Econ 8822 — Feb. 2, 2024

Shakeeb Khan — Due: Feb. 14, 2024

Spring 2024

Problem Set 3

1. Time to get our hands a bit dirty with Matlab: Generate a random sample of 100 observations from the following model:

$$y_i = \max(\alpha_0 + x_i\beta_0 + \epsilon_i, 0)$$

where x_i is distributed standard normal, as is ϵ_i, and the two random variables are distributed independently of each other. The parameters α_0, β_0 are $0.5, 1$ respectively.

   (a) With the 100 values of y_i, x_i you just generated, evaluate an estimate of α_0, β_0 using the MLE.

   (b) Now replicate this exercise 400 times. Therefore you should now have 400 different estimates of α_0, β_0. With this you can now evaluate the following statistics: mean bias and mean squared error (MSE), which are, respectively, the average discrepancy between your estimates and the true value, and the average squared discrepancy between your estimates and the true values.

   (c) Repeat this exercise but now for 200 and 400 observations. What can you say about the values of the MSE for the different sample sizes (100,200,400)?

2. A *median* of a random variable X is a value m such that

$$P(X \leq m) \geq 1/2$$

and

$$P(X \geq m) \geq 1/2$$

In the special case where X is continuously distributed, we have :

$$\int_{-\infty}^{m} f(x)dx = \int_{m}^{+\infty} f(x)dx = 1/2$$

Show that

(a) $\min_a E[|X - a|] = E[|X - m|]$

(b) Let $Y = g(X)$ where $g(\cdot)$ is a (weakly) monotonic function. Show that $g(m)$ is a median for Y.

(c) Suppose instead of working with the median of X, we are more interested in an α^{th} quantile of X, for some $\alpha \in (0, 1)$ denoted by m_α and assumed to satisfy

$$P(X \leq m_\alpha) \geq \alpha$$

and

$$P(X \geq m_\alpha) \geq 1 - \alpha$$

repeat question a) replacing the absolute value function with the "check function"

$$\rho_\alpha(x) = |x| + (2\alpha - 1)x$$

Does the "equivariance property" you proved in b) hold in this case as well?

3. Assume that ϵ_1, ϵ_2 are i.i.d., continuously distributed random variables, with common density function $f(\cdot)$. Let $\epsilon_3 = \epsilon_2 - \epsilon_1$;

   (a) Show that ϵ_3 is symmetrically distributed around 0.

   (b) Suppose now that ϵ_1, ϵ_2 are independent, and symmetrically distributed around 0, but have **different** distributions. Show that ϵ_3 is still symmetrically distributed around 0.

4. Show that the various versions of the maximum score estimator discussed in class are numerically equivalent.