

Свойства байесовской оценки параметров регрессии на основе гауссовских процессов*

А. А. ЗАЙЦЕВ

ООО «Датадванс», Москва,
Институт проблем передачи информации РАН
e-mail: alexey.zaytsev@datadvance.net

Е. В. БУРНАЕВ

ООО «Датадванс», Москва,
Институт проблем передачи информации РАН,
Лаборатория структурных методов
анализа данных в предсказательном моделировании
при Московском физико-техническом институте
(государственном университете)
e-mail: evgeny.burnaev@datadvance.net

В. Г. СПОКОЙНЫЙ

Институт прикладного анализа и стохастики
им. Вейерштрасса, Берлин, Германия,
Берлинский университет им. Гумбольдта, Германия,
Лаборатория структурных методов
анализа данных в предсказательном моделировании
при Московском физико-техническом институте
(государственном университете)
e-mail: spokoiny@wias-berlin.de

УДК 519.22

Ключевые слова: регрессия, гауссовские процессы, апостериорное распределение, теорема Бернштейна—фон Мизеса, ковариационная функция, неасимптотическая постановка.

Аннотация

В работе приведён ряд результатов о свойствах апостериорного распределения параметров ковариационной функции в регрессии на основе гауссовских процессов. Проведённые статистические эксперименты подтверждают полученные теоретические утверждения для широкого класса ковариационных функций, повсеместно используемых в приложениях.

Abstract

A. A. Zaytsev, E. V. Burnaev, V. G. Spokoiny, *Properties of the Bayesian parameter estimation of a regression based on Gaussian processes*, *Fundamentalnaya i prikladnaya matematika*, vol. 18 (2013), no. 2, pp. 53—65.

*Работа выполнена при поддержке Лаборатории структурных методов анализа данных в предсказательном моделировании, МФТИ, гранта правительства РФ 11.G34.31.0073, грантов РФФИ 13-01-12447 офи_м2 и 13-01-00521.

We consider the regression approach based on Gaussian processes and outline our theoretical results about the properties of the posterior distribution of the corresponding covariance function's parameter vector. We perform statistical experiments confirming that the obtained theoretical propositions are valid for a wide class of covariance functions commonly used in applied problems.

1. Введение

В настоящее время одним из наиболее популярных методов решения задачи восстановления (аппроксимации) неизвестной функции по выборке её значений является регрессия на основе гауссовских процессов [4, 7, 11].

В рамках этого подхода предполагается, что аппроксимируемая функция является реализацией гауссовского процесса, распределение которого полностью определяется априорно задаваемыми функцией математического ожидания и функцией ковариации. Считается, что ковариационная функция между значениями гауссовского процесса зависит только от точек, в которых эти значения получены. В таком случае в качестве прогноза значения аппроксимируемой функции в новой точке используют значение апостериорного среднего, а для оценки неопределённости этого прогноза — соответствующую апостериорную дисперсию, при этом апостериорное среднее и апостериорная дисперсия могут быть вычислены аналитически [11] и полностью определяются ковариационной функцией гауссовского процесса.

Обычно предполагают, что ковариационная функция принадлежит некоторому параметрическому семейству [11] (параметрическое предположение), и поэтому задача построения регрессии на основе гауссовских процессов сводится к задаче оценки параметров ковариационной функции.

На данный момент теоретические результаты о свойствах оценок параметров ковариационной функции получены лишь в очень специальных случаях [6, 8, 10, 12]. В частности, рассматривается асимптотическая постановка (размер выборки стремится к бесконечности) задачи с точно специфицированной моделью. Вместе с тем необходим теоретический анализ свойств метода регрессии на основе гауссовских процессов для данных большой размерности при ограниченном объёме выборки, а также исследование поведения метода при неточной спецификации модели.

Теоретический анализ метода регрессии на основе гауссовских процессов естественно проводить с помощью байесовского подхода, в котором центральным результатом является теорема Бернштейна—фон Мизеса о близости апостериорного распределения неизвестного вектора параметров регрессии к соответствующему нормальному распределению. Эта теорема предоставляет теоретическое основание для различных байесовских процедур, например для байесовских методов оценивания параметров моделей, построения эллиптических доверительных множеств на основе первых двух моментов апостериорного распределения и т. п.

