May 2011

Ph.D. COMPREHENSIVE EXAMINATIONS
DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO

THEORETICAL STATISTICS COMPREHENSIVE EXAMINATION

May 16, 2011, 12:30 p.m. – 4:30 p.m.

Sidney Smith Hall

(# questions = 6); (# pages = 5, including cover page)

1. ATTEMPT ALL OF QUESTIONS 1-6
2. Please work neatly and legibly, and start each question in a **new** book with your name and the number of the question. If you use more than one book for a question then also indicate which book is the first for a question, which is the second, and so on.
3. It is not necessary to completely solve every problem to achieve a good performance: *please emphasize what you know.*
4. The questions are not in any special order, and they are not of equal difficulty.
5. The problems here may be improperly phrased or may contain a misprint. Should this happen, reflect it in your discussion.
6. You are **not** permitted any aids (e.g. books, notes, etc.) aside from a single non-programmable calculator.
7. Good luck!

1. Let $\mathcal{P}_0$ and $\mathcal{P}$ be two families of distributions with $\mathcal{P}_0 \subset \mathcal{P}$. Assume that $P(A) = 0$ for all $P \in \mathcal{P}_0$ implies that $P(A) = 0$ for all $P \in \mathcal{P}$.

   (a) If a statistic T is sufficient for $\mathcal{P}$ and minimal sufficient for $\mathcal{P}_0$, show that T is also minimal sufficient for $\mathcal{P}$.

   (b) Suppose that $\mathcal{P}_0$ is a finite family of distributions with densities $f_0, \cdots, f_k$ with respect to some σ-finite measure μ where $f_i(x) = 0$ for some i implies that $f_j(x) = 0$ for all $j \neq i$. If we observe X from $\mathcal{P}_0$, show that the statistic

   $$T(X) = \left(\frac{f_1(X)}{f_0(X)}, \cdots, \frac{f_k(X)}{f_0(X)}\right)$$

   is minimal sufficient for θ.

   (c) Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with common density (w.r.t. counting measure)

   $$f_\theta(x) = \begin{cases} \theta & \text{if } x = -1 \\ (1-\theta)^2\theta^x & \text{if } x = 0, 1, 2, \cdots \end{cases}$$

   where $0 < \theta < 1$. Show that the statistic

   $$T = \left(\sum_{i=1}^{n} I(X_i = -1), \sum_{i=1}^{n} X_i\right)$$

   is minimal sufficient for θ. (Hint: Write the density f_θ as a two-parameter exponential family. Then show that T is minimal sufficient for some finite sub-family.)

   (d) If X_i has the density f_θ then $E_\theta(X_i) = 0$. (You may assume this is true.) Is the statistic T defined in part (c) complete?

2. Suppose that $X_1, \cdots, X_n$ are independent random variables with density

   $$f_\lambda(x) = \lambda \exp(-\lambda x) \quad \text{for } x \geq 0$$

   (a) Show that $T = \sum_{i=1}^{n} X_i$ is sufficient and complete for λ.

   (b) Show that T is independent of $(X_1/T, \cdots, X_n/T)$.

   (c) Is the independence in part (b) true if the $\{X_i\}$ are i.i.d. Gamma random variables? (The Gamma density is provided in question 4(c).)

   (You can appeal to well-known results for the proofs although you need to show that the hypotheses of these results hold.)

3. Suppose $Y_1, \ldots, Y_n$ are independent and identically distributed from a density $f(y \mid \theta), \theta \in \mathbb{R}$.

(a) Let $\ell(\theta; \underline{y})$ be the log-likelihood function for θ based on the sample $\underline{y} = (y_1, \ldots, y_n)$. Argue using Taylor series expansions that the likelihood ratio statistic

$$w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\},$$

is asymptotically equivalent to

$$w_e(\theta) = (\hat{\theta} - \theta)^2 i(\theta),$$

under the model, and describe the regularity conditions needed to establish this. Here $\hat{\theta}$ is the maximum likelihood estimator and $i(\theta)$ is the total expected Fisher information from the log-likelihood function $\ell(\theta; \underline{y})$.

(b) The Laplace approximation to the integral $\int h(x) \exp\{-n g_n(x)\} dx$ is given by

$$\int h(x) \exp\{-n g_n(x)\} dx = h(\tilde{x}) \exp\{-n g_n(\tilde{x})\} \sqrt{\frac{2\pi}{n}} \{g_n''(\tilde{x})\}^{-1/2} \{1 + O(n^{-1})\},$$

where $\tilde{x}$ satisfies $g_n'(\tilde{x}) = 0$, and it is assumed that $h(\cdot)$ does not depend on n, and that $h(\tilde{x}) \neq 0$.[1]

Suppose $\pi(\theta)$ is a prior probability density function for θ. Define the estimator $\tilde{\theta}$ as the posterior mean:

$$\tilde{\theta}(\underline{y}) = \frac{\int \theta \pi(\theta \mid \underline{y}) d\theta}{\int \pi(\theta \mid \underline{y}) d\theta}.$$

Using the Laplace approximation for integrals, show that $\tilde{\theta} - \hat{\theta} = O_p(1/n)$, stating clearly any regularity conditions needed additional to those in (a).

(c) Suppose $f(y \mid \theta)$ is the Pareto density, given by

$$f(y \mid \theta) = \theta(1+y)^{-(\theta+1)}, \quad y > 0, \theta > 0.$$

Describe the form of the conjugate prior for θ, and give an expression for the posterior mean, $\tilde{\theta}(\underline{y})$.

