Probability and Statistics Preliminary Examination: August 2020

Instructions:

- Work all 6 problems.
- Neither calculators nor electronic devices of any kind are allowed.
- Clearly state any theorem or fact that you use.
- The 6 problems are equally weighted (but the parts within a problem may not be).

Abbreviations/Acronyms:

- pmf – probability mass function
- pdf – probability density function
- cdf – cumulative distribution function
- mgf – moment generating function
- iid – independent and identically distributed (i.e. a random sample)
- MSE – mean squared error
- MoM or MOME – method of moments estimator
- MLE – maximum likelihood estimator
- PBE – posterior Bayes estimator
- UMVUE or MVUE – (uniform) minimum variance unbiased estimator
- UMP – uniformly most powerful
- LRT – likelihood ratio test

Notation:

- $I(x \in A)$ or $I_A(x)$ – Indicator function for the set A; takes on the value 1 if $x \in A$ and 0 otherwise. Alternately $I(S)$ takes on the value 1 if the statement S is true and is 0 otherwise. For example, $I(x > 0)$.
- $\mathbb{E}(X)/\mathbb{V}(X)$ – Expected value/variance of X (respectively).
- "$\sim$" – Interpreted as "is distributed as". For example $X \sim N(\mu, \sigma)$ indicates that X has a normal distribution with mean μ and standard deviation σ.
- $\stackrel{p}{\rightarrow}$, $\stackrel{as}{\rightarrow}$, and $\stackrel{d}{\rightarrow}$ denote convergence "in probability", "almost surely", and "in distribution", respectively.
- $\mathbb{Z}$ – the set of all integers, $\mathbb{N}$ – the set of natural numbers (i.e. positive integers), $\mathbb{R}$ – the set of all real numbers.
- $\exp(x)$ is an alternative notation for e^x, where e is the Euler constant.

Common distributions and other results:

Note: $f(x)$ denotes the pmf/pdf, and $M(t)$ denotes the mgf. $\Gamma(x)$ denotes the gamma function.

Binomial(n, p): $\mathbb{E}(X) = np, \mathbb{V}(X) = np(1-p)$, $f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} I(x \in \{0,1,\dots,n\})$; $n \in \{1,2,\dots\}$, $p \in [0,1]$, $M(t) = [pe^t + (1-p)]^n$.

Discrete Uniform(α, β): $\mathbb{E}(X) = \frac{\alpha+\beta}{2}, \mathbb{V}(X) = \frac{(\beta-\alpha+1)^2-1}{12}$, $f(x) = \frac{1}{\beta-\alpha+1} I(x \in \{\alpha, \alpha+1, \dots, \beta-1, \beta\})$; $\alpha, \beta \in \mathbb{Z}$, $\alpha \leq \beta$, $M(t) = \frac{e^{\alpha t} - e^{(\beta+1)t}}{(\beta-\alpha+1)(1-e^t)}$. Note: $\alpha = 1, \beta = N \in \mathbb{N}$ is a common special case.

Geometric(p): $\mathbb{E}(X) = \frac{1}{p}, \mathbb{V}(X) = \frac{1-p}{p^2}$, $f(x) = p(1-p)^{x-1} I(x \in \mathbb{N})$; $p \in [0,1]$, $M(t) = \frac{pe^t}{1-(1-p)e^t}$; $t < -\ln(1-p)$.

Poisson(λ): $\mathbb{E}(X) = \lambda, \mathbb{V}(X) = \lambda$, $f(x) = \frac{e^{-\lambda}\lambda^x}{x!} I(x \in \{0,1,2, \dots\})$, $M(t) = \exp\{\lambda(e^t - 1)\}$.

Beta(α, β): $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}, \mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(x \in [0,1])$; $\alpha, \beta > 0$, $M(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$.

Exponential(β): $\mathbb{E}(X) = \beta, \mathbb{V}(X) = \beta^2$, $f(x) = \frac{1}{\beta} e^{-x/\beta} I(x \geq 0)$; $\beta > 0$, $M(t) = \frac{1}{1-\beta t}$; $t < \frac{1}{\beta}$.

Gamma(α, β): $\mathbb{E}(X) = \alpha\beta, \mathbb{V}(X) = \alpha\beta^2$, $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I(x \geq 0)$; $\alpha, \beta > 0$, $M(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$; $t < \frac{1}{\beta}$. Note: If $\alpha = p/2, \beta = 2$, then the distribution is known as the **Chi squared**(p) distribution with p being the degrees of freedom (notation: $X \sim \chi_p^2$).

Normal(μ, σ^2): $\mathbb{E}(X) = \mu, \mathbb{V}(X) = \sigma^2$, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} I(x \in \mathbb{R})$; $\mu \in \mathbb{R}, \sigma > 0$, $M(t) = \exp(\mu t + \sigma^2 t^2/2)$. Note: Alternative notation is **N**(μ, σ^2). The N(0,1) distribution is known as the standard normal.

