Statistics 581, Problem Set 9 Solutions

Wellner; 11/29/2018

1. Suppose that $X_1, \ldots, X_n$ are i.i.d. Geometric(θ) random variables; that is, $P_\theta(X_1 = k) = \theta(1-\theta)^{k-1}$ for $k = 1, 2, \ldots$ where $\theta \in (0,1)$.
(i) Show that the Geometric distribution with parameter θ satisfies the conditions of Lemma 7.6 of van der Vaart (1998), page 95.
(ii) Compute the information for θ.
(iii) Suggest three different estimators of θ based on the data.
(iv) Which of your estimators are asymptotically efficient in the sense of Hájek's convolution theorem?

Solution: (i) $s_\theta(x) = \sqrt{p_\theta(x)}$ for $\theta \in (0,1)$ and $x \in \mathbb{N}$, so

$$\dot{s}_\theta(x) = \frac{1}{2} p_\theta(x)^{-1/2} \left\{ (1-\theta)^{x-1} - \theta(x_1)(1-\theta)^{x-2} \right\} = \frac{1}{2} p_\theta(x)^{-1/2} \dot{p}_\theta(x),$$

which is continuous in θ for each $x \in \mathbb{N}$. Furthermore,

$$\dot{p}_\theta(x) = p_\theta(x) \left\{ \frac{1}{\theta} - \frac{x-1}{1-\theta} \right\}$$

and hence

$$\begin{aligned} I(\theta) &= E_\theta \left(\frac{\dot{p}_\theta}{p_\theta}(X) \right)^2 \\ &= E_\theta \left(\frac{1}{\theta^2(1-\theta)^2} (1 - \theta - (\theta X - \theta))^2 \right) \\ &= E_\theta \left(\frac{1}{\theta^2(1-\theta)^2} (1 - \theta X)^2 \right) \\ &= \frac{1}{(1-\theta)^2} E_\theta \left(\frac{1}{\theta} - X \right)^2 = \frac{1}{\theta^2(1-\theta)}. \end{aligned}$$

It follows that the hypotheses of Lemma 7.6 of van der Vaart (1998) page 95 holds and we conclude that $\{p_\theta\}$ is differentiable in quadratic mean.

(ii) $\log p_\theta(X) = (X-1)\log(1-\theta) + \log\theta$, so the score function is

$$\begin{aligned} \dot{\mathbf{l}}_\theta(X) &= -\frac{X-1}{1-\theta} + \frac{1}{\theta} = \frac{1}{\theta(1-\theta)}(1-\theta X), \quad \text{and} \\ \ddot{\mathbf{l}}_{\theta\theta}(X) &= -\frac{X-1}{(1-\theta)^2} - \frac{1}{\theta^2}. \end{aligned}$$

Thus the information for θ is, since $E_\theta X = 1/\theta$,

$$I(\theta) = \frac{E_\theta(X) - 1}{(1-\theta)^2} + \frac{1}{\theta^2} = \frac{1}{\theta^2(1-\theta)}.$$

(iii) Since $E_\theta X = 1/\theta$, a method of moments estimator of θ is simply $\widehat{\theta}_n^{(1)} = 1/\overline{X}_n$. Since $P_\theta(X=1) = \theta$, another method of moments estimator of θ is

$$\widehat{\theta}_n^{(2)} \equiv n^{-1}\sum_{i=1}^n 1_{\{1\}}(X_i) = n^{-1}\sum_{i=1}^n 1\{X_i = 1\}.$$

A third estimator $\widehat{\theta}_n^{(3)}$ of θ is the MLE obtained by solving the score equation

$$0 = \sum_{i=1}^n \dot{\mathbf{l}}_\theta(X_i) = -\frac{n\overline{X}_n - n}{1-\theta} + \frac{n}{\theta}.$$

It is easily seen that the solution $\widehat{\theta}_n^{(3)} = 1/\overline{X}_n = \widehat{\theta}_n^{(1)}$.
(iv) Since $\widehat{\theta}_n^{(3)} = \widehat{\theta}_n^{(1)}$, these estimators are asymptotically efficient in the sense of Hájek's convolution theorem with asymptotic variance $1/I(\theta) = \theta^2(1-\theta)$. The empirical probability estimator $\widehat{\theta}_n^{(2)}$ has asymptotic variance $\theta(1-\theta) > \theta^2(1-\theta)$ and hence is asymptotically inefficient (for $\theta \in (0,1)$).

2. Repeat problem 1 above when the X_i's are i.i.d. Cauchy(θ); i.e. each of the X_i's has the common density $f(x) = 1/(\pi(1+x^2))$ for $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$.