Классические асимптотические методы статистики [3] не позволяют исследовать свойства апостериорного распределения в случае растущей размерности и ограниченного объёма данных, поэтому требуются новые статистические подходы, основанные на современной теории эмпирических процессов (см. [13]).

В данной работе приведён ряд авторских результатов [1, 2], из которых следует, что для регрессии на основе гауссовских процессов в достаточно общих предположениях выполнена неасимптотическая версия теоремы Бернштейна—фон Мизеса для случая, когда исходное параметрическое предположение о ковариационной функции гауссовского процесса может не выполняться. Полученные результаты позволяют установить достаточное условие на соотношение объёма выборки и размерности пространства параметров, которое гарантирует выполнение теоремы. Проведённое масштабное статистическое моделирование, являющееся основной целью данной работы, демонстрирует, что все утверждения доказанной авторами теоремы Бернштейна—фон Мизеса имеют место на практике.

Работа построена следующим образом. В разделе 2 описана процедура восстановления регрессии на основе гауссовских процессов. В разделе 3 приведена полученная теорема Бернштейна—фон Мизеса. В разделе 4 описаны результаты проведённых вычислительных экспериментов.

2. Регрессия на основе гауссовских процессов

Построение регрессии на основе гауссовских процессов проводится следующим образом. Предполагается, что задана выборка значений неизвестной функции $\mathbf{D} = (X, \mathbf{y}) = \{\mathbf{x}_i, y(\mathbf{x}_i) = y_i\}_{i=1}^n$, $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^d$. Необходимо построить по выборке $\mathbf{D}$ размера n аппроксимацию $\hat{y}(\mathbf{x})$ функции $y(\mathbf{x})$.

Будем считать, что функция $y(\mathbf{x})$ является реализацией гауссовского процесса. Без ограничения общности положим среднее значение этого гауссовского процесса равным нулю. В таком случае совместное распределение вектора значений $\mathbf{y}$ имеет вид $\mathbf{y} \propto \mathcal{N}(\mathbf{0}, K)$, где K — некоторая положительно определённая ковариационная матрица, вообще говоря зависящая от выборки $\mathbf{D}$.

Предположим, что ковариация между значениями гауссовского процесса в точках $\mathbf{x}$ и $\mathbf{x}'$ задаётся некоторой ковариационной функцией $\text{cov}(y(\mathbf{x}), y(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}')$. Тогда ковариационная матрица значений выборки $\mathbf{D}$ имеет вид $K = \{k(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$.

Для гауссовского случайного процесса апостериорное распределение значения реализации $y(\mathbf{x})$ в новой точке $\mathbf{x} \in \mathbb{R}^d$ будет нормальным при фиксированной ковариационной функции

$$\text{Law}(y(\mathbf{x}) \mid \mathbf{D}) = \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x})).$$

Выражения для математического ожидания $\mu(\mathbf{x})$ и дисперсии $\sigma^2(\mathbf{x})$ апостериорного распределения $\text{Law}(y(\mathbf{x}) \mid \mathbf{D})$ выписываются в явном виде:

$$\mu(\mathbf{x}) = \mathbf{k}^\top(\mathbf{x})K^{-1}\mathbf{y}, \quad \sigma^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})K^{-1}\mathbf{k}(\mathbf{x}).$$

Здесь $\mathbf{k}(\mathbf{x}) = (k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_n))^T$ — вектор-столбец из ковариаций между значением $y(\mathbf{x})$ случайного процесса в точке $\mathbf{x}$ и значениями $(y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))$ случайного процесса в точках выборки $(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Апостериорное математическое ожидание $\mu(\mathbf{x})$ используется в качестве прогноза $\hat{y}(\mathbf{x})$ значения процесса $y(\mathbf{x})$, а апостериорная дисперсия $\sigma^2(\mathbf{x})$ служит оценкой неопределённости прогноза.

