Stat 611: Homework 2

Moulinath Banerjee

University of Michigan

February 3, 2003

(1) (a) Consider a regular parametric model $\{f(x,\theta) : \theta \in \Theta \subset \mathcal{R}^k\}$. Fix $\theta_0 \in \Theta$ and let $I(\theta_0)$ denote the (positive definite) information matrix at the point θ_0. Let $\theta = (\nu, \eta)$ be a partitioning of θ and let $\theta_0 = (\nu_0, \eta_0)$. Here ν is an m–dimensional parameter. Write the information matrix as:

$$I(\theta_0) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}.$$

As defined in class, let

$$I_{11.2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$$

and let $I_{22.1}$ be defined similarly with 2 and 1 swapped. Write,

$$I(\theta_0)^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

and use the fact that $I(\theta_0)\, I(\theta_0)^{-1} = I = I(\theta_0)^{-1}\, I(\theta_0)$ to show that,

$$I(\theta_0)^{-1} = \begin{pmatrix} I_{11.2}^{-1} & -I_{11.2}^{-1}\, I_{12}\, I_{22}^{-1} \\ I_{22.1}^{-1}\, I_{21}\, I_{11.2}^{-1} & I_{22.1}^{-1} \end{pmatrix}.$$

(b) Use (a) to deduce that if $Z_{k \times 1}$ follows a multivariate normal distribution with dispersion matrix $I(\theta_0)$, then

$$Z^T\, I(\theta_0)^{-1}\, Z^T - Z_2^T\, I_{22}^{-1}\, Z_2 \sim \chi_m^2 .$$

Here Z_1 denotes the first m components of Z and Z_2 the remaining $k - m$. What is the distribution of

$$Z^T\, I(\theta_0)^{-1}\, Z^T - Z_1^T\, I_{11}^{-1}\, Z_1\, ?$$

(c) Consider the Rao/Score statistic for testing $H_0 : \eta = \eta_0$. This is

$$R_n \equiv Z_n(\hat{\theta}_n^0)^T\, I^{-1}(\hat{\theta}_n^0)\, Z_n(\hat{\theta}_n^0)$$

where, $\hat{\theta}_n^0 = (\hat{\nu}_n^0, \eta_0)$ is the MLE of θ under the null hypothesis and

$$Z_n(\theta) = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_\nu(X_i, \nu, \eta) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_\eta(X_i, \nu, \eta) \end{pmatrix}.$$

Show that under H_0, R_n has an asymptotic χ^2_{k-m} distribution.

(d) Consider the null hypothesis $H_0 : \nu = \nu_0$. Let $\hat{\theta} = (\hat{\nu}, \hat{\eta})$ and let $\hat{\eta}_0$ be the MLE of η obtained under H_0. Show that, under H_0,

$$\sqrt{n}(\hat{\eta} - \hat{\eta}_0) = -I_{22}^{-1}\, I_{21}\, \sqrt{n}\,(\hat{\nu} - \nu_0) + o_p(1)\,. \qquad (\star)$$

Now consider the likelihood ratio statistic, 2 log λ_n for testing H_0. Show, that under H_0,

$$2\,\log\,\lambda_n = \sqrt{n}(\hat{\theta} - \theta_0)\, I(\theta_0)\, \sqrt{n}(\hat{\theta} - \theta_0) - \sqrt{n}(\hat{\eta}_0 - \eta_0)\, I_{22}\, \sqrt{n}(\hat{\eta}_0 - \eta_0) + o_p(1)\,.$$

Now, using the representation $(\star)$ or otherwise, show that

$$2\,\log\,\lambda_n = n\,(\hat{\nu} - \nu_0)^T\, I_{11.2}\,(\hat{\nu} - \nu_0) + o_p(1)\,.$$

Hence, deduce the asymptotic distribution of the likelihood ratio statistic.

(e) Show that $I_{12}\, I_{22}^{-1}\, \dot{l}_2$ is the closest element to $\dot{l}_1$ in the span of $\dot{l}_2$ in the sense that

$$\text{argmin}_{\alpha}\, E\,(\dot{l}_1(X) - \alpha^T\, \dot{l}_2(X))^2 = I_{12}\, I_{22}^{-1}\,.$$

(2) (a) Let $X_1, X_2, \ldots, X_n$ be a sample from the exponential distribution with parameter θ and let $Y_1, Y_2, \ldots, Y_n$ be a sample from an exponential μ distribution. Also, let the first sample be independent of the second. Consider testing the null hypothesis $\mu = 2\,\theta$. Use an appropriate reparametrization to recast the null hypothesis into the form $\psi = \psi_0$ for some fixed ψ_0, compute the likelihood ratio and Wald statistics and determine their asymptotic distributions under the null.