Solution: (i) Now $s_\theta = \sqrt{p_\theta} = \sqrt{f(\cdot - \theta)}$, so

$$\dot{s}_\theta(x) = \frac{1}{2}p_\theta(x)^{-1/2}\dot{p}_\theta(x) = -\frac{1}{2}f(x-\theta)^{-1/2}f'(x-\theta)$$

is continuous in θ for each $x \in \mathbb{R}$. Furthermore

$$\dot{p}_\theta(x) = p_\theta(x)\left(-\frac{f'}{f}(x-\theta)\right),$$

and hence

$$I(\theta) = E_\theta\left(\frac{\dot{p}_\theta}{p_\theta(X)}\right)^2 = \int \frac{(f'(y))^2}{f(y)}dy = \frac{1}{2}$$

is (trivially) continuous in θ. Thus the hypotheses of Lemma 7.6 of van der Vaart (1998) hold, and we conclude that $\{p_\theta : \theta \in \mathbb{R}\}$ is differentiable in quadratic mean.
(ii) See (i) above.
(iii) Since p_θ is symmetric about θ one simple estimator of θ is the sample median $\widehat{\theta}_n^{(1)} = \mathbb{F}_n^{-1}(1/2)$. In a similar way $\widehat{\theta}_n^{(2)} = (\mathbb{F}_n^{-1}(p) + \mathbb{F}_n^{-1}(1-p))/2$ is also a consistent estimator of θ: noting that $F_\theta^{-1}(p) = \theta + \tan\{(p-1/2)\pi\}$

$$\begin{aligned}(\mathbb{F}_n^{-1}(p) + \mathbb{F}_n^{-1}(1-p))/2 &\to_p (F_\theta^{-1}(p) + F_\theta^{-1}(1-p))/2 \\ &= \theta + \frac{1}{2}(\tan\{(p-1/2)\pi\} + \tan\{-(p-1/2)\pi\}) = \theta.\end{aligned}$$

A third estimator of θ might be the one-step MLE discussed in Section 4.1:

$$\widehat{\theta}_n^{(3)} = \widehat{\theta}_n^{(1)} + 2n^{-1}\sum_{i=1}^n \dot{\mathbf{l}}_\theta(X_i; \widehat{\theta}_n^{(1)}).$$

(iv) The estimators $\widehat{\theta}_n^{(1)}$ and $\widehat{\theta}_n^{(2)}$ are inefficient; the estimator $\widehat{\theta}_n^{(3)}$ is asymptotically efficient (assuming that the model holds).

3. Ferguson, problem 4, page 124: Let X and Y be independent random variables with densities p_θ and q_θ depending on θ. Assume that the Fisher informations $I_X(\theta)$ and $I_Y(\theta)$ for θ based on observing X or Y both exist. Show that the Fisher information for θ based on observing the pair (X, Y) is given by $I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta)$.

Solution: Since X and Y are independent, the joint density is given by

$$p_{X,Y}(x, y; \theta) = p_X(x; \theta) p_Y(y; \theta).$$

Then

$$\log p_{X,Y}(x, y; \theta) = \log p_X(x; \theta) + \log p_Y(y; \theta),$$

and it follows that the score function for θ based on (X, Y) is

$$\dot{\mathbf{l}}_{\theta,X,Y}(x, y) = \dot{\mathbf{l}}_{\theta,X}(x) + \dot{\mathbf{l}}_{\theta,Y}(y),$$

and the information for θ based on (X, Y) is

$$\begin{aligned} I_{X,Y}(\theta) &= E_\theta \dot{\mathbf{l}}^2_{\theta,X,Y}(X, Y) = E_\theta \left(\dot{\mathbf{l}}_{\theta,X}(X) + \dot{\mathbf{l}}_{\theta,Y}(Y) \right)^2 \\ &= E_\theta \dot{\mathbf{l}}^2_{\theta,X}(X) + E_\theta \dot{\mathbf{l}}^2_{\theta,Y}(Y) \\ &= I_X(\theta) + I_Y(\theta) \end{aligned}$$

since, by the independence of X and Y,

$$E_\theta \dot{\mathbf{l}}_{\theta,X}(X) \dot{\mathbf{l}}_{\theta,Y}(Y) = E_\theta \dot{\mathbf{l}}_{\theta,X}(X) \cdot E_\theta \dot{\mathbf{l}}_{\theta,Y}(Y) = 0 \cdot 0 = 0.$$

4. Consider the Laplace location family $p_\theta(x) = 2^{-1} \exp(-|x - \theta|)$ for $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$.
(a) Does the hypothesis (M5) of Theorem 3.2.22, page 11 of the Course Notes hold in this case? Does the hypothesis (M4) of Theorem 3.2.2 hold?
(b) Show that the Laplace location family is differentiable in quadratic mean. What is the consequence of this for the behavior of the local log-likelihood ratios? What is the resulting information for the location parameter θ?
(c) Apply the methods of section 3.5 to show that with $\theta_0 \in \mathbb{R}$ fixed and $\theta_n = \theta_0 + n^{-1/2} h$, then for any estimator T_n of θ we have

$$\liminf_{n \to \infty} \inf_{T_n} \max \{ E_{\theta_n} n |T_n - \theta_n|^2, \ E_{\theta_0} n |T_n - \theta_0|^2 \} \geq c I(\theta_0)^{-1}$$

for some choice of h and an absolute constant c.

