Second Order Asymptotics for R-estimators and M-estimators for a Simple Linear Regression

M. Omelka

Keywords: M-estimator, R-estimator, von Mises expansion

Suppose that the observations $Y_1, \ldots, Y_N$ follow the simple regression model

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, N,$$

where $x_1, \ldots, x_N$ are known constants, $\epsilon_1, \ldots, \epsilon_N$ are i.i.d. errors with an unknown distribution function $F$ and $\alpha$ and $\beta$ are unknown parameters. We will be interested only in the estimation of the slope parameter $\beta$. One way how to estimate this parameter is to minimize the Jaeckel measure of dispersion $\sum_{i=1}^{N}(Y_i - x_i\beta)(R_i(b) - \frac{N+1}{2})$, where $R_1(b), \ldots, R_N(b)$ is the vector of ranks for random variables $Y_1 - x_1b, \ldots, Y_N - x_Nb$. The resulting estimator $T_R$ can be calculated as the weighted median of the set of pairwise slopes $\frac{Y_i - Y_j}{x_i - x_j}$, where each slope is assigned weight proportional to $|x_i - x_j|$. It is well known that under some mild conditions this ‘generalized’ Hodges-Lehmann estimator $R_N$ admits the representation

$$Q_N(T_R - \beta) = \frac{1}{\gamma} \sum_{i=1}^{N} x_i(F(e_i) - \frac{1}{2}) + R_N, \quad \gamma = \int_{-\infty}^{+\infty} f(x)^2 dx, \quad Q_N^2 = \sum_{i=1}^{N} x_i^2$$

and the remainder term $R_N$ is under some appropriate conditions on $x_i$ and $F$ of order $O_P(N^{-\frac{1}{2}})$ (see Jurečková, Sen (1996)). We will find the asymptotic representation of $\sqrt{N}R_N$. The key tool for derivation of this von Mises expansion up to the second term will be Theorem 2.1. of Jurečková (1973).

Analogously, we get a similar second order asymptotic representation for the M-estimator. At first we will suppose the intercept $\alpha$ in (1) to be zero. In this case the M-estimator is defined as a solution of the equation

$$\sum_{i=1}^{N} x_i\psi(Y_i - b x_i).$$

Similarly as for R-estimators, under some appropriate conditions the first order representation of M-estimator is

$$Q_N(T_M - \beta) = \frac{1}{\gamma} \sum_{i=1}^{N} x_i(\psi(e_i) - \frac{1}{2}) + R_N, \quad \gamma = \int_{-\infty}^{+\infty} \psi'(x)f(x)dx.$$

And we will again find the asymptotic representation of $\sqrt{N}R_N$. Particularly, we will be interested in the special case $\psi(x) = cF(x)$, where $F$ is the distribution function of errors in (1). For this choice of $\psi$ the estimators $T_R$ and $T_M$ are first order asymptotic equivalent, i.e. $\sqrt{N}(T_R - T_M) = o_P(1)$. We will find that under some smoothness conditions on $\psi$ and $F$ the difference $\sqrt{N}(T_R - T_M)$ is of order $O_P(N^{-\frac{1}{2}})$ and derive an asymptotic representation of this difference. This representation also implies that the order $O_P(N^{-\frac{1}{2}})$ is exact. Of course, in practice we do not know the true distribution function $F$. But for the choice $\psi(x) = c(F_0(x) - \frac{1}{2})$ a suitably normalized difference $T_R - T_M$ can be used as a simple measure of goodness of fit, that the errors in the regression (1) have just the distribution $F_0$. The investigation of using our asymptotic results in the goodness of fit testing is in progress.
Unfortunately, unlike the $R$-estimator the $M$-estimator defined by (2) is neither location invariant nor scale equivariant. Making the $M$-estimator location invariant by adding the intercept presents only minor difficulties. No extra conditions are needed, but the second order term in von Mises expansion will be a little more complex. On the contrary, studentization of the $M$-estimator will require not only new stringent conditions on $\psi$ and $F$ but it also complicates the second order term in von Mises expansion in a nontrivial way, especially when the distribution function $F$ of errors in (1) is asymmetric. These results, even in the most simple case of the $M$-estimator defined by (2), generalize the results of Jurečková, Sen (1990).