ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 20 МАТЕМАТИКИ Выпуск 4
2013

Е. В. Шаталин, **А. П. Ковалевский** (Новосибирск, НГТУ, ИМ СО РАН). **Асимптотика эмпирического моста в линейных регрессионных моделях, построенных по порядковым статистикам.**

Наиболее простая модель линейной регрессии имеет вид

$$Y_{ni} = \theta X_{ni} + \varepsilon_i, \quad i = 1, \ldots, n,$$

где X_{ni} (регрессор) обычно предполагаются неслучайными величинами, $\theta \in \mathbf{R}$ — неизвестный параметр, $\varepsilon_1, \ldots, \varepsilon_n$ (регрессионные ошибки) — независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и конечной ненулевой дисперсией σ^2.

Параметр θ обычно оценивают по методу наименьших квадратов $\widehat{\theta} = \overline{XY}/\overline{X^2}$. Далее строятся прогнозные значения $\widehat{Y}_{ni} = \widehat{\theta} X_{ni}$. Остатками линейной регрессии называют случайные величины $\widehat{\varepsilon}_{ni} = Y_{ni} - \widehat{Y}_{ni}$.

Приведем определение эмпирического моста, это кусочно-линейная случайная ломаная $\widehat{Z}_n = \{\widehat{Z}_n(t),\ 0 \leqslant t \leqslant 1\}$ с узлами в точках

$$\left(\frac{k}{n},\ \frac{\widehat{\Delta}_{nk} - \frac{k}{n}\widehat{\Delta}_{nn}}{\sqrt{\widehat{\sigma^2} n}} \right),$$

где $\widehat{\Delta}_{nk} = \widehat{\varepsilon}_{n1} + \cdots + \widehat{\varepsilon}_{nk}$, $k = 1, \ldots, n$, $\widehat{\Delta}_{n0} = 0$, $\widehat{\sigma^2} = \overline{\widehat{\varepsilon}^2} - (\overline{\widehat{\varepsilon}})^2$. При условии сходимости $\widehat{\sigma^2} \xrightarrow{p} \sigma^2$ при $n \to \infty$, отыскание предельного процесса для эмпирического моста сводится к отысканию предельного процесса для случайной ломаной Z_n, построенной по точкам

$$\left(\frac{k}{n},\ \frac{\widehat{\Delta}_{nk}}{\sigma\sqrt{n}} \right).$$

В случае, когда в качестве регрессора используется вектор значений неслучайной гладкой функции в равноотстоящие моменты времени, предельный процесс для Z_n изучен MacNeill в [1]. Мы будем предполагать, что в качестве регрессора выступает набор $\{\xi_{1:n}, \ldots, \xi_{n:n}\}$ порядковых статистик, построенных по выборке из некоторого распределения, т. е. $X_i = \xi_{i:n}$. Случайные величины $\xi_1 \cdots \xi_n$ предполагаются независимыми, одинаково распределенными с функцией распределения F и не зависящими от случайных величин $\varepsilon_1 \cdots \varepsilon_n$.

Обозначим $GL_F(t) = \int_0^t F^{-1}(s)\,ds$ — *теоретическая обобщенная кривая Лоренца* (см. в [2]), $GL_F^0(t) = GL_F(t) - tGL_F(1)$, где $F^{-1}(s) = \sup\{x :\ F(x) < s\}$.

Через $\Longrightarrow$ будем обозначать слабую сходимость в пространстве $C(0,1)$, снабженном равномерной метрикой. Тогда справедлива

Теорема 1. *Если* $0 < \mathbf{E}\,\xi_1^2 < \infty$, *то справедливы следующие утверждения*:

1) $Z_n \Longrightarrow Z_F$, *где* Z_F — *центрированный гауссовский процесс с ковариационной функцией*

$$K_F(t,s) = \min\{t,s\} - GL_F(s)GL_F(t)/\mathbf{E}\,\xi_1^2, \quad s,t \in [0,1].$$

2) $\widehat{Z}_n \Longrightarrow Z_F^0$, *где* Z_F^0 *— центрированный гауссовский процесс с ковариационной функцией*

$$K_F^0(t,s) = \min\{t,s\} - ts - GL_F^0(s)GL_F^0(t)/\mathbf{E}\,\xi_1^2, \quad s,t \in [0,1].$$

Аналогичный результат можно получить и для двухпараметрической модели

$$Y_{ni} = a + bX_{ni} + \varepsilon_i, \quad i = 1,\ldots,n,$$

Для неизвестных параметров a и b также обычно используются оценки наименьших квадратов.

Теорема 2. *Если* $0 < \mathbf{Var}\,\xi_1 < \infty$, *тогда* $Z_n \Longrightarrow \widetilde{Z}_F^0$, $\widehat{Z}_n \Longrightarrow \widetilde{Z}_F^0$, *где* $\widetilde{Z}_F^0$ *центрированный — центрированный гауссовский процесс с ковариационной функцией*

$$\widetilde{K}_F^0(t,s) = \min\{t,s\} - ts - \frac{GL_F^0(t)GL_F^0(s)}{\mathbf{Var}\,\xi_1}, \quad t,s \in [0,1].$$

СПИСОК ЛИТЕРАТУРЫ

1. *MacNeill I. B.* Limit processes for sequences of partial sums of regression residuals. — Ann. Probab., 1978, v. 6, № 4, p. 695–698.
2. *Gastwirth J. L.* A general definition of the Lorenz curve. — Econometrica, 1971, v. 39, p. 1037–1039.