Statistics Preliminary Examination: August 2011

Instructions:

- Work all 6 problems. Begin each problem on a new sheet of paper, and do not use both sides of the paper (continue on a new sheet of paper if the solution will not fit on a single sheet). Neither calculators nor electronic devices of any kind are allowed. Clearly state any theorem or fact that you use. The 23 parts are approximately equally weighted.

- Abbreviations/Acronyms.
  - pmf (probability mass function); pdf (probability density function); cdf (cumulative distribution function); mgf (moment generating function);
  - MOME (method of moments estimator); MLE (maximum likelihood estimator); UMVUE (uniform minimum variance unbiased estimator); UMP (uniformly most powerful test); LRT (likelihood ratio test); MLR (monotone likelihood ratio).

- Notation.
  - $I(x \in A)$ or $I_A(x)$: indicator function for set $A$; takes on value 1 if $x \in A$ and 0 otherwise.
  - $E(X)$: expectation of random variable $X$.
  - $V(X)$: variance of random variable $X$.
  - $X \sim N(a, b)$: $X$ has a normal distribution with mean $a$ and variance $b$.

- Common distributions and other results.

**Poisson($\lambda$):** $E(X) = \lambda$, $V(X) = \lambda$, and pmf

\[
f(x) = \frac{e^{-\lambda} \lambda^x}{x!} I(x \in \{0, 1, \cdots\})
\]

**Exponential($\lambda$):** $E(X) = \lambda$, $V(X) = \lambda^2$, and pdf

\[
f(x) = \frac{1}{\lambda} e^{-x/\lambda} I(x > 0)
\]

**Beta($\alpha, \beta$):** $E(X) = \alpha/(\alpha + \beta)$, $V(X) = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$, and pdf

\[
f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1)
\]

**Gamma($\alpha, \beta$):** $E(X) = \alpha\beta$, $V(X) = \alpha\beta^2$, and pdf

\[
f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta} I(x > 0)
\]

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 $X_{(j)}$ and of $X_{(1)} \leq \cdots \leq X_{(n)}$ is given by:

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

\[
f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n) I(-\infty < x_1 \leq \cdots \leq x_n < \infty)
\]
1. Suppose that each of $N$ men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat. Let $H_i$ be the event that the $i$th man selects his own hat.

(a) Show that $P(H_{i_1} \cap H_{i_2} \cap \cdots \cap H_{i_n}) = (N - n)!/N!$. (Note that this is the probability that each of the $n$ men, $i_1, i_2, \ldots, i_n$, selects his own hat.)

(b) Compute $P(\bigcup_{i=1}^N H_i)$.

(c) Hence, or otherwise, show that for large $N$ the probability that none of the men selects his own hat is approximately equal to $e^{-1}$. [Hint: recall that $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$.]

2. A penny and a dime are tossed. Let $X$ denote the total number of heads up. Then the penny is tossed again. Let $Y$ denote the total number of heads up on the dime (from the first toss) plus the penny from the second toss.

(a) Find the joint pmf of $X$ and $Y$, and hence compute the marginal pmf’s of $X$ and $Y$.

(b) Find the conditional distribution of $Y$ given $X = 1$.

(c) Show that $X$ and $Y$ are not independent. Compute the correlation between $X$ and $Y$.

3. Assume $X_1, \ldots, X_n$ is a random sample from a Poisson($\lambda$) distribution, and let $\overline{X}$ and $S_n^2$ denote the usual sample mean and variance. In addition, suppose that $W, Y, Z$ are independent random variables, with both $W$ and $Y \sim N(\mu, \sigma^2)$, but $Z \sim N(0, \sigma^2)$.

(a) Show that both $\sqrt{n}(\overline{X} - \lambda)/\sqrt{X}$ and $\sqrt{n}(\overline{X} - \lambda)/S_n$ have a limiting standard normal distribution.

(b) Find the limiting distribution of $n(\overline{X} - \lambda)^2$.

(c) Find the limiting distribution of $\sqrt{n}(\overline{X}^2 - \lambda^2)$.

(d) Find the (exact) distribution of $\sqrt{2}(W + Y - 2\mu)/\sqrt{2\sigma^2 + (W - Y)^2}$.
4. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution over the interval $[\theta - 1/2, \theta + 1/2]$, where $\theta \in \mathbb{R}$ is unknown. Denote the order statistics by $X_{(1)} \leq \cdots \leq X_{(n)}$.

(a) Recall that if $Z_{(r)}$ is the $r$-th order statistic, $1 \leq r \leq n$, in a random sample of size $n$ from a uniform on $[0, 1]$, then $\mathbb{E}Z_{(r)} = r/(n + 1)$. Using this fact, show that if $Y_{(r)}$ is the $r$-th order statistic in a random sample of size $n$ from a uniform on $[a, b]$, then $\mathbb{E}Y_{(r)} = a + r(b - a)/(n + 1)$. Produce an expression for $\mathbb{E}X_{(r)}$.

(b) Can a sufficient statistic of dimension 1 for $\theta$ be found? If so find it; if not find one of the smallest possible dimension.

(c) Is the sufficient statistic found in (b) complete? Justify.

(d) Find the Method of Moments estimator of $\theta$. Is it unbiased?

(e) Is the MLE of $\theta$ unique? If so find its bias; if not produce an unbiased MLE.

5. Suppose that we have two independent random samples: $X_1, \ldots, X_n$ are Exponential($\theta$), and $Y_1, \ldots, Y_m$ are Exponential($\mu$).

(a) Show that the LRT of $H_0 : \theta = \mu$ vs. $H_1 : \theta \neq \mu$ can be based on the statistic
\[
T = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}.
\]

(b) When $H_0$ is true, find the distribution of $T$ and show that it is independent of the distribution of $S = \sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j$.

(c) Construct the size $\alpha$ LRT in (a) by using the large sample distribution of $-2 \log \lambda$, where $\lambda$ is the LRT statistic.

(d) Based only on the sample $X_1, \ldots, X_n$, describe an exact (non-asymptotic) procedure to determine a $(1 - \alpha)$ confidence interval for $\theta$.

6. Suppose $X_1, \ldots, X_n$ is a random sample from the pdf
\[
f(x|\theta) = (1/\theta)x^{(1-\theta)/\theta}I(0 < x < 1),
\]
where $\theta > 0$ is unknown.

(a) Show that this family of distributions has MLR in some sufficient statistic $T$.

(b) Derive the size $\alpha$ UMP test of $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$. Give reasons for any claims that you make.

(c) Derive an expression for the power function $\beta(\theta)$ of the above UMP test, that can be evaluated by using $\chi^2$ tables.

(d) Derive a $(1 - \alpha)$ confidence interval for $\theta$ obtained by inverting the asymptotic distribution of the score statistic, $Z_S = S(\theta_0)/\sqrt{I_n(\theta_0)}$, where
\[
S(\theta) = \frac{\partial \log L(\theta|x)}{\partial \theta}, \quad \text{and} \quad I_n(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \log L(\theta|x)}{\partial \theta^2} \right],
\]
is the expected (Fisher) Information Number.