Probability and Statistics Preliminary Examination: May 2015

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).
  - 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

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

  Order Statistics: Let $X_{(1)} \leq \cdots \leq X_{(n)}$ 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. Recall that there are 52 cards in a standard deck, with 4 cards for each denomination. Suppose that you flip over the top 4 cards. This constitutes the first draw from the deck. Discard those cards that are not a 2 and redraw new cards from the remaining deck so that you again have 4 cards flipped over. Continue doing this until all 4 cards are a 2. Let X denote the number of draws made and Y_j denote the number of cards drawn on draw j.

   (a) Calculate $P(X = 3)$.

   (b) Derive the pmf of Y_2.

   (c) Derive the pmf of $Y_2|X \geq 3$.

2. Let X be a random variable from a *contaminated normal* distribution. That is, let $B \sim$Bernoulli(p). Then $X|B = 0 \sim N(\mu, \tau^2)$ and $X|B = 1 \sim N(\mu, \sigma^2)$.

   (a) Derive the pdf for X

   (b) Calculate $\mathbb{E}(X)$ and $\mathbb{V}(X)$.

   (c) Derive an expression for $P(B = 1|X = x)$.

   (d) Calculate $\text{Cov}(X, B)$. Are X and B independent? Justify your answer.

3. Let $X_1, \ldots, X_n$ $(n > 4)$ be independent random variables such that $X_i \sim N(i, i)$ for $i = 1, \ldots, n$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean.

   (a) Calculate $\mathbb{E}(\overline{X})$ and $\mathbb{V}(\overline{X})$. Describe the behavior of both of these quantities as $n \to \infty$.

   (b) Determine the distribution of
   $$\frac{2(X_1 + X_2 - X_3)}{\sqrt{6}(X_4 - 4)}$$

   (c) Calculate the probability that X_n is not the largest observation in the sample. This probability should be expressed in terms of $\Phi(\cdot)$, the cdf of the standard normal distribution.

4. Let $(X_1, Y_1) \dots, (X_n, Y_n)$ be iid random vectors from the pdf:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

   (a) Find a minimal sufficient statistic for ρ.
   (b) Show that the MLE may be found as a solution to the following equation:

$$0 = \rho^3 - \rho^2 \frac{\sum_{i=1}^n x_i y_i}{n} - \rho \frac{n - \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2}{n} - \frac{\sum_{i=1}^n x_i y_i}{n}$$

   (c) Is $(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i)$ a sufficient statistic for ρ? Why or why not?
   (d) Are $\overline{X}$ and $\overline{Y}$ (separately) ancillary statistics for ρ? Why or why not?

5. Let $X_1, \dots, X_n$ $(n > 2)$ be a random sample from a distribution with $\mathbb{E}(X_1) = \mu < \infty$ and $\mathbb{V}(X_1) = \sigma^2 < \infty$. Define the following estimators for μ.

$$\overline{X}_t = \frac{\sum_{i=2}^{n-1} X_{(i)}}{n-2}, \qquad \tilde{\mu}_n = \frac{1}{a}X_1 + \frac{c}{n-2}\sum_{i=2}^{n-1} X_i + \frac{1}{a}X_n,$$

where $a > 0$ and c are constants.

   (a) Calculate $\mathbb{E}(\overline{X}_t)$.
   (b) Under what conditions is $\overline{X}_t$ an unbiased estimator of μ?
   (c) Determine the value of c for which $\tilde{\mu}_n$ is an unbiased estimator of μ.
   (d) If $a = 2$, determine whether $\tilde{\mu}_n$ is a consistent estimator for μ.
   (e) Find the value of a that minimizes $\mathbb{V}(\tilde{\mu}_n)$.

6. Let X_1, X_2, and X_3 be random variables from the following joint pmf:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3},$$

with $0 \le x_i \le n$ for $i = 1, 2, 3$, $\sum_{i=1}^3 x_i = n$, and $0 \le p_i \le 1$ for $i = 1, 3$, and $\sum_{i=1}^3 p_i = 1$.

   (a) Find the MLEs for p_1 and p_2.
   (b) Give the test statistic and rejection region for an asymptotic level α LRT of $H_0 : p_1 = p_2$ vs. $H_1 : p_1 \neq p_2$.
   (c) Conditional on $X_3 = x_3$, construct an asymptotic level $100(1-\alpha)\%$ confidence interval for $\frac{p_1}{p_1+p_2}$.