R300 – Advanced Econometric Methods

PROBLEM SET 3 - QUESTIONS

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

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

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

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

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

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

of μ is unbiased?

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?