Midterm Examination

STA 215: Statistical Inference

Due Wednesday, 2006 Mar 8, 1:15 pm

This is an open-book take-home examination. You may work on it during any consecutive 24-hour period you like; please record your starting and ending times on the lines below.

If a question seems ambiguous or confusing, or even if you're just stuck and need a hint, *please* ask *me*— don't guess, and don't discuss exam questions with others.

Unless a problem states otherwise, you must **show** your **work** to get partial credit. It is to your advantage to write your solutions as clearly as possible, since I cannot give credit for solutions I do not understand. Good luck.

Please detatch (or copy) this sheet and staple it to the top of your solutions before you turn them in.

Print Name: ____________________

Start Time: ____________________

End Time: ____________________

1.	/20
2.	/20
3.	/20
4.	/20
5.	/20
Total:	/100

1. In independent Bernoulli trials, all with success probability $p \in (0, 1)$, the number X of failures before the first success has a Geometric distribution, $X \sim \mathsf{Ge}(p)$, with probability mass function

$$f(x \mid p) = p\,(1-p)^x, \qquad x \in \mathbb{Z}_+ \equiv 0, 1, 2, \cdots .$$

The mean and variance are $\mathsf{E}[X] = (1-p)/p$ and $\mathsf{V}[X] = (1-p)/p^2$, respectively; in particular note that the mean is a *decreasing* function of p, so large values of X are associated with small values of p. Please answer the following questions about this $\mathsf{Ga}(p)$ distribution.

All parts of this question concern a single observation X from $f(x \mid p)$—there is no repeated sampling.

(a) Is this an exponential family?
If so, write the p.d.f. in standard form

$$f(x \mid p) = e^{\eta(p)\cdot T(x) - B(p)}\, h(x)$$

for suitable $q \in \mathbb{N}$, $\eta(p) \in \mathbb{R}^q$, $T(x) \in \mathbb{R}^q$, $B(p) \in \mathbb{R}$, and $h(x) \geq 0$ (specify q, η, T, B, and h); if not, explain why (no proof needed).

(b) Find the Maximum Likelihood Estimator $\hat{p} = \hat{p}(x)$, for the single observation $X = x$.

(c) Find the Fisher Information $I(p)$ for one observation. Simplify!

(d) Find the posterior mean $\bar{p}_\pi$ for a Beta prior $\pi = \mathsf{Be}(\alpha, \beta)$ with p.d.f. $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$, $0<p<1$ for $\alpha, \beta \in \mathbb{R}_+$, for the single observation $X = x$.

(e) Find the Jeffreys prior density $\pi_J(p)$ and corresponding posterior density $\pi_J(p \mid x)$, again for a single observation $X = x$.

(f) Find an exact 90% confidence interval $[L_x, R_x]$ for p on the basis of the single observation $X = x$. Evaluate it numerically, to four decimal places, for $x = 4$. Is L_4 bigger or smaller than L_7?

2. The Kullback-Leibler divergence between two distributions on the same space $\mathcal{X}$ is $K(f : g) \equiv \mathsf{E}_f\{\log[f(X)] - \log[g(X)]\}$, the expectation (under $X \sim f$) of the log ratio of the p.d.f.'s or p.m.f's, $f(X)$ and $g(X)$, *i.e.*,

$$K(f : g) \equiv \int_{\mathcal{X}} -f(x) \log\left[\frac{g(x)}{f(x)}\right] dx$$

for continuous distributions with p.d.f.'s f, g, or

$$K(f : g) \equiv \sum_{x \in \mathcal{X}} -f(x) \log\left[\frac{g(x)}{f(x)}\right]$$

for discrete distributions with p.m.f.'s f, g. No integration is needed, just a few moments, for each of the problems below. For example, if $f_j \sim \mathsf{Ex}(\lambda_j)$ for $j = 0, 1$ then $K(f_0 : f_1) = \mathsf{E}_0[-\log(f_1/f_0)]$ is just:

$$K(f_0 : f_1) = \log(\lambda_0/\lambda_1) + (\lambda_1 - \lambda_0)\mathsf{E}_0 X = \log(\lambda_0/\lambda_1) + (\lambda_1/\lambda_0) - 1.$$

(a) Evaluate the divergence $K(f_0 : f_1)$ for two normal distributions

$$f_0 : \quad \mathsf{No}(\mu_0, \sigma^2) \qquad f_1 : \quad \mathsf{No}(\mu_1, \sigma^2)$$

with the same variance but (possibly) different means.

(b) Evaluate the divergence $K(f_0 : f_1)$ for two normal distributions

$$f_0 : \quad \mathsf{No}(\mu, \sigma_0^2) \qquad f_1 : \quad \mathsf{No}(\mu, \sigma_1^2)$$

with the same mean but (possibly) different variances.

(c) Evaluate the divergence $K(f_0 : f_1)$ for two Poisson distributions

$$f_0 : \quad \mathsf{Po}(\lambda_0) \qquad f_1 : \quad \mathsf{Po}(\lambda_1).$$

(d) Evaluate the divergence $K(f_0 : f_1)$ for two Binomial distributions

$$f_0 : \quad \mathsf{Bi}(n, p_0) \qquad f_1 : \quad \mathsf{Bi}(n, p_1)$$

with the same n.

(e) Let $\lambda_0 = \lambda$ and $\lambda_1 = \lambda e^{\epsilon}$ for the Poisson case above. Is there a number $p \in \mathbb{R}$ such that the limit

$$\lim_{\epsilon \to 0} K(f_0 : f_1)/|\epsilon|^p$$

is finite and non-zero? (Hint: Try a Taylor expansion)
Evaluate the limit (as a function of $\lambda > 0$), or show none exists.

