May 2014

Ph.D. COMPREHENSIVE EXAMINATIONS
DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF TORONTO

THEORETICAL STATISTICS COMPREHENSIVE EXAMINATION
May 16, 2014   12:30 p.m. – 4:30 p.m.
Sidney Smith Hall

1. ***Attempt all questions*** (total # of questions = 7). (total # of pages = 4 including cover page)
2. Please work neatly and legibly.
3. ***Start each question in a new book, with your name and the number of the question on the front cover.*** If there is more than one book for a question, then also indicate which is the first book and which second, e.g., Jane Smith, Question 5, Book 1 of 2.
4. The questions are not in any special order, nor are they all of equal difficulty.
5. The problems may be improperly phrased or may contain a misprint. Should this happen, reflect it in your discussion. Faculty members are *not* available to answer questions during the exam.
6. You are NOT permitted any aids (e.g., books, notes, etc.) **aside from a single non-programmable calculator**.
7. Good luck!

1. Let X be a positive real variable whose density function, given a parameter θ in $(0,1)$, is

$$f(x|\theta) = \begin{cases} 2(x-\theta)\,/\,\theta^2 & \text{if } \theta < x < 2\theta \\ 0 & \text{otherwise} \end{cases}$$

We specify a prior distribution for θ whose density function (over the parameter space $(0,1)$) is $f(\theta) = 3\theta^2$.

(a) Find as simple expression as you can for the posterior density function for θ given an observation x, including its normalizing constant. Remember that the parameter space is $(0,1)$.

(b) Suppose we need to take an action, $a \in \{0,1\}$ after observing x, with the loss we suffer depending on the value of θ, as follows:

$$L(\theta, a) = \begin{cases} 0 & \text{if } \theta < 1/2 \text{ and } a = 0 \\ 1 & \text{if } \theta < 1/2 \text{ and } a = 1 \\ 0 & \text{if } \theta \geq 1/2 \text{ and } a = 1 \\ 2 & \text{if } \theta \geq 1/2 \text{ and } a = 0 \end{cases}$$

Derive a decision rule for this problem that minimizes posterior expected loss, expressing it in as simple a form as you can.

2. Suppose $Y_1, \ldots, Y_n$ are independently and identically distributed from a distribution $F(\cdot)$ on $\mathbb{R}$, which is continuous and symmetric about the median of the distribution, defined as $\nu = F^{-1}(1/2)$. We are interested in testing $H_0 : \nu = \nu_0$, against the alternative that $H_A : \nu > \nu_0$.

(a) Statistician A proposes to use the test statistic $T_A = t_A(Y_1, \ldots, Y_n) = \sum_{i=1}^{n} 1(Y_i > \nu_0)$, which counts the number of observations greater than ν_0. What is the distribution of T_A under the null hypothesis, and what is the p-value associated with a test of significance using T_A?

(b) Statistician B proposes to use the test statistic $T_B = t_B(Y_1, \ldots, Y_n) = \bar{Y}$, on the grounds that if F is symmetric about ν, the mean and median of F are the same, and if $F(\cdot)$ is normal, the test based on T_B is uniformly most powerful for testing H_0 against H_A. Show that the test based on T_B is indeed uniformly most powerful under the normal model with median ν_0 and variance 1.

(c) Suppose that $T_1, \ldots, T_k$ are test statistics computed from a set of k similar measurements on n individuals. For example, the n individuals may be laboratory mice, and we might have n measurements of blood flow change after administering some chemical, at each of k locations on the skin surface. These k test statistics lead to k p-values associated with testing $H_{0j} : \nu_j = \nu_{0j}, j = 1, \ldots, k$, where ν_j is the median of the distribution of blood flow at location k. Show that if we 'reject' H_{0j} when $p_j < \alpha/k$ that the overall type one error for the collection of tests is bounded by α.

(d) Give expressions that would enable computation of the power loss of the test based on T_A, under the normal distribution. To simplify this expression it may be helpful to use the normal approximation to the binomial, and the approximation $\Phi(n^{-1/2}\delta) \doteq \frac{1}{2} + \delta(2\pi n)^{-1/2}$ for large n.

3. Let $Y_1, \ldots, Y_n$, where n is an odd integer greater than one, be i.i.d. observations from an $N(\theta, 1)$ distribution, where θ is an unknown real model parameter. Suppose we wish to estimate θ with our loss being the absolute value of the error — ie, the loss when the estimate is a is $L(\theta, a) = |\theta - a|$.

   Prove that the sample median of $y_1, \ldots, y_n$ is not an admissible estimator with this loss function, and find an estimator that dominates it.

