Statistics 582, Midterm Exam, Solutions

Wellner; 2/15/98

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
   (a) The *Rao or score statistic* for testing a simple null hypothesis $H : \theta = \theta_0$ in a regular parametric model.
   (b) The *Rao or score statistic* for testing a composite null hypothesis $H : \theta_1 = \theta_{10}$ in a regular parametric model. with $\theta = (\theta_1, \theta_2)$, $\theta_1 \in R^m$, $\theta_2 \in R^{k-m}$.
   (c) A *one-step "approximate MLE"* (starting from an $n^{1/4}$-consistent estimator).
   (d) The *Wald statistic* for testing a simple null hypothesis $H : \theta = \theta_0$ in a regular parametric model.
   (e) The *likelihood equations* for estimation of θ in a regular parametric model.
   (f) The *Kullback-Leibler information* $K(P, Q)$ between two probability distributions P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.

   Solution: See the chapter 4 notes.

2. (24 points) **State** *any three* of the following results:
   (a) A theorem about the large sample distribution of the MLE in a regular parametric model with parameter set $\Theta \subset R^k$.
   (b) A theorem about the behavior of the likelihood ratio statistic for testing a simple null hypothesis under a fixed alternative $\theta \neq \theta_0$.
   (c) Any theorem / result about nonparametric nonparametric maximum likelihood estimation.
   (d) A uniform strong law of large numbers (or Glivenko - Cantelli theorem).
   (e) Wald's theorem on strong consistency of maximum likelihood estimators.
   (f) A theorem about the limiting distribution of the likelihood ratio statistic for testing a composite null hypothesis $H : \theta_1 = \theta_{10}$ versus $K : \theta_1 \neq \theta_{10}$ under local alternatives $\theta_{1n} = \theta_{10} + n^{-1/2} t_1$.

   Solution: See the chapter 4 notes.

   Do **either** problem 3 **or** problem 4.

3. (30 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. with mixture density (mass function)
$$p(x; \lambda, \mu, \theta) = \theta \frac{\lambda^x}{x!} e^{-\lambda} + (1 - \theta) \frac{\mu^x}{x!} e^{-\mu}, x = 0, 1, \ldots,$$
   where $0 < \theta < 1$, $0 < \lambda \neq \mu < \infty$; in other words, p is the mixture of two Poisson distributions with parameters λ and μ respectively.

A. Describe an EM - algorithm for estimation of (λ, μ, θ).
B. What is the natural corresponding nonparametric model for the data which were modeled with the parametric mixture distribution in A? What is the natural nonparametric maximum likelihood estimator here?

Solutions: Here it is natural to let the "complete data" $\underline{X}$ be $(X_1, \delta_1), \ldots, (X_n, \delta_n)$ where $\delta_i \in \{0, 1\}$ and (X_i, δ_i) are i.i.d. with density

$$p(x, \delta; \theta, \lambda, \mu) = (\theta \frac{\lambda^x}{x!} e^{-\lambda})^\delta ((1-\theta) \frac{\mu^x}{x!} e^{-\mu})^{1-\delta}$$

for $(x, \delta) \in \{0, 1, \ldots\} \times \{0, 1\}$. Then the incomplete $\underline{Y}$ is $X_1, \ldots, X_n$, which are iid with the mixture distribution

$$p(x; \lambda, \mu, \theta) = \theta \frac{\lambda^x}{x!} e^{-\lambda} + (1-\theta) \frac{\mu^x}{x!} e^{-\mu} .$$

It follows that conditional on $X = x$, δ is Bernoulli$(p(x))$ where

$$p(x) \equiv p(x; \theta, \lambda, \mu) = \frac{\theta \lambda^x e^{-\lambda} / x!}{\theta \frac{\lambda^x}{x!} e^{-\lambda} + (1-\theta) \frac{\mu^x}{x!} e^{-\mu}} . \tag{0.1}$$

Hence $E(\delta | X) = p(X)$; this is the basis of the E - step of an EM algorithm.

