Probability and Statistics Preliminary Examination: January 2021

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.
- $\overset{p}{\to}$, $\overset{as}{\to}$, and $\overset{d}{\to}$ 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): $\mathbb{E}(X)=\frac{\alpha+\beta}{2}, \mathbb{V}(X)=\frac{(\beta-\alpha+1)^2-1}{12}, f(x)=\frac{1}{\beta-\alpha+1}I(x\in\{\alpha,\alpha+1,\dots,\beta-1,\beta\})$; $\alpha,\beta\in\mathbb{Z}, \alpha\le\beta, M(t)=\frac{e^{\alpha t}-e^{(\beta+1)t}}{(\beta-\alpha+1)(1-e^t)}$. Note: $\alpha=1,\beta=N\in\mathbb{N}$ is a common special case.

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

Negative Binomial(r,p): $\mathbb{E}(X)=\frac{r(1-p)}{p}, \mathbb{V}(X)=\frac{r(1-p)}{p^2}, f(x)=\binom{r+x-1}{x}p^r(1-p)^xI(x\in\{0,1,\dots\});\ p\in[0,1], M(t)=\left(\frac{p}{1-(1-p)e^t}\right)^r; t<-\ln(1-p)$.

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

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

Uniform(α,β): $\mathbb{E}(X)=\frac{\alpha+\beta}{2}, \mathbb{V}(X)=\frac{(\beta-\alpha)^2}{12}, f(x)=\frac{1}{\beta-\alpha}I(x\in[\alpha,\beta]); \alpha,\beta\in\mathbb{R}, \alpha\le\beta$, $M(t)=\frac{e^{\beta t}-e^{\alpha t}}{(\beta-\alpha)t}$. 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 $X_{(1)}\le\cdots\le X_{(n)}$. Then the pdf of $X_{(j)}, (X_{(i)},X_{(j)})$ $(i<j)$, and $(X_{(1)},\dots,X_{(n)})$ are given by:

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

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

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

1. Suppose that $X_1, \dots, X_n$ is a random sample of lifetimes (in hours) of 100 ohm resistors put on test by the manufacturer. It is assumed that the population has an exponential distribution with unknown mean $\beta > 0$. The manufacturer's testing procedure only records the lifetimes rounded to the next whole hour. That is, only $Y_1, \dots, Y_n$ are observed, where $Y_i = \lceil X_i \rceil$. ($\lceil z \rceil$ is the *ceiling function*, which is defined to be the smallest integer greater-than-or-equal-to z. That is, if k is the integer such that $k - 1 < z \le k$ then $\lceil z \rceil = k$.)
   (a) Show that $Y_1, \dots, Y_n$ is then a random sample from a Geometric(p) population for some p. Give the value of p.
   (b) Find the MLE of β, $\hat{\beta}$. (Remember, only $Y_1, \dots, Y_n$ is observed!)
   (c) Find a minimal sufficient statistic for β, $T(Y_1, \dots, Y_n)$. Is it also complete?
   (d) Propose a level-α test for $H_0: \beta \le \beta_0$ vs. $H_1: \beta > \beta_0$, including describing the rejection region. Is it the UMP level-α test? Explain. (A large-sample test of approximately level-α is acceptable.)
   (Note: $\sum_{i=1}^{n} Y_i - n$ has a Negative Binomial(n, p) distribution.)

2. Suppose that $X_1, \dots, X_n$ is a random sample from a Uniform($0, \theta_1$) population and that $Y_1, \dots, Y_m$ is a random sample from a Uniform($0, \theta_2$) population such that X_i and Y_j are independent for all i, j.
   (a) Find the joint pdf of the maximum order statistics $X_{(n)}$ and $Y_{(m)}$.
   (b) Assume $\theta_1 = \theta_2 = \theta$. Find the pdf of $Z = \max\{X_{(n)}, Y_{(m)}\}$.
   (c) Assume $\theta_1 = \theta_2 = \theta$. Define $T = \frac{X_{(n)}^n Y_{(m)}^m}{Z^{n+m}}$. Find $P(T \le t)$ for $0 < t < 1$ (which is the support for T). What is the distribution of T?

3. A common gift-exchanging method at parties is to require each of the n persons attending to bring a wrapped present. When it is time to open presents, each person is randomly given one of the presents. Let $X_i = 1$ if the i^{th} person gets his or her own present to open, and 0 otherwise. Let $S_n = \sum_{i=1}^{n} X_i$ be the total number of people who are randomly given the present they originally brought to the party. Show that:
   (a) $\mathbb{E}(X_i^2) = \frac{1}{n}$.
   (b) $\mathbb{E}(X_i X_j) = \frac{1}{n(n-1)}$ for $i \ne j$.
   (c) $\mathbb{E}(S_n^2) = 2$.
   (d) $\mathbb{V}(S_n) = 1$.

4. Suppose that X and Y are two independent Exponential(β) random variables. Let $T = \min(X, Y)$ and $U = \max(X, Y) - \min(X, Y)$.
   (a) Find the pdf of T.
   (b) Find the joint pdf of (T, U), $f_{T,U}(t, u)$. Are they independent? Explain why or why not.
   (c) If you found them to be independent, does this result hold for X and Y from any continuous population? If you found them to be not independent, is there a continuous population such that T and U are independent?

5. Let $X_1, \dots, X_n$ be a random sample from a population with pdf $f(x|\mu) = e^{-(x-\mu)} I(x > \mu);\ \mu \in \mathbb{R}$.
   (a) Find the MLE for μ.
   (b) Show that $Q = n(X_{(1)} - \mu)$ is a pivotal quantity for μ, and give its distribution.
   (c) Based on (b), find a general form for a $(1-\alpha) \times 100\%$ confidence interval for μ.
   (d) Find the $(1-\alpha) \times 100\%$ confidence interval that is of the shortest length.

6. Let $X_1, \dots, X_n$ be a Uniform$(0, \frac{1}{\theta})$ random sample. Suppose that θ has a Beta$(\alpha, \beta = n+1)$ prior distribution, where α, n are known.
   (a) Find the posterior pdf for $\theta, \pi(\theta|\boldsymbol{x}, \alpha, n)$, where $\boldsymbol{x} = (x_1, \dots, x_n)$ is the realization of $\boldsymbol{X} = (X_1, \dots, X_n)$.
   (b) Find the Bayes estimator for θ under squared error loss, $\mathbb{E}(\theta|\boldsymbol{x})$.
   (c) Find the Bayes estimator for θ under absolute error loss, which is the median of $\pi(\theta|\boldsymbol{x}, \alpha, n)$.
   (d) Give the endpoints of a $(1-\gamma) \times 100\%$ credible interval for $\theta, (\theta_L, \theta_U)$.