MICHIGAN STATE UNIVERSITY

Department of Statistics and Probability

COLLOQUIUM

Jana Jurečková[1]

Charles University, Prague,

Averaged Regression Quantiles

Tuesday, September 17, 2013, 10:20am -11:10am

Refreshments 10:00am, C405 Wells Hall

Abstract

Consider the linear regression model

$$\mathbf{Y}_n = \mathbf{X}_n\boldsymbol{\beta} + \mathbf{U}_n$$

with observations $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^\top$, i.i.d. errors $\mathbf{U}_n = (U_1, \ldots, U_n)^\top$ with an unknown distribution function F, and unknown parameter $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^\top$. The $n \times (p+1)$ matrix $\mathbf{X} = \mathbf{X}_n$ is known or observable and $x_{i0} = 1$ for $i = 1, \ldots, n$ (i.e., β_0 is an intercept). The α-regression quantile $\widehat{\boldsymbol{\beta}}_n(\alpha)$ is a minimizer $\text{argmin}_{\mathbf{b}\in\mathbb{R}^\mathbf{n}} \sum_{i=1}^n \rho_\alpha(Y_i - \mathbf{x}_i^\top \mathbf{b})$, where $\mathbf{x}_i^\top$ is the i-th row of $\mathbf{X_n}$, $i = 1, \ldots, n$ and $\rho_\alpha(z) = |z|\{\alpha I[z > 0] + (1-\alpha)I[z < 0]\},\ z \in \mathbb{R}^1$.

The scalar statistic $\bar{B}_n(\alpha) = \bar{\mathbf{x}}_n^\top \widehat{\boldsymbol{\beta}}_n(\alpha)$, is called *averaged regression quantile*, where $n\bar{\mathbf{x}}_n = \sum_{i=1}^n \mathbf{x_{ni}}$. The statistic $\bar{B}_n(\alpha)$ is scale equivariant and regression equivariant. Some other properties of $\bar{B}_n(\alpha)$ are surprising; indeed, $\bar{B}_n(\alpha)$ is asymptotically equivalent to the $[n\alpha]$-quantile of the location model:

$$n^{1/2}\left[\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}) - U_{n:[n\alpha]}\right] = O_p(n^{-1/4}), \quad n \to \infty, \tag{1}$$

where $U_{n:1} \leq \cdots \leq U_{n:n}$ are the order statistics corresponding to $U_1, \ldots, U_n$. We shall illustrate this approximation numerically.

The statistics of type $\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha_2) - \widehat{\boldsymbol{\beta}}_n(\alpha_1))$ are invariant to the regression with design $\mathbf{X_n}$ and equivariant with respect to the scale. As such, they provide a tool for studentization of M-estimators in linear regression model, and whenever one needs to make a statistic scale-equivariant.

The approximation (1) remains true under a sequence of local alternative distributions, contiguous with respect to the sequence $\{\prod_{i=1}^n F(u_{ni})\}$, e.g. under the local heteroscedasticity.

Based on observations $Y_{n1}, \ldots, Y_{nn}$, we can estimate the quantile density function $q(u) = 1/f(F^{-1}(u))$ at a fixed point, even under nuisance regression. The estimator can be either of histogram or of kernel types, and it provides a useful tool for an inference.

[1]The talk is based on joint work with Jan Picek.

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.