Probability and Statistics Preliminary Examination: August 2013

Instructions:

- Work all 6 Problems. Neither calculators nor electronic devices of any kind are allowed. Clearly state every 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);
  - MOME (method of moments estimator); MLE (maximum likelihood estimator); UMVUE (uniform minimum variance unbiased estimator); UMP (uniformly most powerful); LRT (likelihood ratio test); MLR (monotone likelihood ratio).
- 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.

  Hypergeometric(N, n, r)**:** $\mathbb{E}(X) = np$, $\mathbb{V}(X) = np(1-p)c$, $p = nr/N$, $c = (N-n)/(N-1)$, and pmf
  $$f(x) = \binom{r}{x}\binom{N-r}{n-x}\bigg/\binom{N}{n}, \qquad \text{for } \max\{0, n-(N-r)\} \le x \le \min\{r, n\}$$

  Exponential(λ)**:** $\mathbb{E}(X) = \lambda$, $\mathbb{V}(X) = \lambda^2$, and pdf
  $$f(x) = \frac{1}{\lambda} e^{-x/\lambda} I(x > 0)$$

  Beta(α, β)**:** $\mathbb{E}(X) = \alpha/(\alpha+\beta)$, $\mathbb{V}(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, and pdf
  $$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1)$$

  Gamma(α, β)**:** $\mathbb{E}(X) = \alpha\beta$, $\mathbb{V}(X) = \alpha\beta^2$, and pdf
  $$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I(x > 0)$$

  Order Statistics: Let $X_{(1)} \le \cdots \le X_{(n)}$ 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. Consider a sequence of n tosses of a fair coin. Let X denote the number of heads, and Y denote the number of *isolated* heads, that come up. (A head is an "isolated" head if it is immediately preceded and followed by a tail, except in: position 1 where a head is only followed by a tail, and position n where the head is only preceded by a tail.) Additionally, let $X_i = 1$ if the ith coin toss results in heads, and $X_i = 0$ otherwise, $i = 1, \ldots, n$. Similarly, $Y_j = 1$ if the jth coin toss results in an isolated head, and $Y_j = 0$ otherwise, $j = 1, \ldots, n$.

   (a) Argue that X_i and Y_j are both Bernoulli random variables, and determine the success probability for each, i.e. find $P(X_i = 1)$ and $P(Y_j = 1)$, for each $i = 1, \ldots, n$ and each $j = 1, \ldots, n$.

   (b) Find $\mathbb{E}(X)$ and $\mathbb{E}(Y)$.

   (c) Find the distribution of $Y|X = 2$.

   (d) Find $\text{Cov}(X, Y)$, the covariance between X and Y, for the case $n = 5$.

2. A box contains N balls: N_1 white, N_2 black, and N_3 red ($N = N_1 + N_2 + N_3$). A random sample of n balls is selected from the box (without replacement). Let Y_1, Y_2, and Y_3 denote the number of white, black, and red balls, respectively, observed in the sample. Define $c = (N - n)/(N - 1)$, and $p_i = N_i/N$, for $i = 1, 2, 3$. Assume $3 \leq n \leq \min\{N_1, N_2, N_3\}$.

   (a) Deduce the marginal distributions of Y_1, Y_2, and Y_3, and compute their respective means and variances.

   (b) Show that $\text{Cov}(Y_1, Y_2) = -np_1p_2c$.

   (c) Hence, or otherwise, compute the correlation between Y_1 and Y_2.

3. Let the random vector $\mathbf{X} = (X_1, \ldots, X_4)$ have a *multivariate normal* distribution with mean and variance-covariance matrix as follows:

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim \text{Normal} \left(\begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & \rho\sigma^2 & \sigma^2 \end{bmatrix} \right).$$

   In particular, this means $X_i \sim \text{N}(\mu, \sigma^2)$, $i = 1, \ldots, 4$, and $\text{Cov}(X_1, X_2) = \rho\sigma^2$. Let $\overline{X} = \sum_{i=1}^{n} X_i/4$ and $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/3$. Recall that any set of linear combinations of the elements of $\mathbf{X}$ are also (jointly) multivariate normally distributed.

