Math-Net.Ru

Общероссийский математический портал

Е. В. Бурнаев, А. А. Зайцев, В. Г. Спокойный, Теорема Бернштейна–фон Мизеса для регрессии на основе гауссовских процессов, *УМН*, 2013, том 68, выпуск 5, 179–180

DOI: <https://doi.org/10.4213/rm9546>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением <http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.172.240.79

26 апреля 2023 г., 18:14:38

Теорема Бернштейна–фон Мизеса для регрессии на основе гауссовских процессов

Е. В. Бурнаев, А. А. Зайцев, В. Г. Спокойный

Регрессия на основе гауссовских процессов (ГП) широко используется в приложениях [1]. Теоретические результаты о ее свойствах получены лишь в очень специальных случаях [2], [3]. Необходим теоретический анализ метода ГП-регрессии для данных большой размерности при ограниченном объеме выборки и неточной спецификации модели. В методе ГП-регрессии предполагается, что неизвестная функция порождена случайным процессом. Значит, теоретический анализ этого метода естественно проводить с помощью байесовского подхода, в котором центральной является теорема Бернштейна–фон Мизеса (БФМ) о близости апостериорного распределения параметров регрессии к соответствующему нормальному распределению. Классические асимптотические методы [4] не позволяют исследовать свойства апостериорного распределения в случае растущей размерности и ограниченного объема выборки, поэтому требуются новые подходы, основанные на современной теории эмпирических процессов, см. [5]. В работе получена неасимптотическая граница точности аппроксимации в теореме БФМ.

Построение ГП-регрессии проводится следующим образом. Задана выборка значений неизвестной функции: $\mathbf{D} = (X, \mathbf{y}) = \{\mathbf{x}_i, y_i = y(\mathbf{x}_i)\}_{i=1}^n$, $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^d$. Предполагается, что функция $y(\mathbf{x})$ является реализацией ГП с нулевым средним, при этом ковариация между значениями ГП определяется некоторой ковариационной функцией $\text{cov}(y(\mathbf{x}), y(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}')$. Тогда распределение вектора значений $\mathbf{y}$ будет нормальным: $\mathbf{y} \propto \mathcal{N}(\mathbf{0}, K)$, где $K = \{k(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$ – ковариационная матрица.

Ковариационную функцию $k(\mathbf{x}, \mathbf{x}')$ моделируют с помощью некоторого параметрического семейства $k_\theta(\mathbf{x}, \mathbf{x}')$, $\theta \in \Theta \subset \mathbb{R}^p$, где Θ – компактное множество. Функция регрессии задается апостериорным средним $\mathbf{E}_\theta\{y(\mathbf{x}) \mid \mathbf{D}\} = \mathbf{k}_\theta^\top(\mathbf{x}) K_\theta^{-1} \mathbf{y}$, где $\mathbf{k}_\theta(\mathbf{x}) = (k_\theta(\mathbf{x}, \mathbf{x}_1), \dots, k_\theta(\mathbf{x}, \mathbf{x}_n))^\top$, $\mathbf{E}_\theta\{\cdot\}$ – математическое ожидание по распределению $\mathcal{N}(\mathbf{0}, K_\theta)$, $K_\theta = \{k_\theta(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$. Для оценки вектора параметров θ часто используют оценку максимального (квази-) правдоподобия $\tilde{\theta} = \arg\max_{\theta \in \Theta} L(\theta)$, где логарифм (квази-) правдоподобия равен $L(\theta) = -[n \log(2\pi) + \log \det(K_\theta) + \mathbf{y}^\top K_\theta^{-1} \mathbf{y}]/2$.

Пусть задано априорное распределение $\Pi(d\theta)$ вектора параметров θ . Тогда апостериорное распределение $\text{Law}(\theta \mid \mathbf{D}) \propto \exp\{L(\theta)\} \Pi(d\theta)$ описывает условное распределение случайного вектора θ . Цель работы – изучить свойства распределения $\text{Law}(\theta \mid \mathbf{D})$. Мы рассматриваем случай неинформативного априорного распределения $\Pi(d\theta)$ и возможной неправильной спецификации модели, т. е., вообще говоря, $k(\mathbf{x}, \mathbf{x}') \notin \{k_\theta(\mathbf{x}, \mathbf{x}')\}$, $\theta \in \Theta$.

