Statistics 581, Problem Set 9 Solutions

Wellner; 11/29/2000

1. (a) Lehmann and Casella, problem 6.3.1, page 501: Let X have the binomial distribution $Bin(n,p)$, $0 \le p \le 1$. Determine the MLE of p:
   (i) by the usual calculus method determining the maximum of a function.
   (ii) by showing that $p^x q^{n-x} \le (x/n)^x [(n-x)/n]^{n-x}$.

   (b) Lehmann and Casella, problem 6.3.2, page 501: In the preceding problem, show that the MLE does not exist when p is restricted to $0 < p < 1$ and when $X = 0$ or $X = 1$.
   (c) Lehmann and Casella, problem 6.3.5, page 501: Here is a (slight) rephrasing of the problem: Let $X \sim$ Bernoulli(p): $P(X = 0) = 1 - p = q$, $P(X = 1) = p$. Suppose that it is known that $1/3 \le p \le 2/3$.
   (i) Find the MLE;
   (ii) Show that the expected squared error of the MLE is uniformly larger than that of $\delta(X) = 1/2$.

   Solution: (a)(i) Since $\log P_p(X = x) = x \log p + (n-x)\log(1-p)$, we have $l(p|X) = X \log p + (n - X)\log(1 - p)$; differentiating this with respect to p yields

   $$\dot{l}(p|X) = \frac{X}{p} - \frac{n-X}{1-p} = \frac{X(1-p) - (n-X)p}{p(1-p)}$$

   and this equals 0 if $p = \hat{p} \equiv X/n$. Since the second derivative is

   $$\ddot{l}(p|X) = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2} < 0$$

   it follows that $\hat{p} = X/n$ is the MLE of $p \in [0,1]$.
   (a)(ii) Since $(\prod_{i=1}^n y_i)^{1/n} \le n^{-1}(y_1 + \cdots + y_n)$ for any numbers $y_i \ge 0$, it follows, with $y_i \equiv np/X$ for $i = 1, \ldots, X$, and $y_i \equiv nq/(n - X)$, $i = X + 1, \ldots, n$, that

   $$\left\{ \left(\frac{np}{X}\right)^X \left(\frac{nq}{n-X}\right)^{n-X} \right\}^{1/n} \le n^{-1} \left\{ X\frac{np}{X} + (n-X)\frac{nq}{n-X} \right\} = 1\,,$$

   or, equivalently,

   $$p^X (1-p)^{n-X} \le \left(\frac{X}{n}\right)^X \left(\frac{n-X}{n}\right)^{n-X},$$

   with equality if and only if $p = X/n \equiv \hat{p}$. Thus $\hat{p} = X/n$ is the MLE of $p \in [0,1]$.
   (b) When the closed interval $[0,1]$ is replaced by the open interval $(0,1)$, then the MLE exists if $0 < X < n$ and is $\hat{p} = X/n \in (0,1)$ in this case. If $X = 0$, then the log-likelihood equals $n\log(1-p)$, so $\sup_{p\in(0,1)} l(p) = 0$, but this supremum is not achieved (in the set $(0,1)$). Thus the MLE does not exist in this case. Similarly, if $X = n$, the the log-likelihood equals $n \log p$, so $\sup_{p\in(0,1)} l(p) = 0$, but this supremum is not achieved (in the set $(0,1)$).