Solution: (a) For the Laplace location family, $\log p_\theta(x) = -\log 2 - |x - \theta|$ is not differentiable at $\theta = x$, while it is differentiable at all other values of θ with derivative $\dot{\mathbf{l}}_\theta(x) = \text{sign}(x - \theta)$. Thus the hypothesis M4 of Theorem 3.2.2 fails. The hypothesis M5 also fails since $\dot{\mathbf{l}}_\theta(x)$ is not twice differentiable at $\theta = x$.
(b) The Laplace location family is differentiable in quadratic mean: with $\dot{\mathbf{l}}_\theta(x) = \text{sign}(x - \theta)$ as in (a), it suffices to prove that

$$\int \left\{ \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - 2^{-1} h \dot{\mathbf{l}}_\theta \sqrt{p_\theta} \right\}^2 dx = o(|h|^2), \tag{0.1}$$

and then it follows that $I(\theta) = E_\theta \dot{\mathbf{l}}_\theta(X)^2 = 1$. To show that (0.1) holds, it suffices to prove the claim for $\theta = 0$ and $h > 0$. Then we want to show that

$$\int \left(\sqrt{p_h(x)} - \sqrt{p_0(x)} - (1/2)h\text{sign}(x) \cdot \sqrt{p_0(x)}\right)^2 dx = o(h^2). \tag{0.2}$$

But the left side of the last display can be rewritten as

$$\begin{aligned}
&\int \left(\sqrt{\frac{p_h(x)}{p_0(x)}} - 1 - (1/2)h\text{sign}(x)\right)^2 2^{-1}e^{-|x|}dx \\
&= \int \left(\sqrt{\frac{e^{-|x-h|}}{e^{-|x|}}} - 1 - (1/2)h\text{sign}(x)\right)^2 2^{-1}e^{-|x|}dx \\
&= \int \left(\sqrt{e^{-(|x-h|-|x|)}} - 1 - (1/2)h \cdot \text{sign}(x)\right)^2 2^{-1}e^{-|x|}dx \\
&= \int_{-\infty}^0 \left(e^{-h/2} - 1 + (1/2)h\right)^2 \cdot 2^{-1}e^{-|x|}dx \\
&\qquad + \int_h^\infty \left(e^{h/2} - 1 - (1/2)h\right)^2 \cdot 2^{-1}e^{-|x|}dx \\
&\qquad + \int_0^{h/2} \left(e^{-(h-2x)/2} - 1 - (1/2)h\right)^2 \cdot 2^{-1}e^{-|x|}dx \\
&\qquad + \int_{h/2}^h \left(e^{-(2x-h)/2} - 1 - (1/2)h\right)^2 \cdot 2^{-1}e^{-|x|}dx
\end{aligned}$$

where the third and fourth terms are clearly of order $O(h^5)$. The first and second terms are easily bounded by

$$\begin{aligned}
&2^{-1}(e^{-h/2} - (1 - h/2))^2 = O(h^4) \quad \text{and} \\
&2^{-1}(e^{h/2} - (1 + h/2))^2 = O(h^4).
\end{aligned}$$

Thus we conclude that (0.2) holds with $O(h^4)$ instead of $o(h^2)$ on the right side. But this easily yields the desired conclusion.
(c) Let $\theta_n = \theta_0 + hn^{-1/2}$. The differentiability in quadratic mean established in (b) implies that $nH^2(p_{\theta_n}, p_{\theta_0}) \to (1/8)h^2 I(\theta_0) = h^2/16$. We conclude from Proposition 3.5.1 that

$$\liminf_{n\to\infty} \inf_{T_n} \max\{E_{\theta_n} n|T_n - \theta_n|^2,\ E_{\theta_0} n|T_n - \theta_0|^2\} \geq (h^2/16)\exp(-h^2/16).$$

Maximizing this bound with respect to h yields a lower bound of e^{-1} by choosing $h = 4$.

5. Ferguson, problem 6, page 125: What was thought to be a certain species of moth is attracted to a capture tank at rate λ per day. On the first day, the number X of moths caught was recorded. It is assumed that X has a Poisson distribution with mean λ. Later it was pointed out that this species is, in fact, two different similar species, so a second day of capture was undertaken. This time, the numbers Y_1 and Y_2 of moths caught of these species separately were noted. It is assumed that these

are Poisson random variables with means λ_1 and λ_2 where $\lambda_1 + \lambda_2 = \lambda$, and it is assumed that X, Y_1, and Y_2 are independent.
(a) Using X, Y_1, and Y_2, find the maximum likelihood estimate of λ_1 and λ_2.
(b) Assuming λ_1 and λ_2 large, what is the approximate variance of your estimator?

