R300 – Advanced Econometric Methods

PROBLEM SET 3 - SOLUTIONS

Due by Mon. October 26

1. When the score is nonlinear in θ, i.e.,

$$\frac{\partial \log f_\theta(x)}{\partial \theta}$$

is not a linear function of θ, the maximum-likelihood estimator (MLE) of θ is biased, in general. The bias is typically n^{-1}, i.e.,

$$E_\theta(\hat{\theta} - \theta) = \frac{b_\theta}{n} + o(n^{-1})$$

for some constant b_θ.

(i) Derive the bias for the MLE of σ^2 when $x_i \sim N(\mu, \sigma^2)$.

(ii) Let $\hat{\theta}^{(-i)}$ be the MLE computed from the subsample of size $n-1$ obtained on omitting the ith observation. You have n such MLEs. Show in the case of (i) that the jackknife estimator

$$\breve{\theta} := n\,\hat{\theta} - (n-1)\,\overline{\theta}, \qquad \overline{\theta} := n^{-1} \sum_i \hat{\theta}^{(-i)},$$

is exactly unbiased.

(i) The MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \overline{x})^2.$$

We have

$$E_\theta(\hat{\sigma}^2) = E_\theta((x_i - \mu)^2) - E_\theta((\overline{x} - \mu)^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\,\sigma^2.$$

So, here, the bias is $-\sigma^2/n$.

(ii) We have

$$\hat{\sigma}^2_{-i} = \frac{1}{n-1} \sum_{j \neq i} (x_j - \overline{x}_{-i})^2, \qquad \overline{x}_{-i} := \frac{1}{n-1} \sum_{j \neq i} x_j.$$

As before,

$$E_\theta(\hat\sigma^2_{-i}) = \sigma^2 - \frac{\sigma^2}{n-1}$$

and, therefore,

$$n\,\hat\sigma^2 - (n-1)\,\hat\sigma^2_{-i}$$

has expectation

$$E_\theta(n\,\hat\sigma^2 - (n-1)\,\hat\sigma^2_{-i}) = n\left(\frac{n-1}{n}\sigma^2\right) - (n-1)\left(\frac{n-2}{n-1}\sigma^2\right) = ((n-1)-(n-2))\,\sigma^2 = \sigma^2$$

for any i. Therefore, the same holds on averaging over the i to construct the jackknife estimator

$$\frac{\sum_i n\,\hat\sigma^2 - (n-1)\,\hat\sigma^2_{-i}}{n}.$$

In this example the jackknife kills all bias and is exactly unbiased.

---

2. Wage data are often top coded. We wish to estimate the following linear model for log wages (w_i),

$$w_i = x_i'\beta + \varepsilon_i, \qquad \varepsilon_i \sim N(0,\sigma^2),$$

but, while we observe actual (log) wage w_i when $w_i \leq c$ we only observe c when $w_i > c$. Thus, our actual data are a random sample on (y_i, x_i) where

$$y_i = \begin{cases} w_i & \text{if } w_i \leq c \\ c & \text{if } w_i > c \end{cases}.$$

(i) Set up the likelihood function for this problem.

(ii) Derive an expression for the conditional mean of $w_i|x_i$ in the subpopulation with $w_i \leq c$.

(iii) What does your response to (ii) imply for the suitability of a least-squares regression of the non-coded outcomes on x_i to recover β? Recall that this least-squares estimator is the solution to

$$\min_b \sum_{i:y_i<c} (y_i - x_i b)^2.$$

(i) The outcome variable (conditional on x_i) is mixed continuous/discrete, with a mass point at the threshold c. below that point, the data follow a normal distribution $N(x_i'\beta, \sigma^2)$. So, with $\theta = (\beta, \sigma^2)$, the density is

$$f_\theta(y_i|x_i) = \begin{cases} \frac{1}{\sigma}\,\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) & \text{if } y_i < c \\ 1 - \varPhi\left(\frac{c - x_i'\beta}{\sigma}\right) & \text{otherwise} \end{cases}.$$

The likelihood function becomes

$$\prod_{y_i<c} \frac{1}{\sigma} \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \times \prod_{y_i=c} \left(1 - \Phi\left(\frac{c - x_i'\beta}{\sigma}\right)\right).$$

(ii) If $v_i \sim N(0, \sigma^2)$ then $v_i | v_i \le c$ is truncated normal. Its distribution function is

$$P(v_i \le v | v_i \le c) = \frac{P(v_i \le v)}{P(v_i \le c)} = \frac{\Phi(v/\sigma)}{\Phi(c/\sigma)},$$

and differentiating gives the density function as

$$\frac{1}{\sigma} \frac{\phi(v/\sigma)}{\Phi(c/\sigma)}.$$

Then (integrating by parts)

$$E(v_i | v_i \le c) = \frac{\int_{-\infty}^{c} \frac{v}{\sigma} \phi(v/\sigma)\, dv}{\Phi(c/\sigma)} = -\sigma \frac{\phi(c/\sigma)}{\Phi(c/\sigma)} =: -\sigma\lambda(c/\sigma).$$

We then have

$$E(w_i | x_i, w_i < c) = x_i\beta + E(\varepsilon_i | \varepsilon_i < c - x_i\beta) = x_i\beta - \sigma\lambda((c - x_i'\beta)/\sigma).$$

(iii) The above calculation shows that, in the subpopulation of workers whose wages are not top coded, the conditional mean is nonlinear in x_i. Hence, a linear regression is not appropriate for estimating the slope β.

