Probability and Statistics Preliminary Examination: May 2014

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 $X_{(1)} \le \cdots \le X_{(n)}$ 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:

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

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

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

1. Recall that there are 4 aces and 48 "non-aces" in a standard 52-card deck, and suppose that each non-ace card is assigned a unique number from 1 to 48. An experiment consists of shuffling the deck and stacking it on the table. Let N denote the number of non-ace cards that are stacked above all 4 aces. Also, for $i = 1, \ldots, 48$, let $X_i = 1$ if non-ace card i is stacked above all 4 aces, and $X_i = 0$ otherwise. Define $p = P(X_i = 1)$ and $q = P(X_i = 1, X_j = 1)$, for $i \neq j$.

   (a) Reason that the X_i are identically distributed, and find their common distribution. Are the X_i independent?

   (b) Find $\mathbb{E}(N)$ in terms of p.

   (c) Find $\mathbb{V}(N)$ in terms of p and q.

   (d) Determine the values of p and q.

2. Consider the following hierarchical specification for the joint distribution of (X, Λ):

$$X \mid \Lambda \sim \text{Poisson}(\Lambda), \qquad \text{and} \qquad \Lambda \sim \text{Gamma}(\alpha, \beta).$$

   (a) Show that the marginal pmf of X is given by

$$f_X(x|\alpha, \beta) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\beta^\alpha x!} \left(\frac{\beta}{\beta + 1}\right)^{x+\alpha} I_{\{0,1,2,\ldots\}}(x).$$

   (b) Compute the mgf of X.

   (c) Compute $\mathbb{E}(X)$ and $\mathbb{V}(X)$.

   (d) If $X_1|\Lambda$ and $X_2|\Lambda$ are iid from the conditional distribution of $X|\Lambda$, show that the correlation coefficient between X_1 and X_2 is $\rho = \beta/(1 + \beta)$.

3. Let $X_1, \ldots, X_n$ be a random sample from $X \sim \text{N}(0, \sigma^2)$, whose cdf is $F(x) = P(X \leq x)$. Define $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x)$ to be the *empirical cdf*, which is an estimator of $F(x)$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ be the usual sample mean and sample variance.

   (a) What is the *exact* distribution of $n\overline{X}^2/S^2$?

   (b) Just for this part of the Problem, if $X_1, \ldots, X_n$ are a random sample from X, not necessarily normal, where $\mathbb{E}(X) = 0$, $\mathbb{V}(X) = \sigma^2$, and $\mathbb{E}(X^4) < \infty$, what is the limiting (asymptotic) distribution of $n\overline{X}^2/S^2$?

   (c) Find the distribution of $F_n(0)$ when $n = 4$.

   (d) Let $\xi_p = F^{-1}(p)$ be the pth quantile of X, i.e., $P(X \leq \xi_p) = p$. Determine the limiting (asymptotic) distribution of

$$\sqrt{n}\left(F_n(\xi_p) - p\right).$$

4. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution depending on the unknown parameters $\nu > 0$ and $\theta > 0$, with pdf

$$f(x|\nu, \theta) = \frac{\nu}{\theta^\nu} x^{\nu-1} I(0 < x < \theta).$$

   (a) Show that the two-dimensional statistic $(S, T) = (\prod_{i=1}^n X_i, X_{(n)})$ is *minimal sufficient* for (ν, θ).
   (b) Find a MOME of (ν, θ).
   (c) Find the MLE of (ν, θ).
   (d) Find the LRT statistic λ for testing the hypothesis $H_0 : \theta = 1$ vs. $H_1 : \theta \neq 1$. Is it possible to appeal to the asymptotic distribution of λ in order to find the rejection region in this case? Justify your answer.

5. Let $X_1, \ldots, X_n$ be a random sample from a *Rayleigh* distribution depending on the unknown parameter $\theta > 0$, with pdf

$$f(x|\theta) = \frac{x}{\theta} \exp\left\{-\frac{x^2}{2\theta}\right\} I(x > 0).$$

Note that $\mathbb{E}(X_1) = \sqrt{\pi\theta/2}$, $\mathbb{E}(X_1^2) = 2\theta$, and define the statistic: $T = \sum_{i=1}^n X_i^2$.

   (a) Find the MLE of θ, and show that it is unbiased.
   (b) Compute the Cramer-Rao Lower Bound for the variance of unbiased estimators of θ. Does the result allow you to make a statement about the optimality of the MLE (i.e., is the MLE UMVUE)?
   (c) Compute $(1 - \alpha)100\%$ confidence intervals for θ based on both the asymptotic distribution of the MLE, and on the asymptotic distribution of the *score* statistic. Compare the lengths of the intervals; which interval is shorter? (Recall that the MLE-based interval uses the *observed* Information, while the score-based interval uses the *expected* Information.)
   (d) Using the (improper) prior

$$\pi(\theta) \propto \theta^n \exp\left\{\frac{t}{2\theta} - \frac{\theta}{t}\right\} I(\theta > 0),$$

   where t is the observed value of T, find a $(1 - \alpha)100\%$ *credible interval* for θ of the shortest possible length.

6. Let $X_1, \ldots, X_n$ be a random sample from $X \sim$ Geometric(θ), where $0 \leq \theta \leq 1$ is unknown. Recall that if $Y = X - 1$, then $Y \sim$ Negative-Binomial$(1, \theta)$, and note the definition of these distributions given on page 1.

   (a) For $n = 1$, find the level $\alpha = 1/8$ UMP test of $H_0 : \theta = 1/2$ vs. $H_1 : \theta = 1/5$, and compute the power of the test when $\theta = 1/5$.
   (b) For $n \geq 1$, give the *rejection region* and derive an expression for the *power function* $\beta(\theta)$ of the UMP size α test of $H_0 : \theta \leq 1/2$ vs. $H_1 : \theta > 1/2$.
   (c) If $n = 2$, compute the rejection region and power function explicitly for the UMP test of (b) when $\alpha = 11/16$.
   (d) Compute the *p-value* for the situation in (c) if we observe $\sum_{i=1}^n x_i = 3$. Should H_0 be rejected?