To find the M - step, note that

$$\begin{aligned} l(\theta, \lambda, \mu | X, \delta) &= \delta \{\log \theta + X \log \lambda - \lambda\} + (1-\delta)\{\log(1-\theta) + X \log \mu - \mu\} \\ &\qquad + \text{constant} , \end{aligned}$$

so that the scores (for a sample of size one) are

$$\dot{l}_\theta(X, \delta) = \frac{\delta}{\theta} - \frac{1-\delta}{1-\theta} ,$$

$$\dot{l}_\lambda(X, \delta) = \delta \left\{ \frac{X}{\lambda} - 1 \right\} ,$$

$$\dot{l}_\mu(X, \delta) = (1-\delta) \left\{ \frac{X}{\mu} - 1 \right\} .$$

Thus the score equations are solved by

$$\widehat{\lambda}_n = \frac{\sum \delta_i X_i}{\sum \delta_i} , \qquad \widehat{\mu}_n = \frac{\sum (1-\delta_i) X_i}{\sum (1-\delta_i)} , \qquad \widehat{\theta} = \frac{\sum \delta_i}{n} .$$

This is the basis of an M - step.

Set $\theta^{(0)} = 1/2$, $\widehat{\lambda}^{(0)} = \widehat{\mu}^{(0)} = \overline{X}$. Then, for $m = 0, 1, \ldots,$ define

$$\widehat{\delta}_i^{(m)} \equiv p(X_i; \widehat{\theta}^{(m)}, \widehat{\lambda}^{(m)}, \widehat{\mu}^{(m)}) \tag{0.2}$$

where $p(x; \theta, \lambda, \mu)$ is given by (0.1), and

$$\widehat{\lambda}^{(m+1)} = \frac{\sum \widehat{\delta}_i^{(m)} X_i}{\sum \widehat{\delta}_i^{(m)}}, \tag{0.3}$$

$$\widehat{\mu}^{(m+1)} = \frac{\sum (1 - \widehat{\delta}_i^{(m)}) X_i}{\sum (1 - \widehat{\delta}_i^{(m)})}, \tag{0.4}$$

$$\widehat{\theta}^{(m+1)} = \frac{\sum \widehat{\delta}_i^{(m)}}{n}. \tag{0.5}$$

Iteration of (0.2) and (0.3,0.4,0.5) yields an EM algorithm for estimation of (θ, λ, μ).

B. The natural nonparametric model for this data would be $\mathcal{P} = \{\underline{p} = (p_0, p_1, p_2, \ldots) : \sum_{x=0}^{\infty} p_x = 1\}$. The nonparametric maximum likelihood estimator is just $\mathbb{P}_n$ where

$$\mathbb{P}_n(\{x\}) = \frac{\#\{i \le n : \ X_i = x\}}{n}.$$

4. (30 points) Suppose that (X_i, Y_i), $i = 1, \ldots, n$ are independent pairs of random variables with

$$X_i \sim \text{exponential}(\beta_i/\alpha), \qquad Y_i \sim \text{exponential}(1/\beta_i \alpha)$$

independent. Here $\alpha > 0$ and $\beta_i > 0$ for $i = 1, \ldots, n$ are all unknown. Thus the joint density of (X_i, Y_i) is

$$f_{X_i, Y_i}(x_i, y_i) = \alpha^{-2} \exp(-\beta_i x/\alpha) \exp(-y_i/\alpha\beta_i) 1_{[0,\infty)}(x_i) 1_{[0,\infty)}(y_i).$$

A. Find the maximum likelihood estimator $\widehat{\alpha}$ of α.

B. Do our theorems about consistency and asymptotic normality of maximum likelihood estimators apply to $\widehat{\alpha}$? Why or why not? To what (famous) model is the above model analogous?