На практике для моделирования ковариационной функции обычно используют параметрическое семейство ковариационных функций $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$, где Θ — некоторое компактное множество. В таком случае для построения регрессии на основе гауссовских процессов необходимо оценить вектор параметров $\boldsymbol{\theta}$ ковариационной функции $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$. Естественно, нет никаких оснований полагать, что параметрическое предположение о ковариационной функции гауссовского процесса выполняется, т. е., вообще говоря, $k(\mathbf{x}, \mathbf{x}') \notin \{k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}'), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$.

Совместное распределение вектора известных значений $\mathbf{y}$ будет нормальным. Тогда логарифм (квази)правдоподобия данных имеет вид

$$L(\boldsymbol{\theta}) = -\frac{1}{2}[n \log 2\pi + \ln |K_{\boldsymbol{\theta}}| + \mathbf{y}^T K_{\boldsymbol{\theta}}^{-1} \mathbf{y}], \quad (1)$$

где $K_{\boldsymbol{\theta}} = \{k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$.

В качестве оценки вектора параметров $\boldsymbol{\theta}$ часто используют оценку максимального (квази)правдоподобия

$$\tilde{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}).$$

Пусть задано некоторое априорное распределение $\Pi(d\boldsymbol{\theta})$ для вектора параметров $\boldsymbol{\theta}$. Апостериорное распределение при заданной выборке $\mathbf{D}$ будет описывать условное распределение случайного вектора $\boldsymbol{\vartheta}$. Обычно это соотношение записывают в виде

$$\operatorname{Law}(\boldsymbol{\vartheta} | \mathbf{D}) \propto \exp\{L(\boldsymbol{\theta})\} \Pi(d\boldsymbol{\theta}). \quad (2)$$

Отметим, что максимум апостериорного распределения может использоваться в качестве характерного значения (оценки) вектора параметров $\boldsymbol{\theta}$ и в случае неинформативного априорного распределения $\Pi(d\boldsymbol{\theta})$ будет совпадать с оценкой максимального (квази)правдоподобия $\tilde{\boldsymbol{\theta}}$.

Цель данной работы — с помощью методов статистического моделирования проверить теоретические утверждения о свойствах апостериорного распределения $\operatorname{Law}(\boldsymbol{\vartheta} | \mathbf{D})$, полученные авторами в [1, 2].

3. Свойства апостериорного распределения

Обозначим через $\|\cdot\|_2$ спектральную норму матрицы, через $I_n \in \mathbb{R}^{n \times n}$ — единичную матрицу, через $\mathbb{E}\{\cdot\}$ — математическое ожидание и через $\operatorname{Var}\{\cdot\}$ — ковариацию по распределению $\mathcal{N}(\mathbf{0}, K)$.

Будем рассматривать случай неинформативного априорного распределения $\Pi(d\boldsymbol{\theta})$ и возможной неправильной спецификации модели, т. е., вообще говоря, $k(\mathbf{x}, \mathbf{x}') \notin \{k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}'), \boldsymbol{\theta} \in \Theta\}$.

Пусть выполнены следующие предположения.

- (A1) $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ трижды непрерывно дифференцируема по $\boldsymbol{\theta} \in \Theta$ при $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$.
 (A2) $\max\{\|K\|_2, \|K_{\boldsymbol{\theta}}\|_2\} \leq \bar{\lambda} < \infty$, $\max\{\|K^{-1}\|_2, \|K_{\boldsymbol{\theta}}^{-1}\|_2\} \leq \lambda_0 < \infty$ при $\boldsymbol{\theta} \in \Theta$.
 (A3) при $\boldsymbol{\theta} \in \Theta$, $i, j, k = \overline{1, p}$

$$\left\| \frac{\partial K_{\boldsymbol{\theta}}}{\partial \theta_i} \right\|_2 \leq \lambda_1 < \infty, \quad \left\| \frac{\partial^2 K_{\boldsymbol{\theta}}}{\partial \theta_i \partial \theta_j} \right\|_2 \leq \lambda_2 < \infty, \quad \left\| \frac{\partial^3 K_{\boldsymbol{\theta}}}{\partial \theta_i \partial \theta_j \partial \theta_k} \right\|_2 \leq \lambda_3 < \infty.$$

- (A4) Центральная точка

$$\boldsymbol{\theta}^* = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \mathbb{E}L(\boldsymbol{\theta})$$

существует.

