(x > 0), \qquad M(t) = \left(\frac{1}{1 - \lambda t}\right), \quad t < 1/\lambda.$$

Beta(α, β): $\mathbb{E}(X) = \alpha/(\alpha + \beta)$, $\mathbb{V}(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, and pdf

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1)$$

Gamma(α, β): $\mathbb{E}(X) = \alpha\beta$, $\mathbb{V}(X) = \alpha\beta^2$, and pdf and mgf

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I(x > 0), \qquad M(t) = \left(\frac{1}{1 - \beta t}\right)^\alpha, \quad t < 1/\beta.$$

Order Statistics: Let $X_{(1)} \le \cdots \le X_{(n)}$ denote the order statistics from a random sample $X_1, \ldots, X_n$. If X_1 is continuous with pdf $f(x)$ and cdf $F(x)$, the pdf of $X_{(j)}$, $(X_{(i)}, X_{(j)})$, and $(X_{(1)}, \ldots, X_{(n)})$, are given by:

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} [F(x)]^{j-1} [1 - F(x)]^{n-j} f(x) I(-\infty < x < \infty)$$

$$f_{X_{(i)}, X_{(j)}}(x_i, x_j) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(x_i) f(x_j) [F(x_i)]^{i-1} [F(x_j) - F(x_i)]^{j-1-i} [1 - F(x_j)]^{n-j} \times I(-\infty < x_i \le x_j < \infty)$$

$$f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n) I(-\infty < x_1 \le \cdots \le x_n < \infty)$$

19

1. Consider a *forced binary choice test* where a subject is forced to choose between two possible choices, A & B, only one of which is correct (e.g., can you tell Coke from Pepsi in a taste test?). It may be that the subject cannot actually discriminate between A & B (which occurs with probability $1-\theta$), but a choice must be made, in which case the subject will guess correctly with probability 1/2. On the other hand, if a subject can discriminate between A & B (which occurs with probability θ), a correct choice will be made. Suppose that n randomly selected subjects participate in a forced binary choice test, each subject being asked to discriminate between A & B once. Letting Y denote the number of subjects that make the correct choice, and noting that θ (modeled as random variable Θ) varies from subject to subject, a plausible hierarchical model for this situation is as follows:

$$c = P(\text{subject makes correct choice}), \qquad Y|C=c \sim \text{Binomial}(n,c), \qquad \Theta \sim \text{Beta}(a,b).$$

   (a) Show that $c=(1+\theta)/2$, and hence conclude that $1/2 < c < 1$ if $0<\theta<1$.

   (b) Compute $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.

   (c) Show that

$$\mathbb{E}[(1+\Theta)^n] = \sum_{k=0}^{n} \binom{n}{k} \frac{\Gamma(a+k)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+k)}.$$

   (d) Find the (marginal) distribution of Y. [Hint: use (c) to get a closed form for the integral.]

2. For parameters $0<\lambda_1<\lambda_2<\infty$, let the bivariate random vector (X,Y) have joint pmf given by

$$f(x,y|\lambda_1,\lambda_2) = \frac{e^{-\lambda_2}}{(y-x)!x!}\lambda_1^x(\lambda_2-\lambda_1)^{y-x} I_{\{0,1,\ldots,y\}}(x) I_{\{0,1,\ldots\}}(y).$$

Assume that a random sample of size n from this joint pmf, $(X_1,Y_1),\ldots,(X_n,Y_n)$, is available to make inference on (λ_1,λ_2). Define the parameter $\theta=\lambda_1/\lambda_2$.

   (a) Find the marginal distributions of X and Y.

   (b) Show that $X|Y=y \sim \text{Binomial}(y,\theta)$, for $y=1,2,\ldots$. What is the distribution of $X|Y=0$?

   (c) Compute the correlation coefficient between X and Y.

   (d) Does an UMVUE of θ exist? Justify. [Note that you are not being asked to find the UMVUE!]

3. Let $X_1,\ldots,X_n$ be a random sample from a $\text{N}(0,\sigma^2)$. Let $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ be the usual sample mean, and define the statistic $T=n\overline{X}^2/\sum_{i=1}^n X_i^2$.

   (a) What is the *exact* distribution of $X_1\sqrt{(n-1)/\sum_{i=2}^n X_i^2}$?

   (b) Find the limiting (asymptotic) distribution of nT.

