

Скорость сходимости оценок параметров линейной логистической регрессии с изменяющимися коэффициентами

Арсений Хапланов

(Московский государственный университет им М.В. Ломоносова)

XXI Международная конференция студентов, аспирантов и молодых
учёных «Ломоносов»–2014
Москва, 9 апреля

Планирование доклада

- ▶ Введение
- ▶ Основные результаты
- ▶ Основные идеи доказательства
- ▶ Список литературы

Логистическая регрессия

Пусть дана пара случайных величин (Y, X) , принимающих значения в множестве $\{0, 1\} \times \mathbb{R}^p$. Предположим, что функция $h : \mathbb{R}^p \rightarrow \mathbb{R}$ такова, что для всех $x \in \mathbb{R}^p$ имеет место равенство

$$P(Y = 1 | X = x) = \frac{\exp\{h(x)\}}{1 + \exp\{h(x)\}}.$$

R. Gramacy, A. Genkin, D. Böhning, E. Makalic, L. Meier,
S. Vishwanathan, J. Friedman, V. Chernozhukov, A. DeMaris,
K.A. Doksum, D. Ruppert, J. Wakefield, Y. Huang.

Модели с варьирующимися коэффициентами

Пусть дана случайная величина Y и случайный вектор $(R_1, \dots, R_p, X_1, \dots, X_p)^T$, где $p \in \mathbb{N}$. Предположим, что распределение Y зависит от *линейного предиктора* η , для которого выполняется следующее равенство:

$$\eta = \beta_0 + X_1\beta_1(R_1) + \dots + X_p\beta_p(R_p),$$

где β_0 является константой, а $\beta_i(\cdot)$, $i = 1, \dots, p$, неизвестными функциями.

Области применения полупараметрического подхода

- ▶ Линейная регрессия
- ▶ Логистическая регрессия
- ▶ Квантильная регрессия
- ▶ Задачи с цензурированием данных

D. Hosmer, S. Hossain, R.-C. Hwang, B. Kai, T. Hoshino,
B.J. Reich, R. Zhang, A. Antoniadis, H. He. J. Wakefield.

Описание модели

Предположим, что дан набор случайных векторов (Y, X, Z, U) , принимающих значения в множестве $\{0, 1\} \times \mathbb{R}^p \times \mathbb{R}^d \times \mathbb{R}$. Пусть существуют функции $\alpha : \mathbb{R} \rightarrow \mathbb{R}$, $\beta : \mathbb{R} \rightarrow \mathbb{R}^p$ и вектор $\gamma_0 \in \mathbb{R}^d$ такие, что следующее равенство выполняется почти всюду:

$$P(Y = 1|X, Z, U) = \frac{\exp\{-\alpha(U) - \beta(U)^\top X - \gamma_0^\top Z\}}{1 + \exp\{-\alpha(U) - \beta(U)^\top X - \gamma_0^\top Z\}}.$$

Построение оценки

Воспользуемся локально-линейным приближением неизвестных функций в точке u_0 .

$$\alpha(u) \approx \alpha(u_0) + \alpha'(u_0)(u - u_0) \text{ и } \beta(u) \approx \beta(u_0) + \beta'(u_0)(u - u_0).$$

Нас будет интересовать оценка вектора параметров $\theta_0(u_0) := (\alpha(u_0), \alpha'(u_0), \beta(u_0)^\top, \beta'(u_0)^\top, \gamma_0^\top)^\top$. Введем симметричную ядерную функцию плотности с ограниченным носителем $K(x)$. Обозначим $\mu_i = \int_{\mathbb{R}} x^i K(x) dx$.

