Stat 513
Homework key 5

February 9, 2016

REGULAR PROBLEMS

1. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. Bernoulli random variables with $Y_i \sim \mathcal{B}(1, p)$, $p \in (0, 1)$. Denote the target of inference by $\theta = Var(Y_i)$.

   (a) What is the Cramér-Rao lower bound for the variance of an unbiased estimator of θ?

   (b) Can you find an estimator that achieves that lower bound? If so, derive that estimator. If not, explain why not.

   (c) Find the maximum likelihood estimate $\hat{\theta}$ of θ.

   (d) Derive an asymptotic distribution for $\hat{\theta}$ in the case that $p \neq \frac{1}{2}$.

   (e) Find the asymptotic relative efficiency of $\hat{\theta}$ to the unbiased nonparametric estimator s^2, the sample variance, when $p \neq \frac{1}{2}$.

   (f) Why is the asymptotic distribution for $\hat{\theta}$ derived in part (d) not useful when $p = \frac{1}{2}$? Derive an asymptotic distribution that could be used when $p = \frac{1}{2}$.

Solution:

(a) We first find the efficient score $\mathcal{U}(p|y)$.

$$\begin{aligned} f_Y(y|p) &= p^y(1-p)^{1-p}; \\ \log f_Y(y|p) &= y \log p + (1-y)\log(1-p); \\ \mathcal{U}(p|y) &= \frac{y}{p} - \frac{1-y}{1-p} \\ &= \frac{1}{p(1-p)}(y-p). \end{aligned}$$

Then, the Fisher's information for p is

$$J(p) = E[\mathcal{U}(p|Y)^2] = \left(\frac{1}{p(1-p)}\right)^2 Var(Y) = \frac{1}{p(1-p)}.$$

Now, $\theta = p(1-p) = g(p)$, so we can write $g'(p) = 1 - 2p$. Then, the CRLB for unbiased estimators of θ is

$$CRLB(\theta) = \frac{g'(p)^2}{J(p)} = (1-2p)^2 p(1-p) = \cdots = (1-4\theta)\theta.$$

(b) We know that a BRUE exists for p because

$$\mathcal{U}(p|y) = \frac{1}{p(1-p)}(y-p) = h(p)(y-p).$$

Since $\theta = p(1-p)$ is a non-linear function of p, there is no BRUE for θ.

(c) The MLE $\hat{p}$ for p solves

$$\frac{1}{p(1-p)} \sum_{i=1}^{n} (Y_i - p) = 0.$$

Hence $\hat{p} = \bar{Y}$. By invariance of the MLE under transformations, we have that $\hat{\theta} = \hat{p}(1-\hat{p})$ is the MLE for θ.

(d) From the CLT, we have that

$$\sqrt{n}(\hat{p} - p) \to_d \mathcal{N}(0, p(1-p)).$$

By the delta method, since $g'(p) \neq 0$ for $p \neq \frac{1}{2}$,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d \mathcal{N}(0, (1-4\theta)\theta).$$

(e) Recall that since Y is binary that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - (\bar{Y})^2\right) = \frac{n}{n-1}\left(\bar{Y}(1-\bar{Y})\right) = \frac{n}{n-1}\hat{\theta}.$$

From Slutsky's theorem, we see that s^2 has the same asymptotic distribution as $\hat{\theta}$ so that it has ARE of 1 (s^2 is asymptotically efficient).

(f) When $p = \frac{1}{2}$, the asymptotic variance in (d) was 0, meaning we obtain an asymptotically degenerate distribution. If we expand $g(\hat{p})$ in its Taylor series around p, we can obtain a valid asymptotic distribution for $p = \frac{1}{2}$. Note that

$$g(\hat{p}) = g(p) + g'(p)(\hat{p} - p) + \frac{g''(p)}{2}(\hat{p} - p) + \dots$$

