18.466 Problem set 4, due Monday, April 1, 2013

1. Bickel and Doksum, p. 197, problem 2.

2. For X having a binomial (n, p) distribution, we have as estimators of p the classical unbiased and maximumum likelihood estimator $\hat{p} := X/n$, and we have the Bayes estimate for a uniform prior, $p_B := (X+1)/(n+2)$. For $n = 20$, find for what values of p each of these has smaller mean-squared error.

3. Bickel and Doksum, p. 203, problem 2. Note: Jensen's inequality also holds for conditional expectations.

4. Consider the family of $N(\theta, 1)$ distributions with the Cauchy prior density

$$\pi(\theta) = \frac{1}{\pi(1+\theta^2)},$$

$-\infty < \theta < +\infty$. Suppose given one observation x from $N(\theta, 1)$ and that we want to estimate θ. Show that the expectation of $|\theta|$ with respect to the prior is infinite, but that a Bayes estimator of θ exists. (You are not asked to find it in closed form.)

5. The Jeffreys prior (Bickel and Doksum, p. 203, Problem 15, but with a part omitted and others added). (By the way this prior was invented by H. Jeffreys, so it should not be called "Jeffrey's prior.") Suppose we have a parametric family of probability mass functions or densities, with a likelihood function $f(\theta, x)$, defined for θ in an open interval (a, b) where $-\infty \le a < b \le +\infty$. The *Fisher information* of the family, at a given θ, is defined as

$$I(\theta) = E_\theta\left[(\partial \log f(\theta, x)/\partial\theta)^2\right]$$

("Information inequalities" handout, non-numbered display in the middle of p. 3). The *Jeffreys prior* measure Π_J on (a, b) has a density $j(\theta) := \sqrt{I(\theta)}$, so that for $a \le c < d \le b$,

$$\Pi_J((c, d)) = \int_c^d j(\theta)d\theta.$$

This may be improper (have infinite total measure), but if it is finite, the *Jeffreys prior probability* is found by normalizing it.

(a) Show that for $(a, b) = (-\infty, \infty)$, and $\theta = \mu$, for the family of $N(\mu, 1)$ densities, the Jeffreys prior density is a positive constant and therefore, the prior is improper (has total measure $+\infty$).

(b) In the general case, suppose we have a parameter ψ satisfying $u < \psi < v$ for some u, v with $-\infty \le u < v \le +\infty$, and a continuously differentiable function h from (u, v) onto (a, b) with $h'(x) > 0$ for all $x \in (u, v)$. Then $g(\psi, x) = f(h(\psi), x)$ gives a reparameterization of the same family of densities. Bickel and Doksum, Problem 3 p. 203, parts (a) and (b), ask one to show that the Fisher information $I(\theta)$ is not equivariant, i.e. as defined with respect to $f(\cdot, x)$ at θ, it may not be the same as $I(\psi)$ defined with respect to $g(\psi, x)$ with $h(\psi) = \theta$, but the information inequality lower bound is equivariant. This problem does not ask you to prove either of those things, rather, it asks you to show that the Jeffreys prior measure is equivariant, in fact, it is invariant under the kind of transformation we're considering, as follows. Let j_f be the Jeffreys prior density as defined above with respect to $f(\theta, x)$, and let j_g be the Jeffreys prior density defined with respect to $g(\psi, x)$. Then show that for any ξ, η with $u < \xi < \eta < v$, we have

$$\int_\xi^\eta j_g(\psi)d\psi = \int_{h(\xi)}^{h(\eta)} j_f(\theta)d\theta.$$

Hint: this is a change of variables in an integral, with use of the chain rule. If you prefer, do the special case in part (d) before this general case.
(c) Consider the Bernoulli (p) distributions for $0 < p < 1$, with likelihood function $p^X(1-p)^{1-X}$ where $X = 0$ or 1. Evaluate its Jeffreys prior probability density (Bickel and Doksum state what it is for a binomial distribution with general n). This is a density in a known family, namely what?
(d) Suppose we consider the same distributions as in (c) but with the exponential family parameterization where $\theta = \log(p/(1-p))$, $p = e^\theta/(1+e^\theta)$. Find explicitly the Jeffreys prior density with respect to θ (for it, part (b) would then hold for $\psi = \theta$).
(e) In "Information inequalities," Proposition 4, it is shown that if a family of distributions P_θ has a finite Fisher information $I_1(\theta)$ and a regularity condition (2) holds, then the Fisher information for n i.i.d. variables with a distribution in the family also has finite Fisher information $I_n(\theta) = nI_1(\theta)$. For $X_1, ..., X_n$ i.i.d. Bernoulli (p), $S_n = X_1 + \cdots + X_n$ is a sufficient statistic, why? Then S_n has a binomial (n, p) distribution. So, check that as Bickel and Doksum say, the Jeffreys prior probability density for this distribution for any n is the same as for $n = 1$ found in (c).