---

3. Recall the problem where $x_i \sim N(\mu, \sigma_i^2)$. We previously considered

$$\check{x} = \sum_{i=1}^{n} w_i x_i, \qquad w_i = \frac{1/\sigma_i^2}{\sum_{i'=1}^{n} 1/\sigma_{i'}^2}$$

as an estimator of μ.

(i) To implement $\check{x}$ we need an estimator of the σ_i^2. Let $\hat{\varepsilon}_i = x_i - \overline{x}$, The *usual* estimator would be

$$\hat{\varepsilon}_i^2.$$

Show that this estimator is biased.

(ii) A *cross-fit* estimator of σ_i^2 is

$$\hat{\sigma}_i^2 = x_i(x_i - \overline{x}_{-i}), \qquad \overline{x}_{-i} = \frac{1}{n-1}\sum_{j \neq i} x_j.$$

Show that $\hat{\sigma}_i^2$ is an unbiased estimator of σ_i^2.

(iii) Does this imply that the plug-in estimator

$$\hat{\hat{x}} = \sum_{i=1}^n \hat{w}_i x_i, \qquad \hat{w}_i = \frac{1/\hat{\sigma}_i^2}{\sum_{i'=1}^n 1/\hat{\sigma}_{i'}^2}$$

of μ is unbiased?

(i) We can always write $x_i = \theta + \varepsilon_i$ for $\varepsilon_i \sim N(0, \sigma_i^2)$. We have

$$\hat{\varepsilon}_i = x_i - \hat{x}_i = (\theta + \varepsilon_i) - (\theta + \sum_j \varepsilon_j/n) = \varepsilon_i - \frac{\sum_{j=1}^n \varepsilon_j}{n}.$$

So,

$$\hat{\varepsilon}_i^2 = \left(\varepsilon_i - \frac{\sum_{j=1}^n \varepsilon_j}{n}\right)^2 = \varepsilon_i^2 + \frac{\sum_{j=1}^n \sum_{k=1}^n \varepsilon_j \, \varepsilon_k}{n^2} - 2\frac{\varepsilon_i \, \sum_{j=1}^n \varepsilon_j}{n},$$

and

$$E(\hat{\varepsilon}_i^2) = E(\varepsilon_i^2) + \frac{\sum_j E(\varepsilon_j^2)}{n^2} - 2\frac{E(\varepsilon_i^2)}{n} = \sigma_i^2 + \frac{n^{-1}\sum_{j=1}^n \sigma_j^2}{n} - 2\frac{\sigma_i^2}{n} \neq \sigma_i^2.$$

(ii) This is straightforward. Write

$$x_i = \theta + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma_i^2).$$

Then

$$\hat{\sigma}_i^2 = x_i(x_i - \hat{x}_{-i}) = (\theta + \varepsilon_i)\,\varepsilon_i - (\theta + \varepsilon_i)\frac{1}{n-1}\sum_{j \neq i} \varepsilon_j.$$

The first term has expectation $E(\varepsilon_i^2) = \sigma_i^2$. The second term has expectation zero. So, $\hat{\sigma}_i^2$ is unbiased.

(iii) The feasible estimator is a nonlinear function of the $\hat{\sigma}_i^2$ and, hence, will not itself be unbiased in general.

---

4. Suppose that you have a random sample from a Geometric distribution with parameter θ, i.e.,

$$f_\theta(x) = \theta\,(1-\theta)^{x-1}$$

for integers x.

(i) Derive the MLE of θ.

(ii) Show that the score at θ has expectation zero.

(iii) Compute the asymptotic variance of the MLE.

(iv) Is the MLE best asymptotically unbiased?

(i) The log-likelihood is

$$L_n(\theta) = \sum_{i=1}^{n} \log\theta + (x_i - 1)\log(1-\theta) = n\,\log\theta + n\,(\overline{x} - 1)\log(1-\theta).$$

A calculation gives the MLE as $\hat{\theta} = \overline{x}^{-1}$.

(ii) The score is

$$\frac{\partial \log f_\theta(x_i)}{\partial\theta} = \frac{1-\theta\,x_i}{\theta(1-\theta)}.$$

We have

$$E_\theta\left(\frac{\partial f_\theta(x_i)}{\partial\theta}\right) = E_\theta\left(\frac{1-\theta x_i}{\theta(1-\theta)}\right) = \frac{1-E_\theta(x_i)\theta}{\theta\,(1-\theta)} = 0,$$

with the last equality following from the fact that

$$E_\theta(x_i) = \theta^{-1}.$$

(iii) The variance of the score is

$$E_\theta\left(\left(\frac{1-x_i\,\theta}{\theta\,(1-\theta)}\right)^2\right) = \frac{1}{\theta^2(1-\theta)}.$$

We obtain the same result on calculating the information as (minus) the expected Hessian.

(iv) This is true for maximum likelihood (under correct specification).
