R300 – Advanced Econometric Methods

PROBLEM SET 1 - SOLUTIONS

Due by Mon. October 12

1. Let y and x be scalar random variables. Show that the following hold.

   (i) $\text{cov}(E(y|x), y - E(y|x)) = 0$.

   (ii) $\text{var}(y - E(y|x)) \leq \text{var}(y)$.

   (iii) $\text{var}(y - E(y|x)) = \text{var}(y)$ when y and x are independent.

(i) Let $\varepsilon = y - E(y|x)$. Then we need to show that

$$\text{cov}(E(y|x), \varepsilon) = 0.$$

By definition of the covariance in the first step, the fact that $E(\varepsilon) = 0$ in the second step, and the law of iterated expectations and the fact that $E(\varepsilon|x) = 0$ in the third and fourth step, we have

$$\begin{aligned} \text{cov}(E(y|x), \varepsilon) &= E(E(y|x)\varepsilon) - E(E(y|x))\, E(\varepsilon) \\ &= E(E(y|x)\varepsilon) \\ &= E(E(y|x)E(\varepsilon|x)) \\ &= 0. \end{aligned}$$

(ii) We have, using the law of total variance and the fact that $E(\varepsilon|x) = 0$ (and so has zero variance)

$$\begin{aligned} \text{var}(\varepsilon) &= \text{var}(E(\varepsilon|x)) + E(\text{var}(\varepsilon|x)) \\ &= E(\text{var}(\varepsilon|x)) \\ &= E(\text{var}(y|x)) \\ &\leq E(\text{var}(y|x)) + \text{var}(E(y|x)) \\ &= \text{var}(y) \end{aligned}$$

(iii) This is immediate from the solution to (ii). By independence, $E(y|x) = E(y)$ and so $\text{var}(E(y|x)) = 0$ from which the equality follows.

---

2. Show that the following functions are probability mass/density functions and compute their first two moments.

(i) $f(x) = ax^{a-1}$, $x \in (0,1)$ and $a > 0$.

(ii) $f(x) = n^{-1}$, $x = 1, 2, \ldots, n$ for n a positive integer.

(iii) $f(x) = (3/2)(x-1)^2$, $x \in (0,2)$.

(i) It suffices to show that f is non-negative and integrates to one. Here, f is clearly non-negative as $a > 0$ and x is positive. Further,

$$\int_0^1 f(x)\, dx = a \int_0^1 x^{a-1}\, dx = x^a|_0^1 = 1.$$

So f is a density function.

Next, the first two moments are

$$\int_0^1 x\,(a\, x^{a-1})\, dx = a \int_0^1 x^a\, dx = \tfrac{a}{1+a}, \qquad \int_0^1 x^2\,(a\, x^{a-1})\, dx = \tfrac{a}{2+a},$$

by straightforward integration.

(ii) Here, f is the uniform mass function on the interval $[1, 2, \ldots, n])$, i.e., it places mass n^{-1} on each of the n masspoints. So, clearly, mass is positive and sums up to one over the support.

The first two moments are

$$\sum_{i=1}^n i/n = (n+1)/2, \qquad \sum_{i=1}^n i^2/n = (n+1)(2n+1)/6.$$

(iii) That f is a pdf is easy to see. The integral is straightforward as f is polynomial (as in (i)),

$$\frac{3}{2}\int_0^2 (x^2 + 1 - 2x)\, dx = \tfrac{3}{2}\,\left(\tfrac{x^3}{3} + x - x^2\right)\Big|_0^2 = 1.$$

The first and second moments are 1 and 8/5, respectively.

---

3. Suppose that, for a scalar random variable x,

$$F(v) = P(x \leq v)$$

is continuous and strictly increasing in v. Define the random variable $y = F(x)$. Derive the distribution of y.

Let $G(v) = P(y \leq v)$. Then, for any $v \in [0,1]$, we have

$$G(v) = P(F(x) \leq v) = P(x \leq F^{-1}(v)) = F(F^{-1}(v)) = v.$$

So,

$$G(v) = \begin{cases} 0 & \text{if } v < 0 \\ v & \text{if } v \in [0,1) \\ 1 & \text{if } v \geq 1 \end{cases}.$$

This is the uniform distribution on the interval $[0,1]$.

---

4. For each of the cases below, show that the score is mean zero and derive the efficiency bound for θ.

(i) Binomial with parameters (n, θ).

(ii) Exponential with mean $1/\theta$.

(iii) Exponential with mean θ.

(i) The sum of m independent Bernoulli random variables, each with success probability θ, is Binomial (m, θ). The mass function at $x = 0, 1, 2 \ldots$ equals

$$f_\theta(x) = \binom{m}{x} \theta^x (1-\theta)^{m-x}.$$

So,

$$\log f_\theta(x) = \log \binom{m}{x} + x \log \theta + (m-x) \log(1-\theta).$$

Therefore,

$$\frac{\partial \log f_\theta(x)}{\partial \theta} = \frac{x}{\theta} - \frac{m-x}{1-\theta} = \frac{x - m\theta}{\theta(1-\theta)}$$

and

$$\frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} = -\frac{x}{\theta^2} - \frac{m-x}{(1-\theta)^2}.$$

Now, $E_\theta(x) = m\theta$ and $\text{var}_\theta(x) = m\theta(1-\theta)$. Note that the score has mean zero. Its variance is

$$E_\theta\left(\frac{\partial \log f_\theta(x)}{\partial\theta}\frac{\partial \log f_\theta(x)}{\partial\theta}\right) = \frac{\text{var}_\theta(x)}{\theta^2(1-\theta)^2} = \frac{m}{\theta(1-\theta)}.$$

Equivalently,

$$-E_\theta\left(\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2}\right) = E_\theta\left(\frac{x}{\theta^2} + \frac{m-x}{(1-\theta)^2}\right) = \frac{m}{\theta} + \frac{m}{1-\theta} = \frac{m}{\theta(1-\theta)}.$$

The efficiency bound then is $\theta(1-\theta)/(nm)$.

(ii) The exponential has

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

for $x \geq 0$. Here,

$$\log f_\theta(x) = \log\theta - x\theta$$

and so

$$\frac{\partial \log f_\theta(x)}{\partial\theta} = \frac{1}{\theta} - x, \qquad \frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} = -\frac{1}{\theta^2}.$$

The efficiency bound is θ^2/n.

(iii) In this parametrization the exponential has

$$f_\theta(x) = \frac{1}{\theta}e^{-x/\theta}$$

for $x \geq 0$. Here,

$$\log f_\theta(x) = -\log\theta - \frac{x}{\theta}.$$

Thus,

$$\frac{\partial \log f_\theta(x)}{\partial\theta} = -\frac{1}{\theta} + \frac{x}{\theta^2} \qquad \frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2x}{\theta^3}.$$

The information becomes

$$1/\theta^2$$

The efficiency bound is θ^2/n.