Solution: A. The likelihood is

$$L_n(\alpha, \underline{\beta}) = \alpha^{-2n} \exp(-\frac{1}{\alpha}\{\sum_{i=1}^n \beta_i X_i + \sum_{i=1}^n \beta_i^{-1} Y_i\}),$$

so

$$l(\alpha, \underline{\beta}) = -2n \log \alpha - \frac{1}{\alpha}\{\sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \beta_i^{-1} Y_i\} ,$$

and

$$\dot{l}_{\alpha}(\alpha, \beta) = \frac{\partial}{\partial \alpha} l(\alpha, \underline{\beta}) = -\frac{2n}{\alpha} + \frac{1}{\alpha^2}\{\sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \beta_i^{-1} Y_i\}$$

and

$$\dot{l}_{\beta_i}(\alpha, \beta) = \frac{\partial}{\partial \beta_i} l(\alpha, \underline{\beta}) = -\frac{1}{\alpha}(X_i - \frac{1}{\beta_i^2} Y_i) ,$$

so $\widehat{\beta}_i = \sqrt{Y_i / X_i}$, $i = 1, \ldots, n$ and

$$\hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \sqrt{X_i Y_i}$$

B. No, since the model involves $(n+1)$ parameters – which increases with the number of observations. This model is a close relative of the famous "Neyman - Scott" example, in which the MLE of σ^2 is nonconsistent.

5. (45 points) Suppose that $X_i = (Y_i, Z_i)$, $i = 1, \ldots, n$ are i.i.d. with $(Y|Z = z) \sim Poisson(\alpha_0 \exp(\beta_0 z))$ where $\theta \equiv (\alpha, \beta) \in R^+ \times R$; 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 $c < \infty$.
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. Find the score statistic for testing $H : \beta = 0$ (and $\alpha =$ anything) versus $K : \beta \neq 0$. What is its asymptotic distribution under H?
D. If β is fixed, show that the likelihood is maximized as a function of α by

$$\hat{\alpha}(\beta) = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} \exp(\beta Z_i)} .$$

Use this to compute the profile log-likelihood function

$$l_n^{prof}(\beta) = l_n(\hat{\alpha}(\beta), \beta) .$$

E. Suppose that an ad hoc consistent estimator $\overline{\beta}_n$ of β_0 is available: $\overline{\beta}_n \to_{a.s.} \beta_0$. Prove that

$$\hat{\alpha}(\overline{\beta}) \to_{a.s.} \alpha_0 .$$

Solution: A. If H has density h, then the joint density for one observation is, with $\lambda_\theta(z) = \alpha \exp(\beta z)$,

$$p(y, z; \theta) = \frac{\lambda_\theta(z)^y}{y!} \exp(-\lambda_\theta(z)) h(z) ,$$

for $x = 0, 1, \ldots$. Hence

$$\begin{aligned} \log p(y, z; \theta) &= y \log \lambda_\theta(z) - \lambda_\theta(z) - \log(y!) \\ &= y \log \alpha + y z \beta - \alpha \exp(\beta z) , \end{aligned}$$

and it follows that the scores for α and β for one observation are given by

$$\dot{l}_\alpha(y, z) = \frac{y}{\alpha} - \exp(\beta z) ,$$

$$\dot{l}_\beta(y, z) = yz - \alpha \exp(\beta z) z = z(y - \alpha \exp(\beta z)) .$$

B. From A we find that

$$I_{\alpha\alpha} = E_\theta(\dot{l}_\alpha^2(Y, Z)) = \frac{1}{\alpha^2} E\{E(Y - \lambda_\theta(Z))^2 | Z\} = \frac{1}{\alpha^2} \alpha E(\exp(\beta Z)) ,$$

$$I_{\beta\beta} = E_\theta(\dot{l}_\beta^2(Y, Z)) = E\left\{ Z^2 E\left((Y - \lambda_\theta(Z))^2 | Z \right) \right\} = \alpha E(Z^2 \exp(\beta Z)) ,$$

and

$$I_{\alpha\beta} = E_\theta(\dot{l}_\alpha \dot{l}_\beta(Y, Z)) = \frac{1}{\alpha} E\left\{ Z E\left((Y - \lambda_\theta(Z))^2 | Z \right) \right\} = E(Z \exp(\beta Z)) .$$