(c)(i) For the more general case in which $X \sim \text{Binomial}(n, p)$ From (a), the MLE $\hat{p}$ of $p \in [1/3, 2/3]$ is

$$\hat{p} = \begin{cases} X/n, & \text{if } X/n \in [1/3, 2/3], \\ 1/3, & \text{if } X/n < 1/3, \\ 2/3, & \text{if } X/n > 2/3\,. \end{cases}$$

For $n = 1$, this implies that the MLE is $1/3$ if $X = 0$ and $2/3$ if $X = 1$.
(ii) Now the estimator $\delta(X) = 1/2$ has expected squared error

$$R_1(p) \equiv E_p(\delta(X) - p)^2 = (1/2 - p)^2, \qquad 1/3 \le p \le 2/3\,.$$

On the other hand the MLE $\hat{p}$ has expected squared error

$$\begin{aligned} R_2(p) \equiv E_p(\hat{p} - p)^2 &= p(2/3 - p)^2 + q(1/3 - p)^2 \\ &> (1/2 - p)^2 = R_1(p) \end{aligned}$$

by noting that $R_2(1/3) = 1/3^3 > 1/6^2 = R_1(1/3)$, and $R_2(1/2) = (1/2)(1/6)^2 + (1/2)(1/6)^2 = 1/6^2 > 0 = R_1(1/2)$. See the following Figure 1 for a comparison of these two mean-square errors.

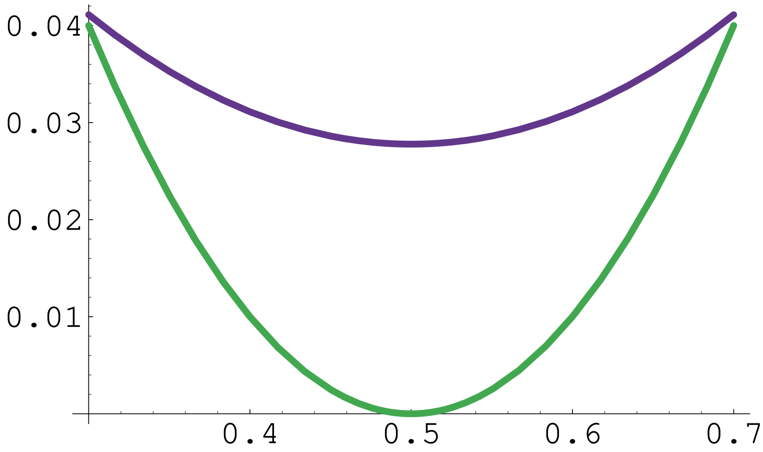

Figure 1: Mean-Square Errors.

2. Suppose that $(Y|Z) \sim \text{Poisson}(\lambda e^{\gamma Z})$, and $Z \sim \mathit{Bernoulli}(\eta)$, and $\theta = (\lambda, \gamma, \eta)$. Let $X = (Y, Z)$, and suppose that we observe $X_1, \ldots, X_n$ i.i.d. as X.
(a) Find the score equations for estimation of θ.
(b) Give conditions on the data $X_1, \ldots, X_n = (Y_1, Z_1), \ldots, (Y_n, Z_n)$ guaranteeing that the score equations have a unique solution which maximizes the likelihood. Call the resulting estimators $\widehat{\theta}_n = (\widehat{\lambda}_n, \widehat{\gamma}_n, \widehat{\eta}_n)$.
(c) What does theorem 4.1.5 (Chapter 4, page 4), say about the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$ when the distribution of the data is given by P_{θ_0}.
(d) Suppose that $\theta_1 \neq \theta_0$ is the "true" value of the parameter θ, and we consider the likelihood ratio $L_n(\theta_1)/L_n(\theta_0)$ where $L_n(\theta) \equiv \prod_{i=1}^n p_\theta(X_i)$. Show that

$n^{-1}\log(L_n(\theta_1)/L_n(\theta_0)) \to_p$ some constant, and identify the constant explicitly in terms of θ_1, θ_0.

