Asymptotic normality of quantile regresssion estimators for samples from a finite population

2010年7月9日　統計数理研究所　オープンハウス

元山　斉　　融合プロジェクト特任研究員

Abstract We derive the asymptotic normality of the quantile regression estimators for a simple random sample from a finite population.

1. Quantile regression

Quantile regression is a method for estimating models of conditional quantile functions introduced by Koenker and Bassett (1978). Koenker and Bassett (1978) also derive the asymptotic normality of these estimators under some regularity conditions. We extend their result to the finite population case via the method of Kato et al. (2009).

Notations:

$\{\mathcal{P}_k, k = 1, 2, \ldots\}$: a sequence of finite populations of size N_k

$(y_j, 1, x_{1j}, \ldots, x_{pj}) = (y_j, \mathbf{x}_j)$, $j = 1, \ldots, N_k$: characteristics of $\mathcal{P}_k$

$(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n_k$: simple random sample without replacement from $\mathcal{P}_k$

Model:

$$\begin{aligned} y_j &= \beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj} + u_j \\ &= \boldsymbol{\beta}^T \mathbf{x}_j + u_j, \quad j = 1, \ldots, N_k, \end{aligned}$$

$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$: an unknown $(p+1) \times 1$ parameter vector

u_j : the unexplained residual terms

$F_k(\cdot) = \Sigma_{j=1}^{N_k} I(u_j \leqq \cdot)/N_k$ with $F_k(0) = \tau$

: the (population) distribution function of u_i

Quantile regression estimator of $\boldsymbol{\beta}$:

$$(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p) = \operatorname{argmin} \sum_{i=1}^{n_k} \rho_\tau(y_i - \boldsymbol{\phi}^T \mathbf{x}_i),$$

over all $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_p)^T$, where

$$\rho_\tau(u) = u(\tau - \mathrm{I}(u \leqq 0))$$

I: indicator function

$$I(u \leqq 0) = \begin{cases} 1 & u \leqq 0 \\ 0 & u > 0 \end{cases}$$

2. Asymptotic Normality

2.1 Assumptions

(A0) $\lim_{k\to\infty} \min n_k, N_k - n_k = \infty$

(A1) There is a sequence of functions $\{f_k\}$ such that

$$\lim_{k\to\infty} \left[\frac{F_{N_k}(0 + \delta_k) - F_{N_k}(0)}{\delta_k} - f_k(0) \right] = 0,$$

for any sequence $\{\delta_k\}$ of order $\sim O(n_k^{-1/2})$ and

$$0 < \inf_k f_k(0) \leqq \sup_k f_k(0) < \infty.$$

(A2) (i)There exists a positive definite matrix C_k such that

$$\lim_{k\to\infty} \left\| \frac{1}{N_k} \frac{N_k - n_k}{N_k - 1} X_{n_k}^T X_{n_k} - \mathrm{C}_k \right\|_{\max} = 0,$$

where $\|\mathrm{A}\|_{\max} = \max |a_{ij}|$.

(ii) $\lim_{k\to\infty} \left(\frac{\Sigma_{S^1_{k\tau}} (x_{lj} - \bar{x}_l)^2}{\Sigma_{j=1}^{N_k} (x_{1j} - \bar{x}_1)^2} \right) = 0$ where $S^l_{k\tau}$: the set of $|x_{lj} - \bar{x}_l| > \tau \sqrt{\frac{n_k}{N_k} (\frac{N_k - n_k}{N_k})\, \Sigma_{j=1}^{N_k} (x_{li} - \bar{x}_l)^2}$, $l = 1, \ldots, p$.

2.2 Main Result

Theorem 1. *Under the assumptions (A0), (A1) and (A2), we have*

$$\frac{\sqrt{n_k}}{f_k(0)} C_k^{-1/2} (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0},\ \tau(1-\tau)\mathbf{I}_{p+1}),$$

as $k \to \infty$.

References

Koenker, R. and G. Bassett (1978). *Econometrica* **46**, 33-50.

Kato, K., Kunitomo, N., and T. Masuda (2009). *Journal of The Japan Statistical Society*, **38**, 121-149 (in Japanese).

Knight, K. (1998). *The Annals of Statistics*, **26**, 755-770.

Pollard, D. (1991). *Econometric Theory*, **7**, 186-199.