UNIVERSITY OF MASSACHUSETTS
DEPARTMENT OF MATHEMATICS AND STATISTICS
ADVANCED EXAM - STATISTICS (II)
January 18, 2017

Work all problems and show all work. Explain your answers. State the theorems used whenever possible. 70 points are required to pass.

1. Let $X_1, \ldots, X_n$ be an independent and identically distributed (i.i.d) random sample from an exponential distribution with mean θ and k-th moment $EX^k = k!\theta^k$. Define

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 = \overline{Y}_n - (\overline{X}_n)^2.$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $Y_i = X_i^2$ and $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$

   (a) (6 points) Derive the joint asymptotic distribution of $\overline{X}_n$ and $\overline{Y}_n$.

   (b) (7 points) Derive the joint asymptotic distribution of $\overline{X}_n$ and S_n^2.

   (c) (7 points) Define the coefficient of variation to be

$$CV_n = \frac{\sqrt{S_n^2}}{\overline{X}_n}.$$

   Show that $\sqrt{n}(CV_n - 1) \xrightarrow{d} Z$ where $Z \sim N(0,1)$ (i.e., $\sqrt{n}(CV_n - 1)$ converges in distribution to Z).

2. Suppose that $X_1, \ldots, X_n$ is an i.i.d random sample from a distribution with the density function $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$ and $\theta > 0$. Note that $E(X_i) = 1/\theta$ and $Var(X_i) = 1/\theta^2$.

   (a) (6 points) Show that the likelihood equation of θ has a unique solution, denoted as $\hat{\theta}_n$, and this solution maximizes the likelihood function. Also check the regularity conditions necessary for consistency of $\hat{\theta}_n$.

   (b) (7 points) Show that $\hat{\theta}_n$ is consistent and asymptotically efficient.

   Consider the prior distribution for the parameter θ as an exponential distribution $\pi(\theta) = e^{-\theta}$ where $\theta > 0$.

   (c) (6 points) Derive the Bayesian estimator (i.e., the posterior mean of θ), denoted as $\hat{\hat{\theta}}_n$.

   (d) (6 points) Derive the asymptotic distribution of $\hat{\hat{\theta}}_n$.

3. Suppose that the random variables $Y_i = \alpha + \beta x_i + \epsilon_i$ for $i = 1, \ldots, n$, where $x_1, \ldots, x_n$ are known constants and $\epsilon_1, \ldots, \epsilon_n$ are i.i.d random variables with mean 0 and variance $\sigma^2 < \infty$. The least-squares estimator of β is

$$\hat{\beta}_n = \sum_{j=1}^{n} Y_j(x_j - \overline{x}_n) \Big/ \sum_{j=1}^{n}(x_j - \overline{x}_n)^2 = \beta + \sum_{j=1}^{n} \epsilon_j(x_j - \overline{x}_n) \Big/ \sum_{j=1}^{n}(x_j - \overline{x}_n)^2$$

where $\overline{x}_n = \frac{1}{n}\sum_{j=1}^n x_j$.

(a) (6 points) Show that $\hat{\beta}_n$ is a consistent estimator of β. Under what condition on $x_1, \ldots, x_n$, is $\hat{\beta}_n$ a consistent estimator of β?

(b) (14 points) Assume that

$$\gamma_n^2 \equiv \frac{\max_{1 \le j \le n}(x_j - \overline{x}_n)^2}{\sum_{j=1}^n (x_j - \overline{x}_n)^2} \to 0 \text{ as } n \to \infty.$$

Prove that

$$\sqrt{n} s_n(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \sigma^2),$$

where $s_n^2 = \frac{1}{n}\sum_{j=1}^n (x_j - \overline{x}_n)^2$. [Hint] Use the Lindeberg-Feller Theorem (extension of the Central Limit Theorem to the independent nonidentically distributed case) by constructing a triangular array of random variables and showing that the Lindeberg condition is satisfied.

4. Suppose $X_1, \ldots, X_n$ are i.i.d random variables with the distribution function $F(x)$. Let $\hat{F}_n(x)$ denote the empirical distribution function $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n I\{X_i \le x\}$.

(a) (5 points) For every value of x, show that $\hat{F}_n(x)$ is a consistent estimator for $F(x)$.

(b) (5 points) For every value of x, find the asymptotic distribution of $\hat{F}_n(x)$.

(c) (5 points) Let $X_1, \ldots, X_n$ be an i.i.d random sample from a distribution with the following density:

$$f(x \mid \theta) = \frac{1}{\pi\left[1 + (x - \theta)^2\right]}$$

where n is odd, θ is the median and $E(X)$ is undefined. Let $\tilde{\theta}_n$ denote the sample median. Suppose we wish to estimate $g(F) = E_F(\tilde{\theta}_n) < \infty$. We use a bootstrap scheme in which we draw B random samples of size n from $\hat{F}_n(x)$, and let M_b be the sample median of the b-th sample, $b = 1, \ldots, B$. Describe (with justification) what happens to $\overline{M}_B = \frac{1}{B}\sum_{b=1}^B M_b$ when n is fixed but with $B \to \infty$.

5. Let $P_0, P_1,$ and P_2 be the space of possible probability distributions assigning to the integers $1, 2, \ldots 6$ the following probabilities:

	1	2	3	4	5	6
P_0	.03	.02	.02	.01	0	.92
P_1	.06	.05	.08	.02	.01	.78
P_2	.09	.05	.12	0	.02	.72

Consider the null hypothesis $H_0 : P = P_0$. Based on a single observation $X \in \{1, 2, \ldots 6\}$:

(a) (6 points) Is there a uniformly most powerful test against the alternatives P_1 and P_2 at level $\alpha = .01$? If so, specify the rejection region of that test.

(b) (6 points) Is there a uniformly most powerful test against the alternatives P_1 and P_2 at level $\alpha = .05$? If so, specify the rejection region of that test.

(c) (8 points) Recall that one way to construct a confidence set is to invert a hypothesis test: (i.e. allowing the confidence set to include all members of the parameter space for which the designated test would not reject.) Suppose $X = 4$ is observed. Give a 99% confidence set, and (briefly) justify your choice.