[Alternatively,

$$\ddot{l}_{\alpha\alpha}(x) = -\frac{x}{\alpha^2} , \quad \ddot{l}_{\alpha\beta}(x) = -z \exp(\beta z) , \quad \ddot{l}_{\beta\beta}(x) = -\alpha z^2 \exp(\beta z) ,$$

and hence

$$I_{\alpha\alpha} = -E_\theta(\ddot{l}_{\alpha\alpha}(X)) = E_\theta \frac{X}{\alpha^2} = \frac{1}{\alpha} E \exp(\beta Z) ,$$

$$I_{\alpha\beta} = -E_\theta(\ddot{l}_{\alpha\beta}(X)) = E_\theta Z \exp(\beta Z) ,$$

and

$$I_{\alpha\beta} = -E_\theta(\ddot{l}_{\beta\beta}(X)) = E_\theta Z^2 \exp(\beta Z)$$

in agreement with the preceding calculations.] Under $\beta = 0$ we easily find that $\hat{\alpha}^0 = \overline{Y}_n$, and $\hat{\theta}^0 = (\hat{\alpha}^0, \hat{\beta}^0) = (\overline{Y}_n, 0)$, while the normalized scores are

$$\underline{Z_n}(\theta) = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^n (\frac{Y_i}{\alpha} - \exp(\beta Z_i)) \\ \sum_{i=1}^n Z_i (Y_i - \alpha \exp(\beta Z_i)) \end{pmatrix} .$$

This yields

$$\underline{Z_n}(\hat{\theta}^0) = \frac{1}{\sqrt{n}} \begin{pmatrix} 0 \\ \sum_{i=1}^n Z_i(Y_i - \overline{Y}_n)) \end{pmatrix} .$$

Now

$$\begin{aligned} I_{22\cdot 1}(\theta) &= I_{22} - I_{21} I_{11}^{-1} I_{12} \\ &= \alpha E(Z^2 \exp(\beta Z) - \frac{\alpha}{E(\exp(\beta Z))} [E(Z \exp(\beta Z))]^2 \end{aligned}$$

so we have

$$\hat{I}_{22\cdot 1}(\hat{\theta}^0) = \overline{Y} \left\{ E(Z^2) - [E(Z)]^2 \right\} = \overline{Y} Var(Z) .$$

It follows that the score statistic for testing $\beta = 0$ is given by

$$R_n = \frac{\left\{ n^{-1/2} \sum_{i=1}^n Z_i(Y_i - \overline{Y}_n) \right\}^2}{\overline{Y} Var(Z)} .$$

Under the null hypothesis we have $R_n \to_d \chi_1^2$.

D. The score equation for α is given by

$$0 = \sum_{i=1}^n \dot{l}_\alpha(Y_i, Z_i) = \sum_{i=1}^n \left(\frac{Y_i}{\alpha} - \exp(\beta Z_i) \right) = \frac{1}{\alpha} \sum_{i=1}^n Y_i - \sum_{i=1}^n \exp(\beta Z_i) .$$

For each fixed β this has the stated solution

$$\widehat{\alpha}(\beta) = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n \exp(\beta Z_i)} . \tag{0.6}$$

Now the log-likelihood is given by

$$l_n(\alpha, \beta) = \sum_{i=1}^n Y_i \log \alpha + \beta \sum_{i=1}^n Y_i Z_i - \alpha \sum_{i=1}^n \exp(\beta Z_i) - \sum_{i=1}^n \log(Y_i!)$$

When we plug the rightside of (0.6) back into the loglikelihood $l_n(\alpha, \beta)$ we find that the profile log-likelihood is given by

$$l_n(\widehat{\alpha}(\beta), \beta) = \sum_{i=1}^n Y_i \left\{ \log \left(\frac{\sum Y_i}{\sum \exp(\beta Z_i)} \right) - 1 \right\} + \beta \sum_{i=1}^n Y_i Z_i .$$