- (A5) Минимальные собственные числа матриц $(1/n)D_0^2$ и $(1/n)V_0^2$ больше $d_0 > 0$ и $v_0 > 0$ соответственно, где $D_0^2 = -\nabla^2 \mathbb{E}L(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ и $V_0^2 = \operatorname{Var}\{\nabla L(\boldsymbol{\theta})\}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$.

Эти предположения аналогичны предположениям из [12].

3.1. Квадратичная экспоненциальная ковариационная функция

Рассмотрим пример параметрического класса ковариационных функций, а именно квадратичную экспоненциальную ковариационную функцию [11], которая широко используется на практике [4, 7]:

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \theta_1^2 \left(\exp \left(-\frac{1}{2} \theta_2^2 \sum_{i=1}^d (x_i - x'_i)^2 \right) + \sigma^2 \delta(\mathbf{x} - \mathbf{x}') \right), \quad (3)$$

где $\delta(\cdot)$ — функция Кронекера. Первое слагаемое в (3) задаёт ковариацию между значениями реализации гауссовского процесса в точках пространства, а второе слагаемое определяет уровень дисперсии нормально распределённого белого шума в данных.

Для такой ковариационной функции выполнение условий, перечисленных выше, обеспечивается выбором достаточно регулярного дизайна X и фиксированной величины уровня шума $\sigma^2 \geq \sigma_0^2 > 0$, играющего роль параметра регуляризации для ковариационной матрицы $K_{\boldsymbol{\theta}}$.

3.2. Теорема Бернштейна—фон Мизеса для регрессии на основе гауссовских процессов

Обозначим через C универсальную абсолютную константу, через $E\{\cdot\}$ — математическое ожидание по распределению $\operatorname{Law}(\boldsymbol{\vartheta} \mid \mathbf{D})$. Определим величины

$$\bar{\boldsymbol{\vartheta}} \stackrel{\text{def}}{=} E\{\boldsymbol{\vartheta} \mid \mathbf{D}\}, \quad \mathbb{S}^2 \stackrel{\text{def}}{=} E\{(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})^\top \mid \mathbf{D}\},$$

играющие роль апостериорного среднего значения и апостериорной ковариационной матрицы случайного вектора $\boldsymbol{\vartheta}$ соответственно.

Положим

$$C_1 = \frac{4}{v_0^2} r_0^2 \bar{\lambda}^2 \lambda_0^4 (2\lambda_1^2 \lambda_0 + \lambda_2)^2, \quad C_4 = \frac{4r_0^2 \Lambda^2}{9d_0^2 v_0}, \quad \Lambda = 4\lambda_0^6 \lambda_1^3 \bar{\lambda}^3 + 5,5\lambda_0^4 \bar{\lambda}^2 \lambda_1 \lambda_2 + \bar{\lambda} \lambda_3,$$

где $r_0 > 0$ — константа, определённая в оценке сверху для экспоненциального момента производной логарифма правдоподобия $L(\boldsymbol{\theta})$ (см. [1]). Справедлива следующая теорема [1, 2].

Теорема 1 (теорема Бернштейна—фон Мизеса). Пусть выполнены предположения (A1)—(A5) и

$$n \geq \max(C_1 p, C_4 p^4). \quad (4)$$

Тогда найдутся β , явно определяемое константами из (A1)—(A5), и случайное событие Ω_n с $\mathbb{P}(\Omega_n) \geq 1 - Ce^{-x_n}$, $x_n = \log n$, такие что для $\tau_n = \beta/\sqrt{n}$ на Ω_n

$$\|D_0(\bar{\boldsymbol{\vartheta}} - \tilde{\boldsymbol{\theta}})\|_2^2 \leq C\tau_n(p + x_n), \quad (5)$$

$$\|\mathbb{I}_p - D_0 \mathfrak{S}^2 D_0\|_\infty \leq C\tau_n(p + x_n). \quad (6)$$

