Statistics 582, Final Exam

Wellner; 3/6/98

Due date: Friday, March 13, 1998 before 4 PM in the Statistics Office

Instructions: This is a "take-home" and "open-book" exam. You are on your honor to do it completely independently from other students in the class with **absolutely no discussion or collaboration** about the problems with other students in the class or with other students, faculty, or other outside assistance.

1. (36 points) (More inequalities involving Hellinger distance and total variation distance.) Suppose that P and Q are two probability measures with densities p and q with respect to some dominating measure μ. So that the Hellinger distance $d_H(P,Q)$ is given by

$$d_H^2(P,Q) = (1/2)\int |\sqrt{p} - \sqrt{q}|^2 d\mu = 1 - \rho(P,Q)$$

where $\rho \equiv \rho(P,Q) = \int \sqrt{pq}\, d\mu$, and the total variation distance $d_{TV}(P,Q)$ is given by

$$d_{TV}(P,Q) = (1/2)\int |p - q| d\mu = 1 - \pi(P,Q)$$

where $\pi \equiv \pi(P,Q) = \int p \wedge q \, d\mu$.

A. Prove the following inequalities:

$$\pi^2 \le \rho^2 \le 1 - (1-\pi)^2 = \pi(2-\pi) \le 2\pi\,.$$

Thus we have

$$(1 - d_H^2(P,Q))^2 \le 1 - d_{TV}(P,Q)^2 = 1 - \left(1 - \int p \wedge q \, d\mu\right)^2 \le 2\int p \wedge q d\mu\,.$$

[Hint: The first inequality is easy. To prove the second inequality, use Cauchy-Schwarz to show that $\rho(P,Q)^2 + d_{TV}(P,Q)^2 \le 1$.]

B. Show that $K(P,Q) \ge d_H^2(P,Q)$. [Hint: Use Jensen's inequality and $\log(1-x) \le -x$.]

2. (56 points) Suppose that $X \sim \text{Hypergeometric}(n, \theta, M)$; that is, X is the number of red balls drawn in n draws without replacement from an urn containing θ red balls and $M - \theta$ black balls, and it has probability mass function

$$P_\theta(X = x) = \frac{\binom{\theta}{x}\binom{M-\theta}{n-x}}{\binom{M}{n}}, \qquad \text{for} \quad x = 0 \vee (\theta + n - M), \cdots, \theta \wedge n\,.$$

Suppose that θ has a Beta-Binomial prior distribution $BetaBin(\alpha, \beta, M)$ with mass function

$$\lambda(\theta) = \binom{M}{\theta} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(M+\alpha+\beta)} \qquad \text{for} \qquad \theta = 0, 1, \ldots, M\,.$$

Here M is a positive integer and $\alpha > 0$, $\beta > 0$. Suppose that the loss function is squared error: $L(\theta, a) = (\theta - a)^2$.

A. What is the maximum likelihood estimator of θ? (Note that differentiation won't work here since θ is integer-valued.) What, if any, of our theory for MLE's applies in this case?

B. Show that the unconditional distribution of X is $BetaBin(\alpha, \beta, n)$.

C. Show that the conditional distribution of $\theta - x$ given $X = x$ is $BetaBin(x + \alpha, n - x + \beta, M - n)$.

D. Show that the Bayes estimate of θ is

$$d(X) = \frac{(M+\alpha+\beta)X + \alpha(M-n)}{n+\alpha+\beta}\,.$$

E. Find the risk function of the estimates $d_{a,b}(X) = aX + b$.

F. Show that the estimate $d_0(X) = a_0 X + b_0$ has a constant risk equal to b_0^2 when

$$a_0 = \frac{M}{n(1+\delta)}, \qquad b_0 = (M - a_0 n)/2$$

where $\delta = \sqrt{(M-n)/(n(M-1))}$.

G. Show that the estimate d_0 is minimax (Bayes wrt $BetaBin(\alpha, \beta, n)$ when $\alpha = \beta = b_0/(a_0 - 1)$). Illustrate with the special case $N = 10$ and $n = 4$.

[Hint: See Ferguson page 100 for properties of the Beta-Binomial distributions.]

H. Compare and contrast this problem with the corresponding sampling with replacement situation in which $X \sim \text{Binomial}(n, \theta)$ and $\theta \sim \text{Beta}(\alpha, \beta)$. To begin this comparison, what is the natural analogue, in the "without replacement" (Hypergeometric) problem, of $\theta \in (0, 1)$ in the "with replacement" problem?

3. (48 points) (Testing in a 2×2 contingency table). Consider a two-dimensional contingency table with two rows and two columns. Let θ_{jk} be the cell probability for the jth row and kth column, so that $\sum_{j=1}^{2}\sum_{k=1}^{2}\theta_{jk} = 1$. Suppose that $\underline{X} = (X_{11}, X_{12}, X_{21}, X_{22}) \sim \text{Mult}_4(n; \underline{\theta} = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}))$. The following questions concern testing hypotheses about the cross-ratio $\psi \equiv \theta_{11}\theta_{22}/(\theta_{12}\theta_{21})$. Note that $\psi = 1$ if and only if $\theta_{jk} = \theta_{j.}\theta_{.k}$ for $j = 1, 2$ and $k = 1, 2$, i.e. independence holds; here $\theta_{j.}$ and $\theta_{.k}$ are the marginal cell probabilities.