   (c) Just for this part of the Problem: assume $X_1,\ldots,X_n$ are a random sample from a $\text{N}(\mu,\sigma^2)$. Compute the LRT statistic for $H_0:\mu=0$ vs. $H_1:\mu\neq 0$, and show that it can be expressed as a function of T.

   (d) Find the limiting (asymptotic) distribution of $-n\log(1-T)$.

19

4. Let $f(x)$ be the pdf of continuous random variable X with support $\mathcal{X}$. A common problem in Monte Carlo integration is to estimate the quantity $\theta = \mathbb{E}[h(X)]$, where $h(\cdot)$ is a known function satisfying $\mathbb{V}[h(X)] < \infty$. To this end, one can simulate $X_1, \ldots, X_n$, a random sample from X, and estimate θ with $\theta_n = \frac{1}{n}\sum_{i=1}^n h(X_i)$. However, when the distribution of X is hard to simulate from, one can instead simulate a random sample $Y_1, \ldots, Y_n$ from Y with pdf $g(y)$ and support $\mathcal{Y} = \mathcal{X}$. Two alternative estimators of θ are then:

$$\tau_n = \frac{1}{n}\sum_{j=1}^n \frac{f(Y_j)}{g(Y_j)} h(Y_j), \qquad \text{and} \qquad \nu_n = \sum_{i=1}^n \left[\frac{f(Y_i)/g(Y_i)}{\sum_{j=1}^n \frac{f(Y_j)}{g(Y_j)}}\right] h(Y_i).$$

The idea here is to use a distribution Y with the same support as X, but which is easier to simulate from. We also need to assume that for every $x \in \mathcal{X}$, $f(x)/g(x) \le M < \infty$.

(a) Show that θ_n is a consistent estimator of θ, and find its MSE.

(b) Show that τ_n is also a consistent estimator of θ, and that it is unbiased (for any n).

(c) Show that ν_n is also a consistent estimator of θ. State one reason (with justification) why it might be advantageous to use the biased estimator ν_n instead of the unbiased τ_n.

5. Let $Y_1, \ldots, Y_N$ be iid from an Exponential(θ) distribution, where Y_j denotes the lifetime of component $j = 1, \ldots, N$, and let $f(y)$ and $F(y)$ denote respectively the pdf and cdf of Y_1. We observe the first n failures, $X_1 = Y_{(1)} \le \cdots \le X_n = Y_{(n)}$, where $n \le N$, and both n and N are known. The observed sample in this problem thus consists of $X_1, \ldots, X_n$, and the aim is to make inference on $\mathbb{E}(Y_1) = \theta > 0$.

(a) Using heuristic arguments about a multinomial with N trials and $n+2$ categories with corresponding success probabilities $\{F(x_1), f(x_1), \ldots, f(x_n), 1 - F(x_n)\}$, or otherwise, show that the joint pdf of $(X_1, \ldots, X_n)$ is given by:

$$g(x_1, \ldots, x_n) = \frac{N!}{0!1!\cdots 1!(N-n)!}[F(x_1)]^0 f(x_1)^1 \cdots f(x_n)^1 [1 - F(x_n)]^{N-n} I(x_1 \le \cdots \le x_n).$$

(b) Show that $T = (N-n)X_n + \sum_{i=1}^n X_i$ is *complete* and *sufficient* for θ.

(c) Using the fact that $X_j = E_1 + \cdots + E_j$, where the E_i are independent Exponentials with $\mathbb{E}(E_i) = \theta/(N-i+1)$, for $i = 1, \ldots, n$, show that T as defined in (b) has a Gamma(n, θ) distribution.

(d) The *Inverted Gamma*(α, β) distribution is defined to be the continuous random variable Z with pdf: $f_Z(z) = \beta^\alpha[\Gamma(\alpha)]^{-1} z^{-\alpha-1} e^{-\beta/z} I(z > 0)$. Show that for $\alpha > 1$ and $\beta > 0$, $\mathbb{E}(Z) = \beta/(\alpha - 1)$.

6. For the setting of Problem 5, answer the following.

(a) Find the MLE of θ and its MSE.

(b) Find the UMVUE of θ and its MSE.

(c) Find the PBE of θ and its MSE under a prior that is proportional to: $\pi(\theta) \propto \theta^{-2} e^{-1/\theta} I(\theta > 0)$.

(d) Find a $(1-\alpha)100\%$ confidence interval for θ using an *exact* (finite n) pivot.

(e) Find the power function of the UMP level α test of $H_0 : \theta \le 1$ vs. $H_1 : \theta > 1$.