Now, $g'(p) = 0$ in our case so we write

$$n(g(\hat{p}) - g(p)) = \frac{g''(p)}{2}\left(\sqrt{n}(\hat{p} - p)\right)^2 + \dots \to_d \frac{p(1-p)g''(p)}{2}\mathcal{N}(0,1)^2 = \frac{p(1-p)g''(p)}{2}\chi_1^2.$$

Furthermore, $g''(p) = -2$, so

$$n(\hat{\theta} - \theta) \to -\theta\chi_1^2.$$

An alternative approach to solve this problem is to guess that the estimator will be n-consistent, and to try to find the asymptotic distribution of

$$n(\hat{p}(1-\hat{p}) - p(1-p))$$

Plugging in $p(1-p) = \frac{1}{4}$ will yield the desired result, where we also note that the negative sign makes sense, because $\hat{p}(1-\hat{p}) \leq 0.25$.

2. Suppose $Y_1, Y_2, \dots, Y_n$ are i.i.d. normal random variables with $Y_i \sim \mathcal{N}(\theta, \theta)$ with $\theta \in (0, \infty)$.

(a) What is the Cramér-Rao lower bound for the variance of an unbiased estimator of θ?

(b) For what function $g(\theta)$ can you derive a best regular unbiased estimator (BRUE)? Justify your answer.

(c) Find the maximum likelihood estimate $\hat{\theta}$ of θ.

(d) Derive an asymptotic distribution for $\hat{\theta}$.

Solution:

(a) We have that by differentiating the log likelihood, the score $\mathcal{U}(\theta|Y)$ is just

$$\begin{aligned}
f_Y(y|\theta) &\propto \theta^{-1/2}\exp\left(y - \frac{y^2}{2\theta} - \frac{\theta}{2}\right)\\
\log f_Y(y|theta) &= -\frac{1}{2}\log\theta + y - \frac{y^2}{2\theta} - \frac{\theta}{2} + constant\\
\mathcal{U}(\theta|\mathcal{Y}) &= \frac{1}{2\theta^2}(Y^2 - \theta(\theta+1)).
\end{aligned}$$

We calculate the information $J(\theta)$ by taking the expected negative derivative of $\mathcal{U}(\theta|Y)$,

$$J(\theta) = E\left[-\frac{\partial}{\partial\theta}\mathcal{U}(\theta|Y)\right] = E\left[\frac{2\theta+1}{2\theta^2} + \frac{Y^2 - \theta(\theta+1)}{\theta^3}\right] = \frac{2\theta+1}{\theta^2}.$$

Hence, the CRLB is given by

$$CRLB(\theta) = J^{-1}(\theta) = \frac{2\theta^2}{2\theta+1}.$$

(b) We have seen that the score was of the form

$$\mathcal{U}(\theta|Y) = h(\theta)(Y^2 - \theta(\theta+1)),$$

so a BRUE exists for $g(\theta) = \theta(\theta+1)$.

(c) We find the MLE $\hat{\theta}$ for θ by solving

$$\sum_{i=1}^{n} \mathcal{U}(\theta|Y_i) = 0.$$

In particular, $\hat{\theta}$ is the positive root of the quadratic equation

$$\theta^2 + \theta - n^{-1}\sum_{i=1}^{n} Y_i^2 = 0,$$

$$\hat{\theta} = \frac{-1 + \sqrt{1 + 4n^{-1}\sum_{i=1}^{n} Y_i^2}}{2}.$$

(d) Since this is a regular problem, we have that

$$\sqrt{n}(\hat{\theta} - \theta) \to_d \mathcal{N}\left(0, CRLB(\theta) = \frac{2\theta^2}{2\theta + 1}\right).$$

3. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables which for some $\theta \in (2, \infty)$ have density

$$f_Y(y|\theta) = \theta y^{-(\theta+1)}\mathbf{1}_{(1,\infty)}(y).$$

(a) Find the maximum likelihood estimate $\hat{\theta}$ of θ.

(b) Prove that $\hat{\theta}$ is biased.

(c) Find the asymptotic distribution of the maximum likelihood estimate for $g(\theta) = Var(Y)$.