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

$$\mathbb{P}(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}}) \in A) \geq e^{-C\tau_n(p+x_n)} \{\mathbb{P}(\gamma \in A) - C\tau_n(p + x_n)^{1/2}\} - Ce^{-x_n}, \quad (7)$$

$$\mathbb{P}(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}}) \in A) \leq e^{C\tau_n(p+x_n)} \{\mathbb{P}(\gamma \in A) + C\tau_n(p + x_n)^{1/2}\} + Ce^{-x_n}. \quad (8)$$

Идея доказательства теоремы состоит в получении ряда оценок для экспоненциальных моментов логарифма правдоподобия $L(\boldsymbol{\theta})$ и его производных и использовании результатов работ [13, 14].

Неравенства (5) и (6) 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 $\tilde{\boldsymbol{\theta}}$ и матрице D_0^{-2} соответственно, а неравенства (7) и (8) описывают близость апостериорного распределения $\text{Law}(\boldsymbol{\vartheta} \mid \mathbf{D})$ к соответствующему нормальному распределению.

4. Вычислительный эксперимент

4.1. Генерация данных

В экспериментах мы будем использовать следующие ковариационные функции:

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \theta_1^2 \left(\exp \left(- \sum_{i=1}^d \theta_{i+1}^2 (x_i - x'_i)^2 \right) + \sigma^2 \delta(\mathbf{x} - \mathbf{x}') \right), \quad (9)$$

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \theta_1^2 \left(\exp \left(- \theta_2^2 \sum_{i=1}^d (x_i - x'_i)^2 \right) + \sigma^2 \delta(\mathbf{x} - \mathbf{x}') \right), \quad (10)$$

$$k_{\theta}(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^d \theta_i^2 (x_i - x'_i)^2\right) + \sigma^2 \delta(\mathbf{x} - \mathbf{x}'). \quad (11)$$

Дисперсию шума будем считать известной и равной $\sigma^2 = 0,001$. В данной работе предполагается, что априорное распределение вектора параметров равномерное на заданном гиперкубе $\Theta = (0, \theta_1^{\max}) \times \dots \times (0, \theta_p^{\max})$. Такое неинформативное априорное распределение не искажает форму исходного правдоподобия в окрестности точки θ^* .

Пусть выбрано истинное значение вектора параметров θ^* и сгенерирована выборка точек $X = \{\mathbf{x}_i\}_{i=1}^n$, $\mathbf{x} \in \mathbb{X} = [0, 1]^d$. Тогда совместное распределение вектора значений $\mathbf{y}$ будет многомерным нормальным с нулевым математическим ожиданием и ковариационной матрицей $K_{\theta^*} = \{k_{\theta^*}(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$.

Генерация отдельной выборки $\mathbf{D}$ для произвольного $\theta \in \Theta$ проводится следующим образом.

- Пусть зафиксирована ковариационная функция $k_{\theta}(\mathbf{x}, \mathbf{x}')$ и её параметры θ .
- Сгенерируем набор точек $X = \{\mathbf{x}_i\}_{i=1}^n$ фиксированного размера n , например используя равномерное распределение на гиперкубе $\mathbb{X} = [0, 1]^d$.
- Сгенерируем нормально распределённый вектор $\mathbf{y}$ с нулевым математическим ожиданием и ковариационной матрицей $K_{\theta} = \{k_{\theta}(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$ в точках X .
- Вектор $\mathbf{y}$ будет реализацией гауссовского процесса с фиксированной ковариационной функцией $k_{\theta}(\mathbf{x}, \mathbf{x}')$.

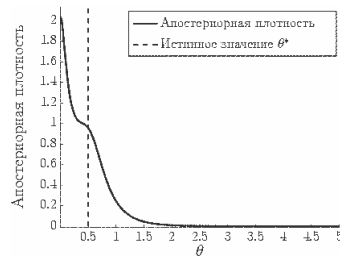
4.2. Вид апостериорного распределения данных

Для широко используемых на практике ковариационных функций характерно достаточно регулярное поведение апостериорной плотности и логарифма правдоподобия, но в некоторых случаях правдоподобие имеет максимум в нуле или несколько локальных экстремумов [9] (например, в случае, если ковариационная матрица K_{θ} «почти» вырожденная). На рис. 1 показаны соответствующие примеры.