Построение первой оценки

Функция правдоподобия с учетом локального сглаживания принимает следующий вид:

$$\begin{aligned} \mathcal{L}_n(a_0, a_1, b_0, b_1, c; u_0) = \\ = \sum_{q=1}^n [I(Y_q = 1)L_q + \log(1 + \exp\{-L_q\})] K\left(\frac{U_q - u_0}{h_n}\right), \end{aligned}$$

где

$$L_q = a_0 + a_1(U_q - u_0) + b_0^\top X_q + b_1^\top X_q(U_q - u_0) + c^\top Z_q, \quad q = 1, \dots, n.$$

Построение первой оценки

$$\begin{aligned}\tilde{\theta}_n(u_0) &:= (\tilde{\alpha}_{0,n}, \tilde{\alpha}_{1,n}, \tilde{\beta}_{0,n}^\top, \tilde{\beta}_{1,n}^\top, \tilde{\gamma}_n^\top)^\top [u_0] = \\ &= \operatorname{argmin} \mathcal{L}_n(a_0, a_1, b_0, b_1, c; u_0)\end{aligned}$$

где $a_0, a_1 \in \mathbb{R}$, $b_0, b_1 \in \mathbb{R}^p$ и $c \in \mathbb{R}^d$. Важно отметить, что $\tilde{\gamma}_n$ зависит от параметра u_0 .

Условия

- A1. Коэффициенты $\alpha(u)$ и $\beta(u)$ имеют равномерно ограниченные первые и вторые производные по u для всех $u \in \mathbb{R}$.
- A2. Плотность $f(x, z, u)$ имеет равномерно ограниченные первые две производные по u .
- A3. Случайные вектора X и Z обладают конечными третьими абсолютными моментами.
- A4. Плотность $f_U(u)$ непрерывна для всех $u \in \mathbb{R}$.

Вспомогательные обозначения

Определим случайную матрицу D и функцию $R : \mathbb{R}^p \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$.

$$D = \begin{pmatrix} 1 & 0 & X^\top & \bar{0}_p^\top & Z^\top \\ 0 & \mu_2 & \bar{0}_p^\top & \mu_2 X^\top & \bar{0}_d^\top \\ X & \bar{0}_p & XX^\top & 0_{p,p} & XZ^\top \\ \bar{0}_p & \mu_2 X & 0_{p,p} & \mu_2 XX^\top & 0_{p,d} \\ Z & \bar{0}_d & ZX^\top & 0_{d,p} & ZZ^\top \end{pmatrix},$$

$$R(x, z; u) = \frac{\exp\{-\alpha(u) - \beta(u)^\top x - \gamma_0^\top z\}}{(1 + \exp\{-\alpha(u) - \beta(u)^\top x - \gamma_0^\top z\})^2}.$$

Теорема 1

Пусть $(h_n)_{n \in \mathbb{N}}$ — последовательность положительных чисел такая, что $nh_n \rightarrow \infty$ и $h_n \rightarrow 0$ при $n \rightarrow \infty$. Предположим, что выполняются условия A1-A4. Тогда, если для некоторого $u_0 \in \mathbb{R}$ матрица $W(u_0) = \mathbb{E}(R(X, Z, U)D|U=u_0)$ не вырождена, то имеет место соотношение

$$\left\| \theta_0(u_0) - \tilde{\theta}_n(u_0) \right\| = O_p(c_n), \quad \text{при } n \rightarrow \infty,$$

где $c_n = \left(\frac{1}{\sqrt{nh_n}} + h_n^2 \right)$.

Построение второй оценки

Введем новую функцию правдоподобия $\hat{L}_n : \mathbb{R}^d \rightarrow \mathbb{R}$

$$\begin{aligned} \hat{L}_n(c) = & \sum_{q=1}^n \mathbb{I}(Y_q = 1) \left(\tilde{\alpha}_{0,n}(U_q) + \tilde{\beta}_{0,n}(U_q)^\top X_q + c^\top Z_q \right) + \\ & + \ln \left(1 + \exp \left\{ -\tilde{\alpha}_{0,n}(U_q) - \tilde{\beta}_{0,n}(U_q)^\top X_q - c^\top Z_q \right\} \right). \end{aligned}$$

Определим аналогично $\hat{\gamma}_n = \operatorname{argmin} \hat{L}_n(c)$, где минимум берется по всем $c \in \mathbb{R}^d$.