Solution: (a) From the solution of problem 1, problem set #8,

$$\begin{aligned}
\dot{l}_\lambda(y,z) &= \frac{y}{\lambda} - e^{\gamma z} = \frac{1}{\lambda}(y-\lambda e^{\gamma z}),\\
\dot{l}_\gamma(y,z) &= yz - \lambda e^{\gamma z} z = z(y-\lambda e^{\gamma z}),\\
\dot{l}_\eta(y,z) &= \frac{z}{\eta} - \frac{1-z}{1-\eta}.
\end{aligned}$$

Thus the score equations for $\theta = (\lambda, \gamma, \eta)$ are

$$\begin{aligned}
0 &= \sum_{i=1}^n \dot{l}_\lambda(Y_i,Z_i) = \frac{1}{\lambda}\sum_{i=1}^n (Y_i - \lambda e^{\gamma Z_i}),\\
0 &= \sum_{i=1}^n \dot{l}_\gamma(Y_i,Z_i) = \sum_{i=1}^n Z_i(Y_i - \lambda e^{\gamma Z_i}),\\
0 &= \sum_{i=1}^n \dot{l}_\eta(Y_i,Z_i) = \sum_{i=1}^n \left\{\frac{Z_i}{\eta} - \frac{1-Z_i}{1-\eta}\right\} = (\eta(1-\eta))^{-1}\sum_{i=1}^n (Z_i - \eta).
\end{aligned}$$

The third equation always has the unique solution $\hat{\eta} = \overline{Z}_n$. For the first and second equations, we compute

$$\begin{aligned}
\ddot{l}_{n,\lambda\lambda} &= -\frac{1}{\lambda^2}\sum_{i=1}^n Y_i,\\
\ddot{l}_{n,\lambda\gamma} &= -\sum_{i=1}^n Z_i e^{\gamma Z_i} = \ddot{l}_{n,\gamma\lambda},\\
\ddot{l}_{n,\gamma\gamma} &= -\lambda \sum_{i=1}^n Z_i^2 e^{\gamma Z_i}.
\end{aligned}$$

Thus the matrix of second partial derivatives (the Hessian) fails to be is negative definite if and only if

$$\left(\lambda \sum_{i=1}^n Z_i^2 e^{\gamma Z_i}\right)\left(\lambda^{-2}\sum_{i=1}^n Y_i\right) = \left(\sum_{i=1}^n Z_i e^{\gamma Z_i}\right)^2.$$

This occurs if:
A. $Z_1 = \ldots = Z_n = 0$ (in which case both sides are 0, the score equation for λ has the solution $\hat{\lambda} = \overline{Y}_n$, and the score equation for γ is identically 0 so we can't estimate γ); or,
B. $Z_1 = \ldots = Z_n = 1$, in which case equality holds if

$$\lambda^{-1} n e^{\gamma} n \overline{Y}_n = n^2 e^{2\gamma},$$

or, equivalently, if $\overline{Y}_n = \lambda e^{\gamma}$. (Note that in this latter case the score equations for λ and γ are identical, and indeed both have the solution $\overline{Y}_n = \lambda e^{\gamma}$ in this case: all

we can estimate is the *product* λe^{γ}.
Thus the likelihood equations have a unique solution if $Z_i \ne Z_j$ for some $i \ne j$.

(b) Theorem 4.1.5 says that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N_3(0, I(\theta_0)^{-1})$$

where

$$I(\theta) = \begin{pmatrix} \lambda^{-1}E(e^{\gamma Z}) & E(Ze^{\gamma Z}) & 0 \\ E(Ze^{\gamma Z}) & \lambda E(Z^2 e^{\gamma Z}) & 0 \\ 0 & 0 & (\eta(1-\eta))^{-1} \end{pmatrix} .$$

When θ_1 is true,

$$\begin{aligned} n^{-1}\log\frac{L_n(\theta_1)}{L_n(\theta_0)} &= n^{-1}\sum_{i=1}^{n}\log\frac{p_{\theta_1}}{p_{\theta_0}}(X_i) \\ &\to_p E_{\theta_1}\log\frac{p_{\theta_1}}{p_{\theta_0}}(X_1) \\ &= K(P_{\theta_1}, P_{\theta_0}) . \end{aligned}$$