Обозначим через $\|\cdot\|_p$ p -норму матрицы, $I_n \in \mathbb{R}^{n \times n}$ – единичную матрицу, $\mathbf{E}\{\cdot\}$ и $\text{Var}\{\cdot\}$ – математическое ожидание и ковариационный оператор по распределению $\mathcal{N}(\mathbf{0}, K)$. Пусть выполнены предположения:

- (A1) $k_\theta(\mathbf{x}, \mathbf{x}')$ трижды непрерывно дифференцируема по $\theta \in \Theta$ при $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$;
- (A2) $\max\{\|K\|_2, \|K_\theta\|_2\} \leq \bar{\lambda} < \infty$, $\max\{\|K^{-1}\|_2, \|K_\theta^{-1}\|_2\} \leq \lambda_0 < \infty$, $\|\nabla_i K_\theta\|_2 \leq \lambda_1 < \infty$, $\|\nabla_{i,j}^2 K_\theta\|_2 \leq \lambda_2 < \infty$, $\|\nabla_{i,j,k}^3 K_\theta\|_2 \leq \lambda_3 < \infty$ при $\theta \in \Theta$, $i, j, k = \overline{1, p}$;
- (A3) центральная точка $\theta^* = \arg\max_{\theta \in \Theta} EL(\theta)$ существует;
- (A4) минимальные собственные числа матриц D_0^2/n и V_0^2/n больше $d_0 > 0$ и $v_0 > 0$ соответственно, где $D_0^2 = -\nabla^2 EL(\theta)|_{\theta=\theta^*}$ и $V_0^2 = \text{Var}\{\nabla L(\theta)\}|_{\theta=\theta^*}$.

Работа выполнена при поддержке ПреМоЛаб, МФТИ (грант Правительства РФ, дог. 11.G34.31.0073 и РФФИ (гранты № 13-01-12447_офи_m2 и 13-01-00521).

DOI: 10.4213/rm9546

Эти предположения аналогичны предположениям из [2] и выполняются, например, для стандартной [1] ковариационной функции $k_{\theta}(\mathbf{x}, \mathbf{x}') = \theta_2^2(e^{-\theta_1^2 \|\mathbf{x} - \mathbf{x}'\|_2^2} + \sigma^2 \delta_{\mathbf{x}, \mathbf{x}'})$, где дисперсия шума $\sigma^2 > 0$ фиксирована, $\|\cdot\|_2$ – евклидова норма, $\theta = (\theta_1, \theta_2) \in \Theta \subseteq [\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]$, $0 < \underline{\theta}_i < \bar{\theta}_i < \infty$ ($i = 1, 2$), $\delta_{\mathbf{x}, \mathbf{x}'}$ – символ Кронекера.

Пусть $\mathbf{C}$ – универсальная абсолютная константа, $E\{\cdot\}$ – математическое ожидание по распределению $\text{Law}(\boldsymbol{\vartheta} \mid \mathbf{D})$, $\bar{\boldsymbol{\vartheta}} = E\{\boldsymbol{\vartheta} \mid \mathbf{D}\}$, $\mathfrak{S}^2 = E\{(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})^\top \mid \mathbf{D}\}$, $\mathbf{C}_1 = (4/v_0^2)\mathbf{r}_0^2\bar{\lambda}^2\lambda_0^4(2\lambda_1^2\lambda_0 + \lambda_2)^2$, $\mathbf{C}_4 = 4\mathbf{r}_0^2\Lambda^2/(9d_0^2v_0)$, $\Lambda = 4\lambda_0^6\lambda_1^3\bar{\lambda}^3 + 5.5\lambda_0^4\bar{\lambda}^2\lambda_1\lambda_2 + \bar{\lambda}\lambda_3$, $\mathbf{r}_0 > 0$ – константа, определенная в оценке сверху для экспоненциального момента производной логарифма правдоподобия $L(\theta)$ (см. [6]).