A5. Для плотности $f_U(u)$ случайной величины U найдется такое число $p > 0$, что $\int_{\mathbb{R}} f_U(u)^{1-p} du < \infty$.

Теорема 2

Пусть выполняются условия предыдущей теоремы и условие A5. Предположим, что существует такое положительное число ϱ , что для всех $v \in \mathbb{R}^{2+2p+d}$ и $u \in \mathbb{R}$ выполняется неравенство $v^\top W(u)v \geq \varrho \|v\|^2$. Тогда если $nh_n^8 = O(1)$ и $\ln^2(h_n)/(nh_n) = O(1)$, то при $n \rightarrow \infty$ выполняется соотношение.

$$\|\hat{\gamma}_n - \gamma_0\| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Представленный результат позволяет улучшить исследования, проведенные в работе J. Li, C. Zhang, K.A. Doksum и E.V. Nordheim. В ней строится только первая оценка. Поэтому предельные теоремы посвящены только следующему вектору разностей.

$$\sqrt{nh_n} \begin{pmatrix} \gamma_0 z_1 - \tilde{\gamma}_n z_1 + \alpha_0(u_1) - \tilde{\alpha}_{0,n}(u_1) \\ \vdots \\ \gamma_0^\top z_L - \tilde{\gamma}_n^\top z_L + \alpha_0(u_L) - \tilde{\alpha}_{0,n}(u_L) \end{pmatrix}$$

R.-C. Hwang, R. Zhang, C.J. Huang, J. Wakefield, N. Müller-Lenke, K. Bendfeldt, . . .

Продолжение исследований

Следствие

Пусть выполняются условия последней теоремы. Тогда существует невырожденная матрица Σ такая, что при $n \rightarrow \infty$ имеет место предельное соотношение:

$$\sqrt{n}(\gamma_0 - \hat{\gamma}_n) \xrightarrow{law} V \sim N(0, \Sigma).$$

Лемма (Y. Mack, B. Silverman)

Пусть дана последовательность н.о.р.с.в. $(\xi, \zeta), (\xi_q, \zeta_q)_{q=1}^{\infty}$.

Предположим, что для некоторого $r \geq 2$ выполняется

$\mathbb{E} |\zeta|^r < \infty$ и $\sup_{x \in \mathbb{R}} \int_{\mathbb{R}} |y|^r f_{\xi, \zeta}(x, y) dy < \infty$. Тогда если $n^{2\varepsilon-1} h_n \rightarrow \infty$ для некоторого $\varepsilon < 1 - r^{-1}$, то при $n \rightarrow \infty$ имеем

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{nh_n} \sum_{q=1}^n \left[K \left(\frac{\xi_q - x}{h_n} \right) \zeta_q - \mathbb{E} \left(K \left(\frac{\xi - x}{h_n} \right) \zeta \right) \right] \right| = O_p(r_n),$$

Где $r_n = [\ln(1/h_n)/(nh_n)]^{1/2}$. Если также для некоторого $\lambda > 0$ выполняется неравенство $\sum_n h_n^\lambda < \infty$, то сходимость по вероятности может быть заменена на сходимость почти всюду.

Идея доказательства первой теоремы

Разложение функции $L_n(\theta)$ по формуле Тейлора до второго слагаемого с остаточным членом в форме Лагранжа дает

$$L_n(\theta) = L_n(\theta_0(u)) + S_n(\theta_0(u_0), u_0)^\top (\theta - \theta_0(u_0)) + \\ + \frac{1}{2}(\theta - \theta_0(u_0))^\top W_n(\theta_0(u_0), u_0)(\theta - \theta_0(u_0)) + \frac{1}{3}F_n((\theta - \theta_0(u_0), \theta_0(u_0)).$$

Введем вспомогательное обозначение:

$$H(h) = \text{diag}\left(1, h, \underbrace{1, \dots, 1}_p, \underbrace{h, \dots, h}_p, \underbrace{1, \dots, 1}_d\right), \quad h \in \mathbb{R}.$$