Here

$$\begin{aligned} \log\frac{p_{\theta_1}}{p_{\theta_0}}(x) &= y\log\left(\frac{\lambda_1 e^{\gamma_1 z}}{\lambda_0 e^{\gamma_0 z}}\right) - \lambda_1 e^{\gamma_1 z} + \lambda_0 e^{\gamma_0 z} \\ &\quad + z\log\frac{\eta_1}{\eta_0} + (1-z)\log\frac{1-\eta_1}{1-\eta_0} , \end{aligned}$$

so

$$\begin{aligned} K(P_{\theta_1}, P_{\theta_0}) &= E_{\theta_1}\log\frac{p_{\theta_1}}{p_{\theta_0}}(X_1) \\ &= E_{\theta_1}\left(Y\log\frac{\lambda_1 e^{\gamma_1 Z}}{\lambda_0 e^{\gamma_0 Z}}\right) + \eta_1(\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1-\eta_1)(\lambda_0 - \lambda_1) \\ &\quad + \eta_1\log\frac{\eta_1}{\eta_0} + (1-\eta_1)\log\frac{1-\eta_1}{1-\eta_0} \\ &= E_{\theta_1}\left(\lambda_1 e^{\gamma_1 Z}\log\left(\frac{\lambda_1 e^{\gamma_1 Z}}{\lambda_0 e^{\gamma_0 Z}}\right)\right) + \eta_1(\lambda_0 e^{\gamma_0} - \lambda_1 e^{\gamma_1}) + (1-\eta_1)(\lambda_0 - \lambda_1) \\ &\quad + \eta_1\log\frac{\eta_1}{\eta_0} + (1-\eta_1)\log\frac{1-\eta_1}{1-\eta_0} . \end{aligned}$$

3. For the same set-up as in problem 2, consider taking a "profile likelihood" approach to the estimation of γ as follows:
(a) Let $l_n(\theta) = l_n(\gamma, \lambda, \eta)$: consider first maximizing this as a function of λ and η for each fixed value of γ to find

$$(\widehat{\lambda}(\gamma), \widehat{\eta}(\gamma)) \equiv \text{argmax}_{(\lambda,\eta)} l_n(\lambda, \gamma, \eta) .$$

Compute the maximizer $(\widehat{\lambda}(\gamma), \widehat{\eta}(\gamma))$ as explicitly as possible, and then form the "profile log-likelihood" $l_n^{profile}(\gamma)$ defined by

$$l_n^{profile}(\gamma) \equiv l_n(\widehat{\lambda}(\gamma), \gamma, \widehat{\eta}(\gamma)) .$$

(b) Now maximize $l_n^{profile}(\gamma)$ with respect to γ. Find the resulting "profile likelihood" score equation for γ.
(c) Does the equation you derived in (b) follow from the original score equations?
(d) Does the "profile score function" which appears in (b) correspond to or relate to the efficient score for γ in any way?

Solution: (a) The value of γ doesn't influence the maximization with respect to η and we find $\hat{\eta}(\gamma) = \hat{\eta} = \overline{Z}_n$ for all γ. Solving the score equation for λ for a fixed γ yields

$$\sum_1^n Y_i = \lambda \sum_1^n e^{\gamma Z_i}, \qquad \text{or} \qquad \hat{\lambda}(\gamma) = \frac{\sum_1^n Y_i}{\sum_1^n e^{\gamma Z_i}} .$$

Substitution of these into the log-likelihood yield the profile log-likelihood

$$\begin{aligned} l_n^{profile}(\gamma) &= l_n(\hat{\lambda}(\gamma), \gamma, \hat{\eta}) \\ &= \sum_{i=1}^n \left\{ Y_i \log(\hat{\lambda}(\gamma) e^{\gamma Z_i}) - \hat{\lambda}(\gamma) e^{\gamma Z_i} - \log(Y_i!) \right\} \\ &\qquad + n\overline{Z} \log \overline{Z} + (n - n\overline{Z}) \log(1 - \overline{Z}) \\ &= n\overline{Y} \log \hat{\lambda}(\gamma) + \gamma \sum_{i=1}^n Y_i Z_i - \hat{\lambda}(\gamma) \sum_{i=1}^n e^{\gamma Z_i} + \text{constant in } \gamma . \end{aligned}$$