A. What is the likelihood ratio test of $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$? What are the large-sample properties of this test under:
(a) the null hypthesis;
(b) local alternatives of the form $\psi_n = 1 + tn^{-1/2}$;
(c) fixed alternatives?

B. Describe at least one other test for testing the hypotheses in A. Compare and contrast it to the LR test.

C. How would you construct an approximate $1 - \alpha$ confidence interval for the parameter ψ?

D. Let $\Theta_0 \equiv \{\underline{\theta} : \psi \leq 1\}$, and $\Theta_1 \equiv \Theta_0^c = \{\underline{\theta} : \psi > 1\}$, Now suppose that we assume the Dirichlet($\underline{\alpha}$) prior distribution for $\underline{\theta}$; here $\underline{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$ and the prior density is of the form

$$\lambda(\underline{\theta}) = c(\underline{\alpha}) \prod_j \prod_k \theta_{jk}^{\alpha_{jk}-1} 1_{[\sum_j \sum_k \theta_{jk}=1]}(\underline{\theta}) \, .$$

Derive the Bayes test for $0-1$ loss of Θ_0 versus Θ_1.
[Hint: Find the posterior density of $\underline{\theta}$, and use this to show that the two variables $\theta_{11}/\theta_{1.}$ and $\theta_{21}/\theta_{2.}$ are independent (aposteriori) with posterior densities which are Beta$(N_{11} + \alpha_{11} - 1, N_{12} + \alpha_{12} - 1)$ and Beta$(N_{21} + \alpha_{21} - 1, N_{22} + \alpha_{22} - 1)$. Use this to show that the posterior probability that $\psi < 1$ is

$$P(\psi < 1 \big| \underline{N}) = P(\frac{\theta_{11}}{\theta_{1.}} < \frac{\theta_{21}}{\theta_{2.}} \big| \underline{N}) = \sum_{k=k_0}^{N_{12}+\alpha_{12}-1} \binom{N_{.1} + \alpha_{.1} - 1}{k} \binom{N_{.2} + \alpha_{.2} - 1}{N_{2.} + \alpha_{2.} - k} \Big/ \binom{N_{..} + \alpha_{..} - 2}{N_{1.} + \alpha_{1.} - 1}$$

where $k_0 \equiv 0 \vee (N_{21} + \alpha_{21} - N_{12} - \alpha_{12})$.]

4. (48 points) Suppose that $X_i = (Y_i, Z_i)$, $i = 1, \ldots, n$ are i.i.d. with

$$(Y|Z = z) \sim Binomial(m, p(z, \alpha, \beta)$$

where $\theta \equiv (\alpha, \beta) \in R \times R$ and

$$p(z, \alpha, \beta) = (1 + \exp(-(\alpha + \beta z)))^{-1},$$

suppose that the distribution H of Z is known and that Z is not degenerate at a single point. You may assume that Z is bounded: $|Z| \le c$ with probability 1 for some $c < \infty$. Consider two different procedures for estimating the parameters α and β as follows:
(i) Maximum Likelihood;
(ii) Minimum Logit-χ^2 defined as follows: Define $\text{logit}(p) = \log(p/(1-p))$, and $\widehat{p}_i = Y_i/m$, $i = 1, \ldots, n$. Choose $\theta = (\alpha, \beta)$ to minimize Berkson's "minimum logit-χ^2 $B_n(\alpha, \beta)$ defined by

$$B_n(\alpha, \beta) \equiv \sum_{i=1}^{n} \widehat{p}_i(1 - \widehat{p}_i) \left\{\alpha + \beta Z_i - \text{logit}(\widehat{p}_i)\right\}^2 .$$

Call the resulting "minimum logit-χ^2" estimator $\widetilde{\theta}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)$. [This is a reformulation of the problem discussed on pages 123 and 124 of Ferguson, MS.]
A. Find the scores for (α, β) based on one observation $X = (Y, Z)$.
B. What is the information matrix for θ? Hint: compute conditionally on Z, and leave the information matrix in terms of expectations of functions of Z.
C. Compute the "minimum logit-χ^2" estimator $\widetilde{\theta}_n$, and show that it is consistent and asymptotically normal.
D. Prove or disprove the claim in Ferguson, page 124, lines 7 - 10: "... these estimates have all the large sample optimal properties of the maximum likelihood and minimum χ^2 estimates."
[Hint: one way to prove the claim would be to show that the maximum likelihood estimator $\widehat{\theta}_n = (\widehat{\alpha}_n, \widehat{\beta}_n)$ is asymptotically equivalent to the "minimum logit-χ^2" estimator $\widetilde{\theta}_n$ in the sense that $\sqrt{n}(\widehat{\theta}_n - \widetilde{\theta}_n) = o_p(1)$.]