Solution:

(a) The log-likelihood for Y is just $\ell(\theta, y) = \log\theta - (\theta + 1)\log y$, so

$$\mathcal{U}(\theta|Y) = -\left(\log Y - \frac{1}{\theta}\right).$$

This means that $\hat{\theta}$ solves

$$-\sum_{i=1}^{n}\left(\log Y_i - \frac{1}{\theta}\right) = 0,$$

leading to $\hat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n}\log Y_i\right)^{-1}$.

(b) Since this is a regular problem, we know that $\mathcal{U}(\theta|Y)$ has mean 0. This implies that $E[\log Y] = 1/\theta$, so

$$E\frac{1}{\hat{\theta}} = \frac{1}{\theta}.$$

By Jensen's inequality, $E[\hat{\theta}] > \theta$, so the MLE is biased.

(c) To find the information $J(\theta)$ for θ, we take the mean negative derivative of the score,

$$-\frac{\partial}{\partial\theta}\mathcal{U}(\theta|y) = \frac{1}{\theta^2}.$$

It follows that $J(\theta)^{-1} = \theta^2$. Using our result for MLEs of regular parametric models,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d \mathcal{N}(0, \theta^2).$$

4. Consider a "dose-response" regression in which we presume that the mean "response" Y_i is related to "dose" X_i by

$$(Y_i|X_i = x_i) = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where we presume the ϵ_i's are i.i.d. with $\epsilon_i \sim (0, \sigma^2)$. We are interested in estimating the dose θ_{50} such that $E[Y|X = \theta_{50}] = 50$.

(a) Derive expressions for the best linear unbiased estimator $\hat{\vec{\beta}}$ of $\vec{\beta}$.

(b) What is the asymptotic distribution of $\hat{\vec{\beta}}$? Be sure to provide the assumptions under which you derived the asymptotic distribution.

(c) Derive an expression for an estimator $\hat{\theta}_{50} = g(\hat{\vec{\beta}})$, along with its asymptotic distribution.

Solution:

(a) By the Gauss-Markov theorem, the BLUE is given by

$$\hat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y},$$

where $\mathbf{X} = (\vec{1}, \vec{X})$, and $\vec{X} = (X_1, \ldots, X_n)^T$. Supposing that the X_i have been centered,

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^n X_i^2 \end{pmatrix},$$

$$\hat{\vec{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \end{pmatrix}.$$

(b) We have seen that if

$$\max_{1 \le i \le n} \frac{|X_i|^2}{\sum_{i=1}^n |X_i|^2} \to 0,$$

then

$$(\mathbf{X}^T\mathbf{X})^{1/2}(\hat{\vec{\beta}} - \vec{\beta}) = \begin{pmatrix} \sqrt{n}(\hat{\beta}_0 - \beta_0) \\ \sqrt{\sum_{i=1}^n X_i^2}(\hat{\beta}_1 - \beta_1) \end{pmatrix} \to_d \mathcal{N}(\vec{0}, \sigma^2 \mathbf{I}_{2\times 2}).$$

If we have further that, for some $\tau > 0$, $n^{-1}\sum_{i=1}^n X_i^2 \to \tau^2$, then

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \to_d \mathcal{N}\left(\vec{0}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/\tau^2 \end{pmatrix}\right).$$

(c) Note that $\theta_{50} = g(\vec{\beta}) = \frac{50-\beta_0}{\beta_1}$. Hence, a consistent estimator can be obtained by defining

$$\hat{\theta}_{50} = \frac{50 - \hat{\beta}_0}{\hat{\beta}_1}.$$

To obtain an asymptotic distribution, we appeal to the delta method, which requires calculating the gradient

$$grad(\theta_{50})^T = \left(\frac{-1}{\beta_1}, \quad \frac{\beta_0 - 50}{\beta_1^2}\right).$$