(b) Differentiating the profile log-likelihood with respect to γ yields

$$\dot{l}_{n,\gamma}^{profile}(\gamma) = \frac{d}{d\gamma} \log \hat{\lambda}(\gamma) n\overline{Y}_n + \sum_{i=1}^n Y_i Z_i = \sum_{i=1}^n Y_i Z_i - n\overline{Y} \frac{\sum_1^n Z_i e^{\gamma Z_i}}{\sum_1^n e^{\gamma Z_i}}$$

since

$$\frac{d}{d\gamma} \log \hat{\lambda}(\gamma) = -\frac{\sum_1^n Z_i e^{\gamma Z_i}}{\sum_1^n e^{\gamma Z_i}} .$$

Thus the profile score equation for γ becomes: $\hat{\gamma}^{profile} = \hat{\gamma}$ satisfies

$$\frac{\sum_1^n Y_i Z_i}{\sum_1^n Y_i} = \frac{\sum_1^n Z_i e^{\hat{\gamma} Z_i}}{\sum_1^n e^{\hat{\gamma} Z_i}} . \tag{0.1}$$

(c) If we solve the original score equation for λ for fixed γ, then we obtain

$$\hat{\lambda}(\gamma) = \frac{\sum_1^n Y_i}{\sum_1^n e^{\gamma Z_i}}$$

as in (a). Substitution of this into the score equation for γ yields

$$\begin{aligned} 0 &= \sum_{i=1}^n Z_i Y_i - \hat{\lambda}(\gamma) \sum_{i=1}^n Z_i e^{\gamma Z_i} \\ &= \sum_{i=1}^n Z_i Y_i - \frac{\sum_1^n Y_i}{\sum_1^n e^{\gamma Z_i}} \sum_{i=1}^n Z_i e^{\gamma Z_i} , \end{aligned}$$

and this implies that (0.1) holds.
(d) To see the connection between the profile score function for γ and the efficient score function for γ, note that

$$\dot{l}_n^{profile}(\gamma) \;=\; \dot{l}_{n,\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\frac{d}{d\gamma}\hat{\alpha}(\gamma)+\dot{l}_{n,\gamma}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\,, \tag{0.2}$$

and, since $0=\dot{l}_{n,\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})$, by differentiating with respect to γ we have

$$0=\ddot{l}_{n,\alpha\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\frac{d}{d\gamma}\hat{\alpha}(\gamma)+\ddot{l}_{n,\gamma\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\,,$$

and hence

$$\frac{d}{d\gamma}\hat{\alpha}(\gamma)=-\left(\ddot{l}_{n,\alpha\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\right)^{-1}\ddot{l}_{n,\gamma\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\,. \tag{0.3}$$

Substitution of (0.3) into (0.2) yields

$$\begin{aligned}
\dot{l}_n^{profile}(\gamma) &= \dot{l}_{n,\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\frac{d}{d\gamma}\hat{\alpha}(\gamma)+\dot{l}_{n,\gamma}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\\
&= \dot{l}_{n,\gamma}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})-\ddot{l}_{n,\gamma\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\left(\ddot{l}_{n,\alpha\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\right)^{-1}\dot{l}_{n,\alpha}(\hat{\alpha}(\gamma),\gamma,\hat{\eta})\\
&= \sum_{i=1}^{n}\left\{\dot{l}_{\gamma}(X_i)-\hat{I}_{\alpha\gamma}\hat{I}_{\gamma\gamma}^{-1}\dot{l}_{\alpha}(X_i)\right\}\Big|_{\theta=(\hat{\alpha}(\gamma),\gamma,\hat{\eta})}\,.
\end{aligned}$$