Определим матрицу Q следующим образом

$$Q = \text{diag} \left(1, 0, \underbrace{1, \dots, 1}_p, \underbrace{0, \dots, 0}_p, \underbrace{0, \dots, 0}_d \right).$$

Лемма

Пусть выполняются условия Теоремы 2. Тогда для построенной оценки $\tilde{\theta}_n(u)$ выполняется следующее соотношение.

$$H_n Q (\tilde{\theta}_n(U) - \theta_0(U)) = \frac{Q}{nh_n f_U(U)} W^{-1}(U) H_n^{-1} S_n(\theta_0(U); U) + O_p(\eta_n),$$

где $\eta_n = (h_n^2 + r_n)^2$.

Теорема (К. Kato (2009))

Пусть функции $f_n(x, \tau)$ являются выпуклой по x для всех τ и ограниченной по τ для всех x . Предположим, что дана функция $g_n(x, \tau) = -x^\top W_n(\tau) + \frac{1}{2}x^\top Q(\tau)x$. Пусть также для всех $x \in \mathbb{R}^p$

$$\sup_{\tau \in T} |f_n(x, \tau) - g_n(x, \tau)| \xrightarrow{P} 0,$$

и для любого $\varepsilon > 0$ существует константа $M > 0$ такая, что

$$\limsup_{n \rightarrow \infty} P \left(\sup_{\tau \in T} \|W_n(\tau)\| > M \right) \leq \varepsilon.$$

Тогда

$$x_n(\tau) = Q(\tau)^{-1} W_n(\tau) + r_n(\tau),$$

где $x_n(\tau) = \operatorname{argmin}_x f_n(x, \tau)$ и $\sup_{\tau} \|r_n(\tau)\| = o_p(1)$.



D. Böhning, (1992). Multinomial logistic regression algorithm. Annals of the Institute of Statistical Mathematics, 44(1), pp. 197–200.



N. Ding, S. Vishwanathan, (2013). T-Logistic Regression for Binary and Multiclass Classification. Journal of Machine Learning Research, 05, pp. 1–55.



J. Friedman, T. Hastie, R. Tibshirani, (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the



A. Genkin, D.D. Lewis, D. Madigan, (2007). Large-Scale Bayesian Logistic Regression for Text Categorization. Technometrics, 49(3), pp. 291–304.

-  R.B. Gramacy, N.G. Polson, (2012). Simulation-based Regularized Logistic Regression. *Bayesian Analysis*, 7(3), pp. 567–590.
-  T. Hastie, R. Tibshirani, (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4), pp. 757–796.
-  D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, (2013). *Applied Logistic Regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
-  Y. Huang, M. Pepe, Z. Feng, Z. (2013). Logistic regression analysis with standardized markers. *The Annals of Applied Statistics*, 7(3), pp. 1–23.
-  K. Kato, (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100(8), pp. 1816–1829.

-  A. Khaplanov, (2013). Asymptotic normality of the estimation of the multivariate logistic regression. Informatics and Applications, 7(2), pp. 69–74.
-  A. Khaplanov, (2013). The study of the multivariate logistic regression with increasing number of covariates. Ulm.
-  A. Khaplanov, (2014). Semiparametric partially linear varying coefficient logistic regression, in preparation.
-  Y.P. Mack, B.W. Silverman, (1985). Weak and strong uniform consistency of kernel regression estimates. Prob. Theory and Related Fields, 61(3), pp. 405–415.
-  E. Makalic, D. Schmidt, (2012). MML logistic regression with translation and rotation invariant priors. AI 2012: Advances in Artificial Intelligence, (2), pp. 878–889.

-  L. Meier, S. van de Geer, P. Bühlmann, (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society. Series B (Methodological), 70(1), pp. 53–71.
-  D. Pollard, (1991). Asymptotics for least absolute deviation regression estimators. Econometric Theory, 7(02), pp. 186–199.
-  X. Tan, Y. Zhang, S. Tang, J. Shao, G. Wu, Y. Zhuang, (2013). Logistic Tensor Regression for Classification. Lecture Notes in Computer Science, 7751, pp. 573–581.