Секция «Математика и механика»

Скорость сходимости оценок параметров линейной логистической регрессии с изменяющимися коэффициентами

Хапланов Арсений Юрьевич

Аспирант

Московский государственный университет имени М.В. Ломоносова, Механико-математический факультет, Москва, Россия

E-mail: khaplanova@gmail.com

Логистическая регрессия является одной из самых широко применяемых моделей в практических исследованиях. Нас будет интересовать использование в логистической регрессии линейной функции с изменяющимися коэффициентами (см. [1]). При этом мы не предполагаем, что значения предикторов ограничены. Пусть для некоторого случайного вектора $(Y, X^T, Z^T, U)^T$ со значениями в $\{0,1\} \times \mathbb{R}^{p+d+1}$ существуют такие функции $\alpha : \mathbb{R} \to \mathbb{R}$, $\beta : \mathbb{R} \to \mathbb{R}^p$ и вектор $\gamma \in \mathbb{R}^d$, что для всех $x \in \mathbb{R}^p$, $z \in \mathbb{R}^d$, $u \in \mathbb{R}$ справедливо следующее соотношение (см. [2]):

$$\ln\left(\frac{\mathbb{P}(Y=1|X=x, Z=z, U=u)}{\mathbb{P}(Y=0|X=x, Z=z, U=u)}\right) = -\alpha(u) - \beta^T(u)x - \gamma_0^T z,$$

где T обозначает транспонирование. По наблюдениям $(Y_q, X_q^T, Z_q^T, U_q)^T$, $q = 1, \ldots, n$, которые независимы и имеют такое же распределение, как $(Y, X^T, Z^T, U)^T$, должным образом строится оценка $\widehat{\theta}_n(u) = (\widehat{\alpha}_n(u), \widehat{\beta}_n^T(u), \widehat{\gamma}_n)^T$ вектора $\theta_0(u) = (\alpha(u), \beta^T(u), \gamma)^T$. Если для последовательности случайных величин $(\xi_n)_{n=1}^\infty$, $a_n \in \mathbb{R}_+$ $(n \in \mathbb{N})$ и любого $\varepsilon > 0$ найдется такое число $D > 0$, что $\mathbb{P}(|\xi_n| > Da_n) \le \varepsilon$, $n \in \mathbb{N}$, то пишем $\xi_n = O_p(a_n)$. Далее мы также используем евклидову норму $\|\cdot\|$ (для соответствующих евклидовых пространств).

Теорема. *Предположим, что дана некоторая числовая последовательность $(h_n)_{n=1}^\infty$ такая, что $n^{1/3}h_n \to \infty$ и $h_n \to 0+$, когда $n \to \infty$. Тогда при определенных условиях регулярности распределения случайного вектора $(Y, X^T, Z^T, U)^T$ для любого фиксированного $u \in \mathbb{R}$ имеют места соотношения:*

$$\left\|(\widehat{\alpha}_n(u), \widehat{\beta}_n^T(u))^T - (\alpha(u), \beta^T(u))^T\right\| = O_p\left(h_n^2 + \frac{1}{\sqrt{nh_n}}\right),$$

$$\|\gamma_n - \gamma\| = O_p\left(n^{-1/2} + h_n^4\right).$$

Отметим, что метод, используемый для доказательства этой теоремы, может быть использован при исследовании ряда других моделей. К ним относится, например, обобщение квантильной регрессии.

Литература

1. T. Hastie, R. Tibshirani. Varying-coefficient models // J. of the Royal Stat. Society. Series B (Methodological). 1993. vol. 55. No 4. p. 757–796.

2. J. Li, C. Zhang, K.A. Doksum, E.V. Nordheim. Simultaneous confidence intervals for semiparametric logistics regression and confidence regions for the multi-dimensional effective dose // Statistica Sinica. 2010. vol.20. p. 637–659.