Probability and Statistics Preliminary Examination: August 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 and 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.
- $\xrightarrow{p}$, $\xrightarrow{as}$, and $\xrightarrow{d}$ 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. A coin having probability p of coming up heads is continually flipped. The flipping stops when one or more heads (H) and one or more tails (T) is observed. Let X denote the total number of flips.
   (a) What is the probability that the last flip lands heads?
   (b) Conditioning on the first flip being heads, what is the distribution of $X-1$?
   (c) Compute $\mathbb{E}[X]$.

2. Let $Z_1, \dots, Z_n$ be iid Bernoulli with $P(Z_i = 1) = p$ and $P(Z_i = 0) = q = 1 - p$. Define the random variables X and Y to be, respectively, the number of "points" and the number of "lines" in the sequence $Z_1 \cdots Z_n$ arranged in a row.

   A "point" is defined to be an isolated 1. That is:
   - It is a single 1 with 0 on both sides of it (e.g. …010…), or
   - It is a single 1 at the front of the sequence, followed by a 0 (e.g. 10…), or
   - It is a single 1 at the end of the sequence, preceded by a 0 (e.g. …01).

   A "line" is defined similarly, except it is a group (string) of 2 or more 1s, with the additional clarification that a string of n 1s counts as a single line.

   For example, if $n = 5$:

sequence	value of (X, Y) pair
01010	(2,0)
11011	(0,2)
10111	(1,1)
11000	(0,1)
11111	(0,1)

   The purpose of this problem is to find the joint distribution of (X, Y), and ultimately that of $W = X - Y$. (These results have important applications in extreme value theory in the context of random fields.)

   Important: Let $n = 4$ in all of the following:
   (a) Is it possible to have $(X, Y) = (2,1)$; that is, $P(X = 2, Y = 1) > 0$? Explain.
   (b) Show that $P(X = 0, Y = 1) = p^4 + 2p^3q + 3p^2q^2$.
   (c) Give the joint distribution of (X, Y). (Display in a table.)
   (d) Verify that $\mathbb{E}[XY] = 2p^3q$.
   (e) Give the distribution of W.

3. Suppose that $X \sim \text{U}(-\theta, \theta)$ with pdf $f(x|\theta) = \frac{1}{2\theta} I(-\theta < x < \theta), \theta > 0$.
   (a) Show that $T = |X|$ is a complete and sufficient statistic.
   (b) Show that $T = |X|$ and $Z = \text{sign}(X)$ are independent. The sign function is defined as:

$$\text{sign}(x) = \begin{cases} -1; x < 0 \\ 0; x = 0 \\ 1; x > 0 \end{cases}$$

4. Let $X_1, \dots, X_n$ be a random sample with marginal pdf

$$f(x|\lambda,\gamma) = \frac{1}{\lambda}\exp(-(x-\gamma)/\lambda)\, I(x>\gamma);\ \lambda>0, \gamma\in\mathbb{R}.$$

This is the *shifted exponential* distribution. (Fact: $X_i - \gamma \sim \text{Exp}(\lambda)$.)

(a) Find the minimal sufficient statistic for $\boldsymbol{\theta} = (\lambda, \gamma)$.

(b) Assume $\lambda = \lambda_0$ is known. Find the MLE for γ and obtain its pdf. Is it complete? Prove or disprove.

(c) Assume $\gamma = \gamma_0$ is known. Find the MLE for λ and obtain its pdf. Is it complete? Prove or disprove.

(d) Find the (joint) MLE $\widehat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\gamma})$ for $\boldsymbol{\theta}$.

(e) Prove that $\hat{\lambda}$ and $\hat{\gamma}$ in (d) are independent. (We advise you to return to this problem later as time permits.)

5. Let $X_1, \dots, X_n$ be a Uniform$\left(0, \frac{1}{\theta}\right)$ random sample. Suppose that θ has a Gamma(α, β) prior distribution, where α, β are fixed hyperparameters.

Note: The <u>lower incomplete gamma function</u> is defined as

$$\gamma(s,z) = \int_0^z t^{s-1} e^{-t}\, dt,$$

and has the following properties

$$\lim_{z\to\infty} \gamma(s,z) = \Gamma(s),$$
$$\gamma(s+1,z) = s\gamma(s,z) - z^s e^{-z}.$$

(a) Suppose that $\boldsymbol{x} = (x_1, \dots, x_n)$ is a set of Uniform$\left(0, \frac{1}{\theta}\right)$ realizations. Find the full posterior distribution of $\theta|\boldsymbol{x}$. Be sure to completely specify all normalizing constants. (Hint: Make a change of variables $y = \theta/\beta$.)

(b) Find the Bayes Estimator under squared-error loss, $\mathbb{E}[\theta|\boldsymbol{x}]$.

6. Let $X_1, \dots, X_n$ be a random sample from pdf $f(x|\theta) = \frac{2}{\theta} x e^{-x^2/\theta} I(x>0);\ \theta>0$. Consider testing $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$.

(a) Derive the rejection region for the UMP level-α test.

(b) Derive the rejection region for an <u>asymptotic</u> level-α LRT.

(c) Compare the rejection regions from (a) and (b) and comment.