# Statistics Ph.D. Comprehensive
# Wednesday, August 14, 2013

Instructions: The exam has 7 problems. All of the problems will be graded. Write your ID number on your answer sheets. Do not put your name on any of the sheets. Be clear, concise, and complete. **All solutions must be rigorously explained.**

1. Consider a positive continuous random variable $T$ with density $f(t)$ and mean 1, i.e.

$$\int_0^\infty t f(t) dt = 1. \tag{1}$$

   (a) For a fixed value $x > 0$, define a random variable $\tilde{\theta}$ by $T = x/\tilde{\theta}$. Find the density of $\tilde{\theta}$. Use equation (1) to show that

$$\int_0^\infty \frac{x^2}{\tilde{\theta}^3} f(x/\tilde{\theta}) d\tilde{\theta} = 1.$$

   (b) Let $X = \theta T$ so that $X|\theta \sim \frac{1}{\theta} f(x/\theta)$, for $x, \theta > 0$. Show that $E(X|\theta) = \theta$, so $X$ is unbiased for $\theta$. (You can use equation (1) but there are easier arguments.)

   (c) Bickel and Mallows (1998) investigated the relationship between unbiasedness and being a Bayes estimate, specifying conditions under which these properties cannot hold simultaneously. They show that if a prior distribution is improper, then a posterior mean can be unbiased. With $X|\theta \sim \frac{1}{\theta} f(x/\theta)$, for $x, \theta > 0$ and an improper prior on $\theta$ of $\pi(\theta) = \frac{1}{\theta^2}$, for $\theta > 0$, show that application of Bayes formula gives $\pi(\theta|x) = \frac{x^2}{\theta^3} f(x|\theta)$ which is a *proper* (posterior) density.

   (d) Show that $E(\theta|x) = x$, and hence the posterior mean is unbiased.

2. Let $X_1, X_2, \ldots, X_n$ be a sample from the Poisson($\lambda$) distribution truncated on the left at 0, i.e. with sample space $\{1, 2, 3, \ldots\}$.

   (a) Give the density of $X_1$.

   (b) Show that the Cramer-Rao lower bound for the variance of an unbiased estimator of $\lambda$ is

$$\frac{\lambda(1 - e^{-\lambda})^2}{n(1 - e^{-\lambda} - \lambda e^{-\lambda})}.$$

   (The algebra/calculus gets a bit complicated but focus on your methodology. Do not spend *too much* time on the mechanics.)

   (c) Find the asymptotic variance of the MLE estimator of $\lambda$.

3. Suppose that $Y_{(1)} < Y_{(2)} \ldots < Y_{(n)}$ are the order statistics from a random sample $X_1, X_2, \ldots, X_n$ of size $n$ with a continuous pdf $f(x)$ and cdf $F(x)$.

   (a) Derive both the cdf and the pdf of $Y_{(1)}$ based on $f(x)$ and $F(x)$.

(b) Let $\xi_p$ be the $p-th$ quantile of $X_1$ for some $0 < p < 1$, so that $P(X_1 \leq \xi_p) = p$. Show that $P(Y_{(1)} \leq \xi_p) \to 1$ as $n \to \infty$.

4. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson($\theta$) distribution. Notice that any single observation $X_i$ is an unbiased estimator for $\theta$ ($i = 1, \ldots, n$).

   (a) Show that $T = \sum_{i=1}^n X_i$ is a sufficient and complete statistic for $\theta$.

   (b) Show that $T \sim \text{Poisson}(n\theta)$. The mgf is a convenient method.

   (c) Show that $X_1|T \sim \text{Binomial}(T, 1/n)$. Note that $X_1$ and $T - X_1$ are independent.

   (d) According to the Rao-Blackwell theorem $\phi(T) = E[X_i|T = t]$ is the uniformly minimum variance unbiased estimator of $\theta$ and this estimator is unique. Compute $\phi(T)$ and verify that $\phi(T)$ is a well known estimator for this problem.

5. Let $X_1, X_2, \ldots$ be iid with $E(X_j) = \mu \equiv \mu_1$, $Var(X_j) = \sigma^2$, and $E(X_j^k) = \mu_k$, $k = 1, 2, 3, 4$. Find the asymptotic distribution of the sample variance

$$S_n^2 \equiv \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \bar{X}\right)^2,$$

   i.e., of $\sqrt{n}[S_n^2 - \sigma^2]$.

6. Let $X_1, X_2, \ldots$ be independent with $X_n$ taking the values $\sqrt{n}$ and $-\sqrt{n}$ each with probability 1/4 and the value 0 with probability 1/2. Find the asymptotic distribution of $\bar{X}_n$.

7. Consider, for $i = 0, 1, 2, \ldots, n$, a linear model

$$y_i = x_i'\beta + e_i, \qquad e_i \text{ iid } N(0, \sigma^2).$$

   Use observations 1 through $n$ to create a standard linear model,

$$Y = X\beta + e, \qquad e \sim N(0, \sigma^2 I).$$

   For the standard (least squares) estimates, show that if $x_0'\beta$ is estimable, then

$$\frac{y_0 - x_0'\hat{\beta}}{\sqrt{MSE[1 + x_0'(X'X)^- x_0]}} \sim t(n - r(X)).$$

   Use this result to construct a prediction interval for $y_0$. Explain the philosophical basis for the interval.