Uniform(α, β): $\mathbb{E}(X) = \frac{\alpha+\beta}{2}, \mathbb{V}(X) = \frac{(\beta-\alpha)^2}{12}$, $f(x) = \frac{1}{\beta-\alpha} I(x \in [\alpha, \beta])$; $\alpha, \beta \in \mathbb{R}$, $\alpha \leq \beta$, $M(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta-\alpha)t}$. Note: Alternative notation is **U**(α, β).

Order Statistics: Let $X_1, \dots, X_n$ be continuous and iid with pdf $f(x)$ and cdf $F(x)$. Denote the order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$. Then the pdf of $X_{(j)}$, $(X_{(i)}, X_{(j)})$ $(i < j)$, and $(X_{(1)}, \dots, X_{(n)})$ are given by:

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

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

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n!\, f(x_1) \cdot \ldots \cdot f(x_n) I(x_1 \leq \cdots \leq x_n)$$

1. The joint density of (X, Y, Z) is given by
$$f(x, y, z) = \frac{1 - \sin x \sin y \sin z}{8\pi^3}; 0 \le x, y, z \le 2\pi.$$
   (a) Find the joint density for (X, Y).
   (b) Find the marginal density for X.
   (c) Prove that X, Y, and Z are pairwise independent. Justify/explain your answer.
   (d) Are X, Y, and Z mutually independent? Justify/explain your answer.

2. Let $X_1, \dots, X_n$ be iid U(1,2) random variables. Let H_n denote the harmonic mean
$$H_n = \frac{n}{\sum_{i=1}^n X_i^{-1}}.$$
   (a) Show that $H_n \xrightarrow{p} c$ as $n \to \infty$ for some constant c. Identify the constant c (i.e. give its value).
   (b) Show that $\sqrt{n}(H_n - c)$ converges in distribution, and identify the limit distribution. (Be sure to specify all parameters of the resulting limit distribution as well.)

3. Let $X_1, \dots, X_n$ be iid U$(0, \theta)$ random variables. Suppose that θ is a random variable with a U$(0, \beta)$ prior distribution.
   (a) Find the posterior distribution for θ, $\pi(\theta|\underline{\boldsymbol{x}})$, where $\underline{\boldsymbol{x}} = (x_1, \dots, x_n)$ denotes the realizations.
   (b) Let $\beta = 1$. Find a $(1 - \alpha) \times 100\%$ lower Bayesian credible bound for θ. That is, find θ_L such that $P(\theta_L \le \theta \le 1|\underline{\boldsymbol{x}}) = 1 - \alpha$.
   (c) Let $U = X_{(n)}/\theta$. Show that U is a pivotal quantity and use it to find a $(1 - \alpha) \times 100\%$ lower confidence bound for θ. Note that the word "confidence" indicates this is a Frequentist (i.e. non-Bayesian) problem.

4. Let $X_1, \dots, X_n$ be a random sample from a N$(\mu, 1)$ population with unknown μ. Suppose that one forgets to record the values of $X_1, \dots, X_n$ in a study and instead only records $Y_i = I(X_i > 0)$ for $i = 1, \dots, n$.
   (a) Find the MLE of μ based on the observed data, $\underline{\boldsymbol{Y}} = (Y_1, \dots, Y_n)$.
   (b) Let $T = \sum_{i=1}^n Y_i$. Is T a sufficient statistic for μ? Is T a complete statistic for μ? Justify/explain your answers.
   (c) Use the observed data $\underline{\boldsymbol{Y}}$ to construct the level-α UMP test for $H_0: \mu \le \mu_0$ vs. $H_1: \mu > \mu_0$. Please describe the form of the rejection region and simplify the expression as much as you can. You may use a normal approximation to compute the cut-off value for the rejection region.

5. Let $X_1, \dots, X_n$ be a random sample with pdf $f(x|\theta) = \frac{2}{\sqrt{\pi\theta}} \exp\left\{-\frac{x^2}{\theta}\right\} I(x > 0);\ \theta > 0$. Note that $\frac{2X_i^2}{\theta} \sim \chi_1^2$.
   (a) Find the Cramér-Rao Lower Bound for unbiased estimators of θ.
   (b) Find the UMVUE for θ.
   (c) Show that the LRT for $H_0: \theta = \theta_0$ vs. $H_1: \theta \ne \theta_0$ $(\theta_0 > 0)$ has rejection region given by
   Reject H_0 if $W_n \le \theta_0 c_1$ or $W_n \ge \theta_0 c_2$,
   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/critical values to specify c_1 and c_2 for a size-α test.

6. Let $X_1, \ldots, X_n$ be a random sample with cdf $F(x)$. The *empirical distribution function* is defined as

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq t)\,;\ -\infty < t < \infty.$$

Assume t is any fixed constant and consider estimation of $F(t)$.

(a) Show that $\hat{F}_n(t)$ is an unbiased estimator of $F(t)$.

(b) Specify the distribution of $\hat{F}_n(t)$. Be sure to specify the values of any parameters.

(c) For the fixed t, show that $\sqrt{n}\left(\hat{F}_n(t) - F(t)\right) \xrightarrow{d} N(0, v(t))$ and determine the value of $v(t)$.

(d) Based on the result given in (c), give an approximate large-sample $(1-\alpha) \times 100\%$ confidence interval for $F(t)$.