University of Illinois | | Department of Economics
Spring 2016 | Econ 590 | Roger Koenker

Problem Set 1

(1) Consider the exponential family submodels,

$$p_\epsilon^k(z) = c(\epsilon, p^k) \exp(\epsilon D(p^k, z)) p^k(z),$$

where $c(\epsilon, p^k)$ is just the normalizing constant at ϵ for the semiparametric model, p^k, at the kth iteration. And $D(p^k, z)$ is the efficient influence function, EIF, for estimation of the parameter $\psi(p)$ in the semiparametric model, $\mathcal{P}$.

  (a) Given a random sample, $\{z_i, \cdots, z_n\}$, describe how to find the MLE for ϵ first, assuming that you had an explicit form for c, and then if you did not.

  (b) Show that evaluating the log likelihood score equation,

$$\frac{d}{d\epsilon} \sum_{i=1}^{n} \log p_\epsilon^k(z_i)|_{\epsilon=0} = \sum_{i=1}^{n} D(p^k, z_i).$$

  (c) Conclude that iterating the MLE solutions for ϵ, eventually producing a p^k for which $\hat{\epsilon} \approx 0$ yields an efficient estimator of the parameter of interest $\psi(p^k)$.

(2) Reconsider question one for the other Diaz and Rosenblum subfamily

$$p_\epsilon^k(z) = c(\epsilon, p^k)(1 + \exp(-2tD(p^k, z)))^{-1} p^k(z),$$

and comment on the advantages if any of this formulation.

(3) Using the code provided in the lecture for the Diaz and Rosenblum estimator of the standard missing data problem, design a *small* simulation experiment to compare its performance with various naive estimators like the mean of the observed y_i's, and the Horvitz-Thompson estimator. Or, preferably, as an alternative design a small simulation experiment for the Diaz QTET estimator intended to compare approaches. The third installment of the Lecture Notes contains some initial steps in this direction.