QUALIFYING EXAM IN STATISTICS, MAY 2005

Instr. A. Delaigle, L. Haff
MATH 281ABC

Student name:
Student ID (last 5 digits):

---

1. Let $X_1, X_2, \ldots, X_n$ be an iid sample with pdf $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$, with $x > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Consider the following estimator of μ :

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n T_i}{n-1} + 1,$$

where we define $T_i = \ln X_i$. It is well known that $T_i \sim N(\mu, \sigma^2)$ (do not show this).

(a) Show that the bias of $\hat{\mu}_1$ is

$$\frac{\mu}{n-1} + 1.$$

(b) Show that the variance of $\hat{\mu}_1$ is

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

(c) Deduce the mean squared error of the estimator. Is $\hat{\mu}_1$ a consistent estimator of μ? Why?

(d) Transform $\hat{\mu}_1$ to obtain an unbiased estimator $\hat{\mu}_2$ of μ.

2. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the pdf

$$f(x, \theta) = \theta.(1+x)^{-\theta-1},$$

with $0 < x < \infty$ and $\theta > 0$.

(a) Show that the maximum likelihood estimator of θ is

$$\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^n \ln(1 + X_i)}.$$

(b) Show that the Fisher information $I_n(\theta)$ is equal to n/θ^2.
(remember that $I_n(\theta)$ is minus the expectation of the second derivative of the log likelihhod).

3. Suppose $X_1, X_2, \dots, X_n$ are iid from a distribution with pdf $f(x|\theta)$ where T is a sufficient statistic. Also, assume the Bayesian framework with prior pdf's $\pi(\theta|\gamma)$ and $\psi(\gamma)$. Show that the posterior distribution depends on the data only through T.

4. Suppose that

$$X \sim f(x|\lambda) = \frac{\lambda^3}{2} x^2 e^{-\lambda x}; \quad x \geq 0 \text{ and } \lambda > 0 \ (\lambda \text{ a constant}).$$

Give conditions on $g(x)$ under which

$$\mathrm{E}_\lambda\, g'(X) = \lambda\, \mathrm{E}_\lambda\, g(X) - \mathrm{E}_\lambda \left(\frac{2g(X)}{X}\right).$$

5. Let $X_1, X_2, \dots, X_n$ be an iid sample from a distribution F_X and let $Y_1, Y_2, \dots, Y_n$ be an iid sample from a distribution F_Y, and the two samples are independent. The Wilcoxon-Mann-Whitney test statistic is defined by

$$T_X = \sum_{i=1}^{m} r(X_i),$$

where $r(X_i)$ denotes the rank of X_i in the global sample, i.e. in the sample of X's and Y's together.

Suppose we want to test

$$H_0 : F_X = F_Y$$

against

$$H_1^- : F_X < F_Y.$$

Let t be the observed value of the test statistic T_X. Are we going to reject H_0 if t is too large or if t is too small? Justify.

6. Let $X_1, X_2, \dots, X_n$ be an iid sample from an unknown continuous density f. Consider the kernel density estimator of f, i.e.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where the kernel K is bounded and satisfies $\int K(x)\, dx = 1$, $\int xK(x)\, dx = 0$ and $0 < \int x^2 K(x)\, dx < \infty$.

(a) Show that the bias of the estimator is given by

$$\text{Bias}[\hat{f}(x)] = K_h * f(x) - f(x),$$

where $K_h(x) = h^{-1}K(x/h)$.

(b) Suppose now that the density f has two bounded and continuous derivatives. Show that the bias satisfies

$$\text{Bias}[\hat{f}(x)] = \frac{h^2}{2} f''(x) \int x^2 K(x)\, dx + o(h^2).$$

Hint: use Taylor expansion of f.