3. Let $U_1, \cdots, U_n$ be $n \geq 2$ independent draws from the $\mathsf{Un}(a, b)$ distribution, uniform on some interval $(a, b) \subset \mathbb{R}$, with $a < b$ both unknown. Set

$$X \equiv \min(U_1, ..., U_n) \qquad Y \equiv \max(U_1, ..., U_n) \tag{1}$$

It turns out that the vector (X, Y) is sufficient for (a, b).
(**Extra credit**: Prove this.)

(a) Find the joint p.d.f. $f(x, y \mid a, b)$ for X and Y.
Hint: Starting by considering $\mathsf{P}\{X > x, \quad Y \leq y\}$ for all numbers $-\infty < x < y < \infty$; be careful about the ranges of x and y.

(b) Find the M.L.E.'s $\hat{a}$ for a and $\hat{\mu}$ for the mean $\mu \equiv \mathsf{E}[U_j] = \frac{a+b}{2}$. Are either of them biased? How much?

(c) Find an exact 90% equal-tail confidence interval $[L, R]$ for the mean $\mathsf{E}[U_j] = \frac{a+b}{2}$ for a sample $\{U_j\}$ of size n.
Hint: How do you expect $L = L(X, Y)$ to depend on X and Y?

(d) For a sample of size $n \geq 2$, find the conditional probability distribution of each U_j, given X, Y. Describe it in words, or give the conditional CDF $F(u \mid x, y) \equiv \mathsf{P}[U_j \leq u \mid X = x, Y = y]$ correctly for all x, y, u. Calculate $\mathsf{E}[U_j \mid X, Y]$ carefully. Hint: What is $\mathsf{P}[U_1 = X]$?

(e) The statistic $T(\vec{U}) = \bar{U} \equiv \frac{1}{n} \sum U_j$ is an unbiased estimator of $\mu = \frac{a+b}{2}$, but it is not sufficient. The Raô-Blackwell theorem suggests a specific improvement T^* of T. Find T^* explicitly.

4. Now let's adopt a Bayesian perspective for the $\mathsf{Un}(a,b)$ problem. Again take $\{U_j\} \overset{\text{iid}}{\sim} \mathsf{Un}(a,b)$, with parameter $\theta = (a,b)$ unknown.

   (a) Some might advocate the improper prior density function

   $$\pi(a,b) = \frac{1}{(b-a)}, \qquad -\infty < a < b < \infty$$

   as an expression of ignorance about a, b. How large does the sample size n have to be for the posterior distribution with density

   $$\pi(a,b \mid \vec{U}) \propto \pi(a,b) f(U_1,\ldots,U_n \mid a,b)$$

   to be well-defined?[1] You must evaluate the normalizing constant to find $\pi(a,b \mid \vec{U})$; if it turns out to be zero or infinity then the proposed improper prior does not lead to a well-defined posterior, but if it is finite and positive then the posterior is well-defined.

   (b) Find the posterior mean estimators $\bar{a}_\pi \equiv \mathsf{E}_\pi[a \mid \vec{U}]$ and $\bar{\mu}_\pi \equiv \mathsf{E}_\pi[\frac{a+b}{2} \mid \vec{U}]$ for this prior. Are they biased? How much?

   (c) Others might have advocated a uniform improper prior density

   $$\pi_U(a,b) = 1, \qquad -\infty < a < b < \infty$$

   How large must the sample size n be for the posterior with *this* prior to be well-defined? Can you characterize the inferential difference between inference using π and inference using π_U? Are the differences smaller when n is big or small?

[1] Any sufficient statistic $S(\vec{U})$ for (a,b) can be used instead of the whole vector $\vec{U} = (U_1,\ldots,U_n)$ for evaluating posterior densities and expectations.

5. An experiment generates n independent exponentially-distributed random variables $\xi_j \overset{\text{iid}}{\sim} \mathsf{Ex}(\lambda)$ (with p.d.f. $f(\xi) = \lambda e^{-\lambda\xi}$ for $\xi > 0$); alas we observe only truncated versions

$$X_j \equiv \xi_j \wedge 1 = \begin{cases} \xi_j & \text{if } 0 < \xi_j < 1 \\ 1 & \text{if } 1 \le \xi_j < \infty. \end{cases}$$

(Think of readings on a meter that only has a range of $[0, 1]$.)

(a) Evaluate the common CDF for each of the observables $\{X_j\}$:

$$F(x \mid \lambda) = \mathsf{P}[X_j \le x \mid \lambda]$$

correctly for *all* $x \in \mathbb{R}$ and $\lambda > 0$. You might like to consider separately the three cases $x \in (-\infty, 0)$, $x \in [0, 1)$, and $x \in [1, \infty)$.

(b) Find the likelihood function for λ on the basis of a random sample of size n, $\{X_1, \cdots, X_n\}$.
Hint: Some of the X_j's will be exactly 1, the rest will be in $(0, 1)$. What are some suffient statistics?

(c) Find the M.L.E. $\hat{\lambda}$ upon observing $\{X_j\}$, $j = 1 \cdots n$.

(d) Find the Fisher information $I_X(\lambda)$ for observing $\{X_j\}$. Is it larger or smaller than the Fisher information $I_\xi(\lambda)$ for the uncensored observations $\{\xi_j\}$? Is it larger or smaller than the Fisher information $I_T(\lambda)$ for the binomially-distributed number $T = \#\{j : X_j = 1\}$ of truncated observations ($\xi_j \ge 1$)? It is enough to consider the case of a single observation, $n = 1$.

Extra Credit: Use `R`, `S-Plus` or Matlab to plot all three (overlaid) on the two ranges $0.5 < \lambda < 1$ and $1 < \lambda < 4$, using colors or line type to distinguish the curves (include a legend). Describe what happens for large λ and for small λ. Why?