4. (a) Define *sufficient statistic*, *minimal sufficient statistic*, *ancillary statistic*, and *maximal ancillary statistic.*

   (b) Suppose $Y_1, \ldots, Y_n$ are independently distributed from the $U(-\theta, \theta)$ distribution, with density
   $$f(y; \theta) = \frac{1}{2\theta}, -\theta < y < \theta, \theta \in \mathbb{R} \setminus \{0\}.$$
   Identify a (version of the) minimal sufficient statistic, and an ancillary statistic. Is the sufficient statistic complete?

   (c) Suppose $Y_1, \ldots, Y_n$ are independently distributed from the $N(\mu, \sigma^2)$ distribution. Show that $(\bar{y}, s)$ is minimal sufficient for $\theta = (\mu, \sigma)$, and that $a = (y - \bar{y}1)/s$ is ancillary for θ.

   (d) In a general model, is the maximum likelihood estimator $\hat{\theta}$ sufficient for θ? Why or why not?

5. Suppose that Y_{ij} are independent random variables from a normal distribution with
   $$\mathrm{E}(Y_{ij}) = \mu_i, \quad \mathrm{Var}(Y_{ij}) = \sigma^2; \quad j = 1, \ldots, m; i = 1, \ldots, n;$$
   and hence density function
   $$f(y_{ij}; \mu_i, \sigma^2) = \frac{1}{\sqrt{(2\pi)}\sigma} \exp\{-\frac{1}{2\sigma^2}(y_{ij} - \mu_i)^2\}.$$

   (a) Find the profile log-likelihood function for σ^2, and the maximum likelihood estimator of σ^2.

   (b) Show that $\hat{\sigma}^2$ is consistent for σ^2 if $m \to \infty$ with n fixed, but is not consistent for σ^2 if $n \to \infty$ with m fixed.

   (c) Show that the log-likelihood ratio statistic based on the profile log-likelihood function, $w_{\mathrm{p}}(\sigma^2) = 2\{\ell_{\mathrm{p}}(\hat{\sigma}^2) - \ell_{\mathrm{p}}(\sigma^2)\}$ takes the form
   $$w_{\mathrm{p}} = mn\left\{-\log\left(\frac{\hat{\sigma}^2}{\sigma^2}\right) - 1 + \frac{\hat{\sigma}^2}{\sigma^2}\right\}.$$
   What is the asymptotic behaviour of w_{p} as $n \to \infty$ with m fixed?

6. Consider two models for i.i.d. data $Y_1, \ldots, Y_n$. In model 0, each y_i has density function $f_0(y; \theta_0)$, and in model 1, each y_i has density function $f_1(y; \theta_1)$, where θ_0 and θ_1 are the unknown parameters of each model, which need not be from the same parameter space. We also define a mixture model, in which the density function for each y_i is $f_2(y; \theta_0, \theta_1) = \{f_0(y; \theta_0) + f_1(y; \theta_1)\}/2$. In the mixture model, both θ_0 and θ_1 are unknown.

   Consider the following conjecture:

   > If the statistic $s_0(y_1, \ldots, y_n)$ is sufficient for model 0, and the statistic $s_1(y_1, \ldots, y_n)$ is sufficient for model 1, then the statistic (s_0, s_1) is sufficient for the mixture model.

   Prove that this conjecture is true, or give a counterexample showing that it is false. Be sure to keep in mind the precise definitions of the concepts you use. Your argument should assume only the definition of a sufficient statistic, well-known theorems regarding sufficiency, and basic facts regarding probability.

7. Let the distribution of $X_1, X_2, \ldots$ be defined as follows. Let $X_1 \sim U(-1, 1)$. For $n > 1$, conditional on $X_1, \ldots, X_{n-1}$, let $X_n \sim U(-1, 1)$ with probability $\alpha \,/\, (n-1+\alpha)$, and for each $i \in 1, \ldots, n-1$, let $X_n = X_i$ with probability $(1/2) \,/\, (n-1+\alpha)$ and let $X_n = -X_i$ with probability $(1/2) \,/\, (n-1+\alpha)$.

   (a) Prove that this distribution for $X_1, X_2, \ldots$ is exchangeable, justifying each step in your proof in detail, including steps that are the same as in similar proofs you may have seen or done.

   (b) Since the distribution for $X_1, X_2, \ldots$ is exchangeable, by de Finetti's representation theorem, the $X_1, X_2, \ldots$ are i.i.d. conditional on a distribution, D, which has some prior distribution. Say what you can about the characteristics of a distribution D drawn from this prior distribution over distributions.