E. Note that

$$\hat{\alpha}(\overline{\beta}) = \frac{n^{-1} \sum_{i=1}^n Y_i}{n^{-1} \sum_{i=1}^n \exp(\overline{\beta} Z_i)}$$

is a continuous function $g(u,v) = u/v$ of $(\overline{Y}_n, n^{-1}\sum_{i=1}^n \exp(\overline{\beta} Z_i))$ and we know that

$$(\overline{Y}_n, n^{-1}\sum_{i=1}^n \exp(\beta_0 Z_i)) \rightarrow_{a.s.} (E(Y), E(\exp(\beta_0 Z))) = (\alpha_0 E(\exp(\beta_0 Z)), E(\exp(\beta_0 Z)))$$

which clearly has $g(\alpha_0 E(\exp(\beta_0 Z)), E(\exp(\beta_0 Z))) = \alpha_0$. Thus it suffices to show that

$$n^{-1}\sum_{i=1}^n \exp(\overline{\beta} Z_i) - n^{-1}\sum_{i=1}^n \exp(\beta_0 Z_i) \rightarrow_{a.s.} 0\,, \tag{0.7}$$

or, since $h(\beta) \equiv E_0 \exp(\beta Z)$ is continuous in β, that

$$n^{-1}\sum_{i=1}^n \left(\exp(\overline{\beta} Z_i) - g(\overline{\beta})\right) \rightarrow_{a.s.} 0\,. \tag{0.8}$$

It suffices to prove *either* (0.7) or (0.8). First a proof of (0.7) by straightforward Taylor expansion: Set

$$D_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^n e^{\beta Z_i}\,.$$

Then

$$D_n(\overline{\beta}) - D_n(\beta_0) = D_n'(\beta_n^*)(\overline{\beta}_n - \beta_0)$$

where $|\beta_n^* - \beta_0| \le |\overline{\beta_n} - \beta_0| \rightarrow_{a.s.} 0$ and

$$D_n'(\beta) = \frac{1}{n}\sum_{i=1}^n Z_i e^{\beta Z_i}\,.$$

Let $\delta > 0$. Then $|\beta_n^* - \beta_0| \le \delta$ a.s. for $n \ge N_\omega$. Since $|Z| \le c$, for $n \ge N_\omega$ it follows that

$$|D_n'(\beta_n^*)| \le \frac{1}{n}\sum_{i=1}^n |Z_i| \exp((|\beta_0| + \delta)|Z_i|) \le c \exp((|\beta_0| + \delta)c) \equiv C\,.$$

Therefore we have

$$|D_n(\overline{\beta}) - D_n(\beta_0)| \le |D_n'(\beta_n^*)||\overline{\beta}_n - \beta_0| \le C|\overline{\beta}_n - \beta_0| \rightarrow_{a.s.} 0\,.$$

Here is a proof of (0.8) via a uniform strong law of large numbers (Le Cam's Glivenko-Cantelli theorem). Let $\delta > 0$ and consider the collection of continuous functions $\mathcal{F} = \{f(z,\beta) = \exp(\beta z) : |\beta - \beta_0| \le \delta\}$ for some $\delta > 0$: this collection is clearly continuous, the set $[\beta_0 - \delta, \beta_0 + \delta]$ is compact, and $|f(z,\beta)| \le \exp(c(|\beta_0| + \delta))$ which is integrable. Hence

$$\sup_{\beta : |\beta - \beta_0| \le \delta} |n^{-1}\sum_{i=1}^n (\exp(\beta Z_i) - E_0 \exp(\beta Z))\,| \rightarrow_{a.s.} 0\,.$$

Since $\overline{\beta}_n \in [\beta_0 - \delta, \beta_0 + \delta]$ for $n \geq N_\omega$,

$$\begin{aligned}
&\Big| n^{-1} \sum_{i=1}^{n} \Big(\exp(\overline{\beta} Z_i) - g(\overline{\beta}) \Big) \Big| \\
&\quad \leq \sup_{\beta : |\beta - \beta_0| \leq \delta} | n^{-1} \sum_{i=1}^{n} \left(\exp(\beta Z_i) - E_0 \exp(\beta Z) \right) | \rightarrow_{a.s.} 0 \, .
\end{aligned}$$