QUALIFYING EXAM IN STATISTICS, MAY 2006

Instr. A. Delaigle, I. Abramson
MATH 281ABC

Student name:
Student ID (last 5 digits):

---

1. Let $X_1, X_2, \ldots, X_n$ be an iid sample with pdf $f(x) = \dfrac{2x}{\theta^2}$, with $0 < x < \theta$, $\theta > 0$. Let $T = X_{(n)}$, the maximum of $X_1, X_2, \ldots, X_n$.
   (a) Prove that T is sufficient for θ.
   (b) Find the density of T.
   (c) Is T complete? *Hints: using the density found in (b) and using a certain derivative will help.*

2. Let $X_1, X_2, \ldots, X_n$ be an iid sample with pdf
$$f(x) = \frac{\theta^2}{x^3} \exp(-\theta/x),$$
with $0 < x < \infty$, $\theta > 0$.
   (a) Prove that the maximum likelihood estimator of θ is
$$\hat{\theta}_{MLE} = \frac{2n}{\sum_{i=1}^{n} \frac{1}{Y_i}}$$
   (b) Is $\hat{\theta}_{MLE}$ sufficient for θ?
   (c) Find the Fisher information matrix $I_n(\theta)$

Question 3. $w(x) = \frac{15}{16}(1-x^2)^2 \quad (|x| \le 1)$ is a nice smooth bell-shaped density function (the "Tukey biweight").

A bimodal density $f_0(x)$ can be formed by taking

$f_0(x) = \frac{1}{2}\left(w(x+1) + w(x-1)\right) \quad (|x| \le 2)$. Sketch it. Then a location family $\mathcal{L} = \{f_\mu(x)\}_{\mu \in \mathbb{R}}$ can be formed from the shape f_0 by setting $f_\mu(x) = f_0(x-\mu)$, with center of symmetry μ.

(a) Define consistency and $\sqrt{n}$-consistency of an estimator.

(b) We had a result as follows: "$X_1, \ldots, X_n \sim$ iid density f, invertible cdf F. Then the sample p-quantile $\hat{x}_p = X_{(\lceil pn \rceil)}$ $(0<p<1)$ is $\sqrt{n}$-consistent for the true quantile $x_p = F^{-1}(p)$. Specifically $\sqrt{n}(\hat{x}_p - x_p) \xrightarrow{d} \mathcal{N}\left(0, \cancel{E(x_p)}\ \frac{p(1-p)}{f(x_p)^2}\right)$ as long as $f(x_p) > 0$." (We proved the case $p = \frac{1}{2}$ in detail.)

Show that the sample median from a density in $\mathcal{L}$ is not $\sqrt{n}$-consistent for the true median μ. Is the difficulty in the bias or the variance or both? Is it consistent at all?

(c) Show that nevertheless there is still an estimate of μ that is based on the 25th percentile (call it $\hat{\mu}$), that is $\sqrt{n}$-consistent.

(d) Show that the average of the 25th and 75th percentiles is

$\sqrt{n}$-consistent as well, and that in fact its bias and variance are both at least as good as those of $\hat{\mu}$.

(e) How would you compute the ARE of $\bar{X}$ and $\hat{\mu}$ here? (There is some messy polynomial calculus involved - incompleteness or minor errors in the calculation are forgivable, but sketch the method and state the theorem that it's based on.)

(f) Assuming that $\mathcal{L}$ is regular enough to apply the fundamental asymptotic theorems of maximum likelihood to, show that the exact MLE is difficult to solve for, but that there is an equally efficient estimator based on Newton's method. Describe the procedure and write an expression for its asymptotic mean squared error, with brief explanation, but no detailed proofs of the theorems you refer to.

Question 4. (a) Describe the Wilcoxon rank-sum test. Give the usual distributional assumptions and hypotheses for which the theory is derived. No need for detail about the null distribution just give the form of the rejection region.

(b) Find a ~~density in~~ pairing of null and alternative for which the α-level test is inconsistent (power $\not\to 1$). How does the power behave if it doesn't tend to 1?

Question 5. Let $\Theta = \{0, 1\}$ and Z have one of the probability mass functions

z:	z_1	z_2	z_3	z_4	z_5
$f_0(z)$ $(H_0 : \theta = 0)$	.2	.3	.1	.3	.1
$f_1(z)$ $(H_1 : \theta = 1)$	.3	.1	.3	.2	.1

(a) Give the Neyman-Pearson test on Z operating at level .3, and calculate its power.

(b) Add another distibution f_2 to the alternative hypothesis (having the same support), so you test is uniformly most powerful against $H_2 = \{f_1, f_2\}$.

(c) Add a third such distibution f_3, so you test is no longer UMP against $H_3 : \{f_1, f_2, f_3\}$.

(d) If a Bayesian sets a prior on Θ

θ:	0	1
	.4	.6

derive the Bayes rule for estimating θ, and compute the Bayes risk of the Bayes rule (assume your loss function is squared error as usual).