ТЕОРЕМА (БфМ). Пусть выполнены условия (A1)–(A4) и $n \geq \max(\mathbf{C}_1 p, \mathbf{C}_4 p^4)$. Тогда найдутся β , явно определяемое константами из условий (A1)–(A4), и случайное событие Ω_n с $P(\Omega_n) \geq 1 - \mathbf{C}e^{-x_n}$, $x_n = \log n$, такие, что для $\tau_n = \beta/\sqrt{n}$

$$\|D_0(\bar{\boldsymbol{\vartheta}} - \tilde{\boldsymbol{\vartheta}})\|_2^2 \leq \mathbf{C}\tau_n(p + x_n), \quad \|I_p - D_0\mathfrak{S}^2D_0\|_\infty \leq \mathbf{C}\tau_n(p + x_n) \quad \text{на } \Omega_n. \quad (1)$$

При этом для любого измеримого множества $A \subset \mathbb{R}^p$ и $\gamma \propto \mathcal{N}(\mathbf{0}, I_p)$

$$P(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}}) \in A) \geq e^{-\mathbf{C}\tau_n(p+x_n)} \{P(\gamma \in A) - \mathbf{C}\tau_n(p+x_n)^{1/2}\} - \mathbf{C}e^{-x_n}, \quad (2)$$

$$P(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}}) \in A) \leq e^{\mathbf{C}\tau_n(p+x_n)} \{P(\gamma \in A) + \mathbf{C}\tau_n(p+x_n)^{1/2}\} + \mathbf{C}e^{-x_n}. \quad (3)$$

Идея доказательства теоремы состоит в получении ряда глобальных и локальных оценок для экспоненциальных моментов логарифма правдоподобия $L(\theta)$ и его производных и использовании результатов работ [5], [6].

Неравенства (1) показывают, что среднее значение $\bar{\boldsymbol{\vartheta}}$ и матрица ковариаций $\mathfrak{S}^2$ апостериорного распределения $\text{Law}(\boldsymbol{\vartheta} \mid \mathbf{D})$ близки к оценке максимума правдоподобия $\tilde{\boldsymbol{\vartheta}}$ и матрице D_0^{-2} соответственно, а неравенства (2) и (3) описывают близость апостериорного распределения $\text{Law}(\boldsymbol{\vartheta} \mid \mathbf{D})$ к соответствующему нормальному распределению.

Список литературы

- [1] M. Seeger, *Int. J. Neural Syst.*, **14**:2 (2004), 69–106. [2] B. Shaby, D. Ruppert, *J. Comput. Graph. Statist.*, **21**:2 (2012), 433–452. [3] K. V. Mardia, R. J. Marshall, *Biometrika*, **71**:1 (1984), 135–146. [4] И. А. Ибрагимов, Р. З. Хасьминский, *Асимптотическая теория оценивания*, Наука, М., 1979. [5] V. Spokoiny, *Ann. Statist.*, **40**:6 (2012), 2877–2909. [6] V. Spokoiny, *Bernstein–von Mises theorem for growing parameter dimension*, arXiv: 1302.3430.

Е. В. Бурнаев (E. V. Burnaev)
Datadvance; ИППИ РАН; МФТИ (ПреМоЛаб)
E-mail: burnaev@iitp.ru

Представлено В. М. Бухштабером
Принято редколлегией
17.08.2013

А. А. Зайцев (A. A. Zaytsev)
Datadvance; ИППИ РАН
E-mail: alexey.zaytsev@datadvance.net

В. Г. Спокойный (V. G. Spokoiny)
Weierstrass-Institute; Humboldt University, Berlin;
МФТИ (ПреМоЛаб)
E-mail: spokoiny@wias-berlin.de