   (a) Find the exact distribution of $\overline{X}$.

   (b) Find $\mathbb{E}(S^2)$.

   (c) Find the exact distribution of:

$$W = \frac{3(1 + \rho)(X_1 - X_2)^2}{(1 - \rho)[(X_1 + X_2 - 2\mu)^2 + (X_3 + X_4 - 2\mu)^2] + (1 + \rho)(X_3 - X_4)^2}.$$

   (d) Find the exact distribution of:

$$Y = \frac{(X_1 + X_2 + X_3 + X_4 - 4\mu)}{\sqrt{(X_1 - X_2)^2 + (X_3 - X_4)^2}} \sqrt{\frac{1 - \rho}{1 + \rho}}.$$

4. Let $X_1, \ldots, X_n$ be a random sample from the *inverse Gaussian distribution*, X, with pdf

$$f(x|\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\} I(x>0),$$

where $\lambda > 0$ and $\mu > 0$ are unknown parameters. Define the statistics

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \text{and} \qquad T = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i} - n/\overline{X}}.$$

You may use the result that $n\lambda/T$ is known to have a χ^2_{n-1} distribution.

(a) By showing that $\mathbb{E}(X+\mu) = 2\mu$, or otherwise, establish that $\mathbb{E}X = \mu$.

(b) Find the MLEs of μ and λ^{-1}. Are they unbiased?

(c) Find the UMVUEs of μ and λ^{-1}. Compare them to the corresponding MLEs.

(d) By using either exact (finite n) or asymptotic (large n) methods, construct $(1-\alpha)$ confidence intervals for each of the parameters λ and μ.

5. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on the interval $(\theta, 1)$, with $0 < \theta < 1$ unknown. Let $T = X_{(1)}$ be the minimum order statistic.

(a) Show that the distribution of T is a location-scale shifted Beta$(1, n)$, and identify the location and scale parameters.

(b) Find a *complete* and *sufficient* statistic for θ.

(c) Find a MOME, the MLE, and the UMVUE of θ.

(d) For a Uniform$(0,1)$ prior on θ, find a $(1-\alpha)$ *Bayesian credible interval* for θ that has the shortest possible length.

(e) Find the power function of the UMP size α test of $H_0 : \theta \leq 1/2$ vs. $H_1 : \theta > 1/2$.

6. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the bivariate discrete distribution (X, Y) with pmf

$$f(x,y|\boldsymbol{\theta}) = \left(\frac{\theta_2}{2}\right)^{x+y-2xy} \theta_1^{xy}(1-\theta_1-\theta_2)^{1+xy-x-y} I_{\{(1,1),(1,0),(0,1),(0,0)\}}(x,y),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $0 \leq \theta_1$, $0 \leq \theta_2$, with $\theta_1 + \theta_2 \leq 1$, are unknown parameters. Define the statistics $s = \sum_{i=1}^{n}(x_i + y_i)$, and $t = \sum_{i=1}^{n} x_i y_i$. Note that the pmf can be expressed in tabular form as:

(x,y)	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$f(x,y\|\boldsymbol{\theta})$	θ_1	$\theta_2/2$	$\theta_2/2$	$1-\theta_1-\theta_2$

(a) Find the marginal distributions of X and Y.

(b) Show that the MLE of $\tau(\boldsymbol{\theta}) = \theta_1 - (\theta_1 + \theta_2/2)^2$ is:

$$\hat{\tau} = \frac{t}{n} - \left(\frac{s}{2n}\right)^2.$$

(c) Construct a level α test (either asymptotic or exact) of

$$H_0 : X \text{ and } Y \text{ are independent}, \quad \text{vs.} \quad H_1 : X \text{ and } Y \text{ are not independent}.$$