Probability and Statistics Preliminary Examination: August 2019

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 (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); ind (independently distributed).
  - MSE (mean squared error), 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).
- 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.

  Poisson(λ)**:** $\mathbb{E}(X) = \lambda$, $\mathbb{V}(X) = \lambda$, and pmf and mgf given respectively by

  $$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} I(x \in \{0, 1, \cdots\}), \qquad M(t) = \exp\{\lambda(e^t - 1)\}$$

  Geometric(p)**:** $\mathbb{E}(X) = 1/p$, $\mathbb{V}(X) = (1-p)/p^2$, and pmf and mgf given respectively by

  $$f(x) = p(1-p)^{x-1} I(x \in \{1, 2, \cdots\}), \qquad M(t) = \frac{pe^t}{1-(1-p)e^t}, \quad t < -\log(1-p)$$

  Negative-Binomial(r, p)**:** $\mathbb{E}(X) = r(1-p)/p$, $\mathbb{V}(X) = r(1-p)/p^2$, and pmf and mgf, respectively:

  $$f(x) = \binom{r+x-1}{x} p^r (1-p)^x I(x \in \{0, 1, \cdots\}), \qquad M(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r, \quad t < -\log(1-p)$$

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

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

  Multinomial($n, \pi_1, \pi_2, ..., \pi_k$)**:** $\mathbb{E}(X_i) = n\pi_i$, $\mathbb{V}(X_i) = n\pi_i(1-\pi_i)$, and pdf is given by

  $$f(x_1, x_2, ..., x_k) = \frac{n!}{\prod_{j=1}^k x_j!} \prod_{j=1}^k \pi_k^{x_k}, \quad \sum_{j=1}^k \pi_j = 1; \quad \sum_{j=1}^k x_j = n$$

Dirichlet$(\alpha_1, \alpha_2, ..., \alpha_k)$**:** $\mathbb{E}(X_i) = \alpha_1 / \sum_{i=1}^k \alpha_i$, $\mathbb{V}(X_i) = \frac{\tilde{\alpha}_i(1-\tilde{\alpha}_i)}{\tilde{\alpha}+1}$, where $\tilde{\alpha} = \sum_{i=1}^k \alpha_i$ and $\tilde{\alpha}_i = \alpha_i/\tilde{\alpha}$ and the pdf up to a normalizing constant is given by

$$f(x_1, x_2, ..., x_k) \propto \prod_{j=1}^k x_j^{\alpha_j - 1}, \quad \sum_{j=1}^k x_j = 1$$

Order Statistics: Let $X_{(1)} \leq \cdots \leq X_{(n)}$ 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:

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

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

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

1. Let $X_1, X_2 \overset{iid}{\sim} N(0,1)$.
   (a) Find the pdf of $|X_1 - X_2|$.
   (b) Using the result from part (a), or in general, show that $P(|X_1 - X_2| > t) \leq 2P(|X_1| > t/2)$

2. Let $Y_1, Y_2 \overset{iid}{\sim}$ Uniform(0,1)
   (a) Find the pdf for $V = Y_1 - Y_2$.
   (b) Find the pdf of $|V|$.
   (c) Find $E(Y_1/Y_2)$ with $Y_2 > 0$.

3. Let $X_1, X_2, X_3 \overset{iid}{\sim}$ Exponential(1).
   (a) Find the pdf of $Y_1 = \frac{X_1}{X_1+X_2}, \quad X_1 + X_2 > 0$
   (b) Now define $Y_2 = \frac{X_1+X_2}{X_1+X_2+X_3}, \quad X_1 + X_2 + X_3 > 0$ and $Y_3 = X_1 + X_2 + X_3$. Are Y_1, Y_2, Y_3 mutually independent? Why why not?

4. Let $X_1, X_2, ..., X_n$ be n iid samples from Exponential distribution with mean θ.
   (a) Find the UMVUE of θ.
   (b) Define $g(x) = X_i / \sum_{j=1}^{n} X_j$. Find $\mathbb{E}(g(x))$ and $\mathbb{V}(g(x))$.

5. Let $X_1, X_2, ..., X_n \overset{iid}{\sim} Poisson(\lambda), \lambda > 0$. Define

$$Y_i = \begin{cases} 0 & \text{if } X_i = 0, \\ 1 & \text{if } X_i > 0 \end{cases}$$

   (a) Find the MLE of λ based on Y.
   (b) Does the MLE derived in part (a) always exist for finite n? If yes, prove it. If not find the probability that the MLE does not exist for finite n.

6. Consider a discrete random variable X with pmf $P(X = j) = \pi_j, \quad j = 1,2,3, \quad \sum_{j=1}^{3} \pi_j = 1$. A random sample of size n is drawn from this population. Let n_j denote the frequency of j in the sample. Observe that $\sum_{j=1}^{3} n_j = n$. We wish to test $H_0 : \pi_1 = \pi_2 = \pi_3$ Vs. $H_a : H_0$ is false.
   (a) Derive an approximate large sample test for the above pair of hypotheses.
   (b) Use the test statistic obtained in part (a) to derive a $100(1-\alpha)\%$ Confidence Region for (π_1, π_2, π_3).
   (c) Derive an approximate large sample confidence interval for each π_i, separately. Will this CI be equivalent to the confidence region obtained in part (b)? Why, why not?
   (d) Now suppose you impose a Dirichlet$(\alpha_1, \alpha_2, \alpha_3)$ prior on (π_1, π_2, π_3). How can you obtain the Bayesian credible interval for each π_i, separately?