Solution: (a) The joint density of (X, Y_1, Y_2) is

$$p_X(x)p_{Y_1}(y_1)p_{Y_2}(y_2) \quad = \quad e^{-\lambda}\frac{\lambda^x}{x!} \cdot e^{-\lambda_1}\frac{\lambda_1^{y_1}}{y_1!} \cdot e^{-\lambda_1}\frac{\lambda_1^{y_1}}{y_1!}$$

where $\lambda = \lambda_1 + \lambda_2$. Thus

$$l_n(\lambda_1, \lambda_2) = X \log \lambda - \lambda + Y_1 \log \lambda_1 - \lambda_1 + Y_2 \log \lambda_2 - \lambda_2,$$

and hence the score equations are

$$\begin{aligned} \dot{\mathbf{l}}_{\lambda_1}(X, Y_1, Y_2) &= \frac{X}{\lambda_1 + \lambda_2} + \frac{Y_1}{\lambda_1} - 2 = 0, \\ \dot{\mathbf{l}}_{\lambda_2}(X, Y_1, Y_2) &= \frac{X}{\lambda_1 + \lambda_2} + \frac{Y_2}{\lambda_2} - 2 = 0. \end{aligned}$$

Multplying across the first equation by $\lambda_1(\lambda_1+\lambda_2)$ and across the second by $\lambda_1(\lambda_1+\lambda_2)$ yields

$$\begin{aligned} \lambda_1 X + (\lambda_1 + \lambda_2)Y_1 - 2\lambda_1(\lambda_1 + \lambda_2) = 0, \\ \lambda_2 X + (\lambda_1 + \lambda_2)Y_2 - 2\lambda_2(\lambda_1 + \lambda_2) = 0. \end{aligned}$$

Adding these two equations yields

$$(\lambda_1 + \lambda_2)X + (\lambda_1 + \lambda_2)(Y_1 + Y_2) - 2(\lambda_1 + \lambda_2)^2 = 0,$$

and hence (after dividing by $\lambda_1 + \lambda_2$,

$$X + Y_1 + Y_2 \equiv X + Y - 2(\lambda_1 + \lambda_2) = 0.$$

Thus

$$\widehat{\lambda}_1 + \widehat{\lambda}_2 = (1/2)(X + Y).$$

Subtracting the two equations yields

$$(\lambda_1 - \lambda_2)X + (Y_1 - Y_2)(\lambda_1 + \lambda_2) - 2(\lambda_1 - \lambda_2)(\lambda_1 + \lambda_2),$$

or

$$(\lambda_1 - \lambda_2)(X - 2(\lambda_1 + \lambda_2)) = -(Y_1 - Y_2)(\lambda_1 + \lambda_2),$$

which in turn yields

$$\widehat{\lambda}_1 - \widehat{\lambda}_2 = \frac{-(Y_1 - Y_2)(X + Y)/2}{X - (X + Y)} = \frac{(Y_1 - Y_2)(1/2)(X + Y)}{Y}.$$

Adding and subtracting again yields

$$\widehat{\lambda_1} = \frac{Y_1}{Y}\frac{X+Y}{2} \equiv g(X, Y_1, Y_2), \quad \text{and} \quad \widehat{\lambda_2} = \frac{Y_2}{Y}\frac{X+Y}{2}.$$

(b) The function g given in the last display by

$$g(x, y_1, y_2) = \frac{y_1}{y_1+y_2}\left(\frac{x+y_1+y_2}{2}\right) = \frac{y_1}{2}\left(\frac{x}{y}+1\right)$$

has gradient given by

$$\dot{g}(x, y_1, y_2) = \left(\frac{y_1}{2y}, \frac{1}{2}\left(\frac{x+y}{y}\right) - \frac{y_1 x}{2y^2}, -\frac{y_1}{2}\frac{x}{y^2}\right)^T,$$

so that

$$\dot{g}(\lambda, \lambda_1, \lambda_2) = \left(\frac{\lambda_1}{2\lambda}, 1 - \frac{\lambda_1}{2\lambda}, -\frac{\lambda_1}{2\lambda}\right)^T.$$

Thus the approximate variance of $\widehat{\lambda_1}$ is

$$\begin{aligned} &\dot{g}(\lambda, \lambda_1, \lambda_2)^T \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \dot{g}(\lambda, \lambda_1, \lambda_2) \\ &= \lambda_1 - \frac{\lambda_1^2}{2\lambda} = \lambda_1\left(1 - \frac{\lambda_1}{2\lambda}\right). \end{aligned}$$