May 2013

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

THEORETICAL STATISTICS COMPREHENSIVE EXAMINATION

1. ***Attempt all questions*** (total # of questions = 7).
   It is not necessary to completely solve every problem to achieve a good performance. *Emphasize what you do know.*

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. Probability tables that may be useful are appended at the end of the exam paper, after all questions.

8. Good luck!

1. Suppose that $x = (x_1, \dots, x_n) \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ and $\mu \mid \sigma^2 \sim N(\mu_0, \tau_0^2\sigma^2), 1/\sigma^2 \sim \text{Gamma}(\alpha_0, \beta_0)$ with $\mu_0, \tau_0^2, \alpha_0$ and β_0 fixed.

   (a) Show that $T(x) = (\bar{x}, ||x - \bar{x}1||^2)$ is a minimal sufficient statistic and determine its distribution.

   (b) Determine the posterior distribution of (μ, σ^2).

   (c) Determine the posterior distribution of μ and the form of hpd intervals for μ.

   (d) Determine the posterior predictive distribution of a future $y \sim N(\mu, \sigma^2)$.

   (e) Discuss how $\mu_0, \tau_0^2, \alpha_0$ and β_0 are to be chosen in an application.

2. Let $X_n \sim \text{Multinomial}(n, p_n)$ and $Y_n \sim \text{Multinomial}(c_n n, q_n)$, where X_n and Y_n are independent with the same dimension k, c_n is chosen so that $c_n n$ is an integer with $c_n \to c > 0$ as $n \to \infty$. Assume that $\sqrt{n}(p_n - p) \to 0$ and $\sqrt{n}(q_n - q) \to 0$.

   (a) Derive the asymptotic distribution of a properly scaled form of $\left(\frac{X_n}{n} - \frac{Y_n}{c_n n}\right)$.

   (b) Construct an appropriate statistic W_n in the quadratic form of the quantity in (a), such that W_n asymptotically follows a Chi-sqaure distribution under the null hypothesis that $p = q$. Justify all steps and specify the limiting distribution. You may use the fact on projection matrix.

3. Suppose that $y \sim N_n(X\beta, \sigma^2 I)$ where $X \in R^{n \times k}$ is of rank k and known, while $(\beta, \sigma^2) \in R^k \times (0, \infty)$.

   (a) Determine the MLE $(\hat{\beta}, \hat{\sigma}^2)$ of (β, σ^2) and its joint distribution.

   (b) Show that $\hat{\beta}$ is an optimal unbiased estimator of β and explain what the criterion for optimality is.

   (c) Determine the optimal unbiased estimator of σ^2.

4. Let $X_i \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(1/2)$ and $\{Y_i | X_i = c\} \overset{\text{i.i.d.}}{\sim} \text{Poisson}(c\,\theta)$, and one only can observe Y_i's.

   (a) Find an estimator of θ based on $\bar{Y}_n = \sum_{i=1}^n Y_i$ and its asymptotic distribution.

   (b) Find an asymptotically efficient estimator of θ, and specify the asymptotic distribution of this estimator.

5. Let $g : R^r \to R$ be symmetric in its arguments, $F \in \mathcal{F} = \{G : \int \cdots \int |g^2(y_1, \dots, y_r)| \prod_{i=1}^r dG(y_i) < \infty\}$, consider the V-functional and the corresponding estimator,

$$T(f) = \int \cdots \int g(y_1, \dots, y_r) \prod_{i=1}^r dF(y_i), \quad T(F_n) = n^{-r} \sum_{i_1=1}^n \cdots \sum_{i_r=1}^n g(X_{i_1}, \dots, X_{i_r}),$$

   where $\{X_1, \dots, X_n\}$ is a random sample from F.

(a) Calculate the influence function $T'(x; F)$.

(b) Drive the asymptotic distribution of the properly scaled $T(F_n)$. Note that it is necessary to show the remaining term is asymptotically negligible.

6. Suppose that $X_1, \ldots, X_n$ are independently and identically distributed with the probability density function $f_{\theta,j}$, where $f_{\theta,1}$ is the density of $N(0, \theta^2)$ and $f_{\theta,2} = (2\theta)^{-1}e^{-|x|/\theta}$, $\theta > 0$, $j = 1, 2$.

   (a) Obtain the MLE of (θ, j).

   (b) Is the MLE of j obtained in (a) is consistent? Justify.

   (c) Show that the MLE of θ is consistent and find its asymptotic distribution.

7. Suppose G is a symmetry group of a model $\{f_\theta : \theta \in \Theta\}$ with sample space $\mathcal{X}$.

   (a) If $S : \mathcal{X} \to \mathcal{S}$ is onto and equivariant under G, then show that this implies an action of G on $\mathcal{S}$ given by $T_g^{\mathcal{S}} s = S(T_g x)$ when $S(x) = s$.

   (b) If $\Psi : \Theta \to \Psi$ is onto (so Ψ denotes both the map and its range), define what it means for $\Psi(\theta)$ to be equivariant under G and give the action T_g^* this induces on Ψ, namely, what does $T_g^* \psi$ equal.

   (c) Give the full definition of what it means for a decision problem to be invariant under G.

   (d) Suppose $\{f_{(\mu,\sigma)} : \mu \in R^1, \sigma > 0\}$ is a location-scale family on R^n, $\Psi(\mu, \sigma) = \mu$ and $L((\mu, \sigma), \psi) = (\mu - \psi)^2/\sigma^2$. Show that this decision problem is invariant under the location-scale group.

   (e) For $g \in G$ and $\delta \in \mathcal{D}$, where $\mathcal{D}$ is the set of all decision functions for a problem, show that $T_g^{**}\delta$ defined by $T_g^{**}\delta(x, A) = \delta(T_g^{-1}x, T_g^{*-1}A)$ defines an action on $\mathcal{D}$.

   (f) Prove that a nonrandomized δ is invariant if and only if the corresponding estimator d is equivariant under G.

   (g) Prove that if G leaves the decision problem invariant, then

   $$R(\bar{T}_g\theta, T_g^{**}\delta) = R(\theta, \delta)$$

   for every g, θ and δ. Conclude that, if δ is invariant then δ has constant risk on orbits of G acting on Θ. What do you conclude if G is transitive on Θ? Explain why this transitivity simplifies the problem of obtaining the optimal invariant decision function.

   (h) Suppose that G leaves the decision problem invariant, acts transitively on Θ and acts freely on $\mathcal{X}$. Let $[\cdot]$ be a transformation variable and D be the corresponding maximal invariant. Prove that the optimal invariant estimator d is obtained by minimizing

   $$\int_G L(\theta_0, T_g^* d(D(x))) P_{\theta_0[\cdot]}(dg \mid D)(D(x))$$

for $d(D(x)$ where θ_0 is any fixed value of the full parameter. This estimator is called the *Pitman estimator.*

(i) Obtain the Pitman estimator of μ for the problem described in (d) when $f_{(0,1)}$ is the $N_n(0, I)$ density.