Then, an asymptotic distribution for our estimator is

$$\sqrt{n}\left(\hat{\theta}_{50} - \theta_{50}\right) \to_d \mathcal{N}\left(0, grad(\theta_{50})^T \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/\tau^2 \end{pmatrix} grad(\theta_{50})\right) = \mathcal{N}\left(0, \frac{\sigma^2}{\beta_1^2}\left(1 + \frac{1}{\tau^2}\left(\frac{50-\beta_0}{\beta_1}\right)^2\right)\right).$$

MORE INVOLVED PROBLEMS

5. Suppose $Y_1 Y_2, \ldots, Y_n$ are i.i.d. random variables with the double exponential distribution having density for $\theta > 0$
$$f_Y(y|\theta) = \frac{1}{2}\theta e^{-\theta|y|}.$$

   (a) Find a method of moments estimate $\tilde{\theta}$ for θ and derive its asymptotic distribution.

   (b) Find a maximum likelihood estimate $\hat{\theta}$ for θ.

   (c) Derive the asymptotic distribution of $\hat{\theta}$. (Hint: Is this a regular problem? What can you use if it is not?)

   (d) What is the asymptotic relative efficiency of the MME $\tilde{\theta}$ compared to MLE $\hat{\theta}$?

Solution:

(a) We note that since the density of Y is symmetric that our method of moments estimator (MME) should be based on the second moment. The population moment is
$$E[Y^2] = \int_{-\infty}^{\infty} y^2 \frac{1}{2}\theta e^{-\theta|y|} dy = \int_0^{\infty} y^2 \theta e^{-\theta y} dy = \frac{2}{\theta^2},$$
where we recognize the integral as the expection $E[T^2]$, where $T \sim \mathcal{E}(\theta)$. Hence, our MME is
$$\tilde{\theta} = \left(\frac{1}{2n}\sum_{i=1}^n Y_i^2\right)^{-1/2}.$$
The asymptotic distribution of $\tilde{\theta}$ can be found via the delta method, where we start with
$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \frac{2}{\theta^2}\right) \to_d \mathcal{N}(0, Var(Y^2) = 20/\theta^4).$$
We note that for $T \sim \mathcal{E}(\theta)$, $E[T^m] = m!\theta^{-m}$, so that
$$E[Y^4] = \int_{-\infty}^{\infty} y^4 \frac{1}{2}\theta e^{-\theta|y|} dy = \int_0^{\infty} y^4 \theta e^{-\theta y} dy = 4!\theta^{-4} = \frac{24}{\theta^4},$$
and $Var(Y) = E[Y^4] - E[Y^2]^2 = 20/\theta^4$.

We now apply the delta method with $g(t) = \left(\frac{t}{2}\right)^{-1/2}$ and $g'(t) = -\frac{\sqrt{2}}{2}t^{-3/2}$, so that $g'(2/\theta^2) = \theta^6/16$. This yields
$$\sqrt{n}\left(\tilde{\theta} - \theta\right) \to_d \mathcal{N}\left(0, \frac{5}{4}\theta^2\right).$$

(b) The log likelihood is $\ell(\theta|y) = \log\theta - \theta|y| + constant$, so the efficient score is just
$$\mathcal{U}(\theta|Y_i) = \frac{1}{\theta} - |Y_i|.$$
The MLE for θ is then
$$\hat{\theta} = \left(\frac{1}{n}\sum_{i=1}^n |Y_i|\right)^{-1}.$$

(c) This is a regular parametric model, so we can appeal to the consistent asymptotic normality of the MLE. The asymptotic variance will just be the inverse information, where the information for θ is
$$J(\theta) = -E\left[\frac{\partial}{\partial\theta}\mathcal{U}(\theta|Y)\right] = \frac{1}{\theta^2},$$
so
$$\sqrt{n}(\hat{\theta} - \theta) \to_d \mathcal{N}(0, \theta^2).$$

(d) The asymptotic relative efficiency is
$$ARE(\hat{\theta}, \tilde{\theta}) = \frac{\theta^2}{\frac{5}{4}\theta^2} = \frac{4}{5} = 0.8,$$
so that we see the MME is less efficient than the MLE (which agrees with what we expect from our theory for MLEs in regular models).