(d) Assume now that $\theta = (\psi, \lambda)$ with $\psi \in \mathbb{R}$, $\lambda \in \mathbb{R}^k$. Define the *profile log-likelihood* function for ψ, and explain how it can be used for approximate inference about ψ.

[1] Strictly speaking we should also assume $g_n(x) = g(x) + O(n^{-1})$ for $|x - \tilde{x}| < \delta/\sqrt{n}$, but you can assume the result is correct as stated.

4. Suppose $T = t(Y)$ is a test statistic, for testing a null hypothesis H_0. Assume that the distribution of T under H_0 is continuous, and known, and that the larger the observed value of t, the stronger the evidence against H_0. The p-value or observed level of significance of $t^0 = t(y^0)$, given observed data y^0, is defined as

$$p^0 = \mathrm{Pr}_0(T \geq t^0),$$

where Pr_0 means probability under the null hypothesis H_0.

(a) Show that the random variable $P^0 = p^0(Y)$ follows a uniform distribution under H_0.

(b) The p-value *function* is defined for a scalar parameter model indexed by $\theta_0 \in \mathbb{R}$ by

$$p(\theta_0) = \mathrm{Pr}_{\theta_0}(T \geq t^0),$$

where the null hypothesis is $H_0 : \theta = \theta_0$. Give an expression for the p-value function for inference about θ in the model

$$f(y;\theta) = \theta e^{-y\theta}, \quad y > 0; \theta > 0$$

based on an independent sample $\underline{y} = (y_1, \ldots, y_n)$ from $f(y;\theta)$ and the test statistic given by the standardized score $\ell'(\theta_0)\{i(\theta_0)\}^{-1/2}$, where $i(\theta_0)$ is the expected Fisher information in the sample under H_0.

(c) The gamma distribution with unknown shape and scale,

$$f(y;\theta,\beta) = \frac{1}{\Gamma(\beta)}\theta^\beta y^{\beta-1}e^{-\theta y},$$

has moment generating function

$$M_Y(t) = \mathrm{E}(e^{tY}) = \frac{\theta^\beta}{(\theta+t)^\beta},$$

and hence cumulant generating function

$$K_Y(t) = \log M_Y(t) = \beta\log\theta - \beta\log(\theta+t).$$

The jth cumulant is defined by

$$\kappa_j = K^{(j)}(0),$$

and series expansions of $K_Y(t)$ and $M_Y(t)$ about 0 can be used to derive the equalities

$$\kappa_1 = \mathrm{E}(Y), \quad \kappa_2 = \mathrm{var}(Y), \quad \kappa_3 = \mathrm{E}(Y-\kappa_1)^3.$$

(You do not need to derive these results.)
Show that in the gamma model, $\kappa_3\kappa_1 = 2\kappa_2^2$, and use this to suggest a test statistic for the goodness-of-fit of a gamma model.

5. Suppose that $(X_1, Y_1), \cdots, (X_n, Y_n)$ are independent pairs of correlated Bernoulli random variables. The correlation between X_i and Y_i is induced as follows: Take $\theta_1, \cdots, \theta_n$ to be i.i.d. random variables on $[0, 1]$ with a non-degenerate distribution function G. Then given θ_i, X_i and Y_i are independent random variables with

$$P(X_i = 1|\theta_i) = \theta_i \quad \text{and} \quad P(Y_i = 1|\theta_i) = h(\theta_i, \phi)$$

where h is such that

$$\frac{h(\theta_i, \phi)}{1 - h(\theta_i, \phi)} = \phi \frac{\theta_i}{1 - \theta_i}.$$

Both ϕ and the distribution function G are unknown parameters in this model; given $(X_1, Y_1), \cdots, (X_n, Y_n)$, we want to estimate the parameter ϕ (the odds ratio).

(a) Show that $E[Y_i(1 - X_i)] = \phi E[X_i(1 - Y_i)]$.

(b) Show that the estimator

$$\widehat{\phi}_n = \frac{\sum_{i=1}^n Y_i(1 - X_i)}{\sum_{i=1}^n X_i(1 - Y_i)}$$

is a consistent estimator of ϕ.

(c) Find the limiting distribution of either $\sqrt{n}(\widehat{\phi}_n - \phi)$ or $\sqrt{n}(\ln(\widehat{\phi}_n) - \ln(\phi))$. Suggest an estimator of the standard error of $\widehat{\phi}_n$ (or of $\ln(\widehat{\phi}_n)$).

6. Let $Y_1, \ldots, Y_n$ be independent Bernoulli random variables with probability of success θ.

(a) Compare the risk functions under squared-error loss of the maximum likelihood estimator, $\hat{\theta} = \bar{Y}$, with that of the Bayes estimator using the *Beta*(a, b) prior

$$\pi(\theta) = \frac{1}{B(a, b)} \theta^{a-1}(1 - \theta)^{b-1}, \quad 0 \leq \theta \leq 1,$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$.

(b) Show that the Bayes estimator with the choice $a = b = \sqrt{n/4}$ has constant risk. What is the importance of this property?

(c) Suppose now that each Y_i is Bernoulli with probability of success θ_i, and $\theta_1, \ldots, \theta_n$ are independently distributed as *Beta*(a, b). Show that $\bar{y}$ is an empirical Bayes estimate of $a/(a + b)$.

Note: The mean and variance of the *Beta*(a, b) are $\dfrac{a}{a + b}$ and $\dfrac{ab}{(a + b)^2(a + b + 1)}$, respectively.