(b) For $i = 1, 2, \ldots, k$ let $X_{i1}, X_{i2}, \ldots, X_{in}$ be independent samples from Poisson distributions, Poi(θ_i) respectively. Find the likelihood ratio test and its asymptotic distribution for testing $H_0 : \theta_1 = \theta_2 = \ldots = \theta_k$.

(3) Let X follow the Beta(α, β) distribution. Thus,

$$p(\theta, x) = p(\alpha, \beta, x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha - 1}\,(1 - x)^{\beta - 1}\, 1\,(x \in (0, 1))\,.$$

This is a regular parametric model. Let,

$$\psi(\alpha) \equiv \frac{\Gamma^{'}(\alpha)}{\Gamma(\alpha)}\,.$$

(a) Compute the scores for α and β based on one X following Beta(α, β). Hence, compute the information matrix.

(b) What is the information for α if β is known? What if β is unknown? Draw a picture of the scores to illustrate this geometrically.

(c) Now, suppose that we have i.i.d. observations $X_1, X_2, \ldots, X_n$ from a Beta(α_0, β_0) distribution. Let $q(\theta) = \psi(\alpha) - \psi(\beta)$.

(i) Compute the efficient influence function and the information bound for the estimation of q at the point $\theta_0 \equiv (\alpha_0, \beta_0)$.

(ii) Propose a method of moments estimator of the parameter $q(\theta)$; i.e. find some $h(X)$ such that $E_\theta(h(X))$ is $q(\theta)$ and use $n^{-1} \sum_{i=1}^n h(X_i)$ as an estimate of $q(\theta)$. Compute $\mathrm{Var}_\theta(h(X_1))$ and determine the asymptotic distribution of your estimate.

4 **Contiguity in nonparametric models:** Contiguity is a concept that can also be effectively used in nonparametric problems. This exercise deals with contiguity in a nonregular problem. Consider, a regression model of the form,

$$Y = r(X) + \epsilon$$

where X is a random variable taking values in $[0, 1]$ and ϵ is a $N(0, 1)$ random variable. (More generally, X can be allowed to vary in any compact set and ϵ can have arbitrary variance.) Here X and ϵ are independent and X has a continuously differentiable density f. Also, the regression function r is continuously differentiable in $(0, 1)$.

Let P_{r,f,σ^2} denote the distribution of (X, Y) when the regression function is r. Now, fix a point t_0 in $(0, 1)$ and consider a sequence of regression functions of the form,

$$r_n(t) = r(t) + n^{-1/3}\, B_n(n^{1/3}\,(t - t_0))$$

where B_n is a sequence of continuously differentiable functions converging uniformly to a (continuously differentiable) function B on a compact set $[-K, K]$ and vanishing outside of it. Let P_{r_n,f,σ^2} denote the distribution of (X, Y) when the regression function is r_n. By P^n_{r,f,σ^2}, we denote the joint distribution of i.i.d. random vectors $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ when $(X_1, Y_1) \sim P_{r,f,\sigma^2}$; by $P^n_{r_n,f,\sigma^2}$, we denote the joint distribution of i.i.d. random vectors $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ when $(X_1, Y_1) \sim P_{r_n,f,\sigma^2}$. By $A_n(t)$ we denote the function $B_n(n^{1/3}\,(t - t_0))$.

(a) Show, using the outline given below, that the sequence of probability measures $\{P^n_{r_n,f,\sigma^2}\}$ and $\{P^n_{r,f,\sigma^2}\}$ are mutually contiguous.

Outline: Let $L_n(r)$ denote the likelihood function of the data $\{X_i, Y_i\}_{i=1}^n$ under regression function r and $L_n(r_n)$ denote the likelihood function under r_n. Consider the log likelihood ratio λ_n,

$$\lambda_n = \log\, L_n(r_n) - \log\, L_n(r) \equiv \log \frac{d\,P^n_{r_n,f,\sigma^2}}{d\,P^n_{r,f,\sigma^2}}\,.$$

We will deduce the asymptotic distribution of λ_n under $\{P^n_{r,f,\sigma^2}\}$.

