Секция «Теория вероятностей и математическая статистика»

Предельный процесс для последовательных сумм остатков линейной пуассоновской регрессии

Научный руководитель – Ковалевский Артем Павлович

Веселов Владислав Евгеньевич

Студент (бакалавр)

Новосибирский государственный университет, Механико-математический факультет, Новосибирск, Россия

E-mail: v.veselov1@g.nsu.ru

В работе [1] рассматривались последовательности частичных сумм регрессионных остатков модели линейной регресии, для которых было получено коварционное ядро следующего вида

$$K(s,t) = \mathbf{E}B_f(s)B_f(t) = \min(s,t) - \int_0^s \int_0^t g(x,y)dxdy,$$

где $\{B_f(t), t \in [0,1]\}$ – предельный гауссовский процесс,
$B_f(t) = B(t) - B(1)\int_0^t g(x,1)dx + \int_0^1 B(x)\left\{\int_0^t \frac{\partial}{\partial x} g(x,y)dy\right\}dx$,
$\{B(t), t \in [0,1]\}$ – стандартный винеровский процесс,
$g(x,y) = f(x)^T F^{-1} f(y),\ f(x) \in \mathbb{R}^p,\ F = \left(\int_0^1 f_r(t)f_s(t)dt\right)_{r,s=1}^p$.

Мы хотим получить похожий результат, но для пуассоновской регрессии, и на основе значений частичных сумм регрессионных остатков построить критерий, который бы говорил о правильности выбранной модели в целом. Для этого была сформулирована и доказана теорема 1, рассмотрены частные случаи при разных функциях $g_i(t),\ t \in [0,1],\ i = 1,\dots,m$.

Теорема 1 Пусть $Y_1,\dots,Y_n$ — независимые пуассоновские величины с параметрами $\lambda_1,\dots,\lambda_n$, определяемыми равенством $\lambda_i = \sum_{j=1}^m \theta_j g_j(i/n)$, где g_j — известные функции, а θ_j — неизвестные параметры. Пусть существует $\varepsilon gt;0$ такой, что для любого $t \in [0,1]$ выполнено $\sum_{j=1}^m \theta_j g_j(t) \geq \varepsilon$. Кроме того, пусть функции g_j интегрируемы по Риману на [0, 1], и матрица

$$G = \left(\int_0^1 g_i(t)g_j(t)dt\right)_{i,j=1}^m$$

невырождена.
Обозначим $\overline{Y} = \sum_{i=1}^n Y_i/n,\ \ \widehat{\lambda}_i = \sum_{j=1}^m \widehat{\theta}_j g_j(i/n),\ i = 1,\dots,n$, где $\widehat{\theta}_j$ – оценки θ_j методом наименьших квадратов, $\widehat{\varepsilon}_i = Y_i - \widehat{\lambda}_i$, $i = 1,\dots,n$, $\widehat{S}_0 = 0,\ \widehat{S}_k = \sum_{i=1}^k \widehat{\varepsilon}_i,\ k = 0,\dots,n$. Тогда случайная ломаная $\widehat{Z}_n = \{\widehat{Z}_n(t),\ 0 \leq t \leq 1\}$, построенная по точкам $\left(\frac{k}{n},\ \frac{\widehat{S}_k}{\sqrt{n\overline{Y}}}\right)$, сходится по распределению в равномерной метрике на [0, 1] к центрированному гауссовскому процессу
$\widehat{Z} = \{\widehat{Z}(t),\ 0 \leq t \leq 1\}$ с ковариационной функцией

$$K(s,t) = \mathbf{E}\widehat{Z}(s)\widehat{Z}(t) =$$

$$= \frac{1}{a}\Bigg[\Big(\mathbf{h}(\min(s,t)) - \mathbf{h}(s)G^{-1}G(t) - \mathbf{h}(t)G^{-1}G(s)\Big)\theta +$$

$$+ \mathbf{h}(s)G^{-1}\left(\int_0^1 g_{i_1}(u)g_{i_2}(u)\mathbf{g}(u)\,du\,\theta\right)_{i_1,i_2=1}^m G^{-1}\mathbf{h}^T(t)\Bigg]$$

где $G(t) = \left(\int_0^t g_i(s)g_j(s)ds\right)_{i,j=1}^m$ и $\mathbf{g}(t) = (g_1(t),\dots,g_m(t))$,
$\mathbf{h}(t) = \int_0^t \mathbf{g}(s)\,ds$, $a = \mathbf{h}(1)\theta$, $\theta = (\theta_1,\dots,\theta_m)^T$.

Источники и литература

1) MacNeill I. B., 1978. Limit processes for sequences of partial sums of regression residuals. Ann. Prob. 6, 695–698. Zbl 0377.60028