Show that,

$$\lambda_n = \frac{1}{2}\left[2\sum_{i=1}^{n} \epsilon_i\,(r_n(X_i) - r(X_i)) - \sum_{i=1}^{n}(r_n(X_i) - r(X_i))^2\right].$$

Before proceeding further, let's introduce some notation. Let $\mathbb{P}_n$ denote the empirical measure of $(X_1, \epsilon_1), (X_2, \epsilon_2), \ldots, (X_n, \epsilon_n)$ i.e. the probability measure that places mass $1/n$ at each point. If we take a function of (X, ϵ), say $g(X, \epsilon)$ then by $\mathbb{P}_n\, g(X, \epsilon)$ we denote the expected value of $g(X, \epsilon)$ under $\mathbb{P}_n$. Thus,

$$\mathbb{P}_n\, g(X, \epsilon) \equiv E_{\mathbb{P}_n}\, g(X, \epsilon) = \frac{1}{n}\sum_{i=1}^{n} g(X_i, \epsilon_i)\,.$$

Also,

$$P\, g(X, \epsilon) \equiv E_{P_{r,f,\sigma^2}}\, g(X, \epsilon)$$

and

$$(\mathbb{P}_n - P)\, g(X, \epsilon) \equiv \mathbb{P}_n\, g(X, \epsilon) - P\, g(X, \epsilon)\,.$$

Starting off from where we left, show,

$$\lambda_n = I_n - II_n$$

where

$$I_n = n^{2/3}\,(\mathbb{P}_n - P)\,(\epsilon\, A_n(X))\,,$$

and

$$II_n = \frac{1}{2}\, n^{1/3}\, \mathbb{P}_n(A_n^2(X))\,.$$

Show that, under P_{r,f,σ^2},

$$II_n \to_p a^2/2 \equiv \frac{1}{2}\int_{-K}^{K} B^2(z)\, dz\;\, f(t_0)\,,$$

by showing that $II_n - E\, II_n$ converges to 0 in probability and studying the limit of $E\, II_n$. Next, note that,

$$I_n = \sqrt{n}(\mathbb{P}_n - P)\, s_n(X, \epsilon)$$

where

$$s_n(X, \epsilon) = n^{1/6}\, \epsilon\, A_n(X)\,.$$

Thus,

$$\lambda_n = \sqrt{n}\,(\mathbb{P}_n - P)\, s_n(X, \epsilon) - \frac{a^2}{2} + o_p(1)\,.$$

Use the Lindeberg–Feller CLT to deduce that,

$$\sqrt{n}\,(\mathbb{P}_n - P)\, s_n(X, \epsilon) \to_d N(0, a^2)$$

under P_{r,f,σ^2}. (A rigorous proof will need verification of the Lindeberg–Feller condition. If you want to be non–rigorous, you can just work out the limiting variance of s_n.) Hence deduce that the two sequences of probability measures are contiguous. Clearly indicate the results that you use.

(b) A key process that needs to be studied in doing likelihood and likelihood ratio based inference in this model, when r is monotone, is

$$\mathbb{M}_n(z) = \sqrt{n}\,(\mathbb{P}_n - P)\, f_{n,z}$$

where

$$f_{n,z} = n^{1/6}\, \frac{1}{f(t_0)} \left[(Y - r(t_0))\, (1\{X \leq t_0 + z\, n^{-1/3}\} - 1\{X \leq t_0\}) \right].$$

Consider a fixed $z > 0$. Compute the limiting distribution (this will be bivariate normal) of

$$(\mathbb{M}_n(z), \log\, L_n(r_n) - \log\, L_n(r))$$

under $\{P^n_{r,f,\sigma^2}\}$. (A rigorous proof will need the Lindeberg–Feller CLT to be used in conjunction with the Cramer–Wold device; if you want to be sketchier, just work out the mean and covariance matrix of the limit distribution) and hence, deduce the limit distribution of $\mathbb{M}_n(z)$ under the sequence $\{P^n_{r_n,f,\sigma^2}\}$.

See the notes on "Basic Large Sample Theory" for the Lindeberg–Feller CLT or a standard probability text (Billingsley, Chung). I will post something on it on the web too.