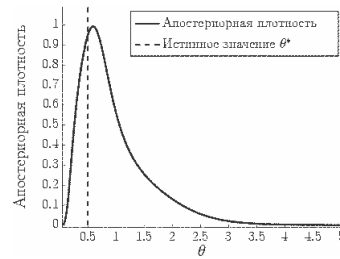
4.3. Распределение оценок параметров

Для выборки $\mathbf{D}$ можно получить оценку максимума правдоподобия $\tilde{\theta}$ и значение апостериорного среднего $\bar{\theta}$. Исследуем свойства распределений значений $\tilde{\theta}$ и $\bar{\theta}$ в зависимости от размера выборки.

Для оценки плотности распределения использовалась ядерная оценка плотности с гауссовским ядром [15]. Ширина ядра подбиралась на основе скользящего контроля. Кроме ядерных оценок плотности, будем подсчитывать и показывать на рисунках 95-процентные доверительные интервалы для полученных оценок.



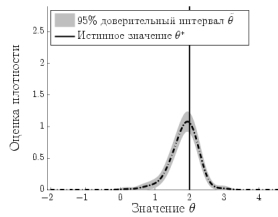
Обычная форма апостериорной плотности. Размер обучающей выборки равен $n = 50$



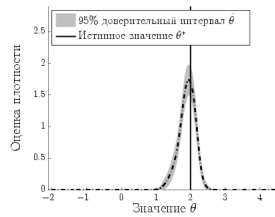
Глобальный максимум апостериорной плотности находится в нуле. Размер обучающей выборки равен $n = 50$

Рис. 1. Возможные виды апостериорной плотности распределения вектора параметров θ в одномерном случае

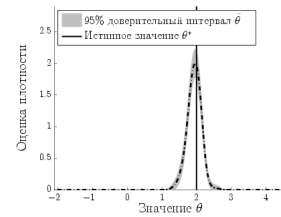
Рассматривался одномерный случай ($p = d = 1$) и ковариационная функция типа (11). Для θ полученные результаты приведены на рис. 2. Для ϑ полученные результаты приведены на рис. 3. Видно, что с увеличением размера выборки плотности распределений величин θ и ϑ концентрируются вокруг истинного значения.



Размер выборки n равен 10

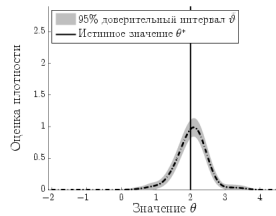


Размер выборки n равен 200

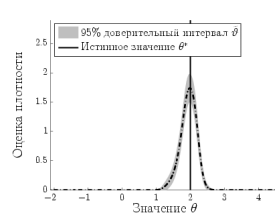


Размер выборки n равен 1000

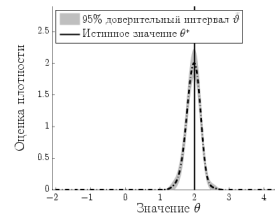
Рис. 2. Ядерная оценка плотности для $\tilde{\theta}$, размерность пространства параметров $p = 1$



Размер выборки n равен 10

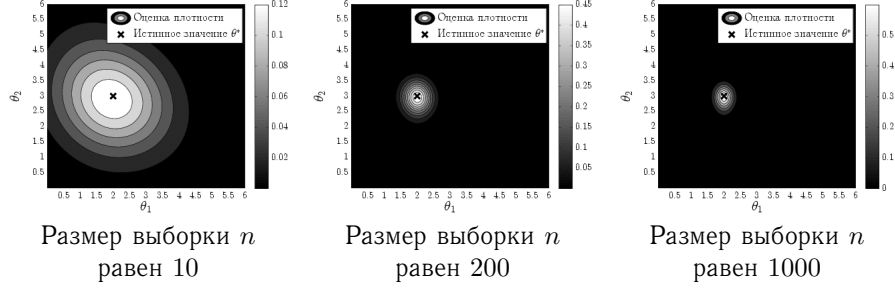
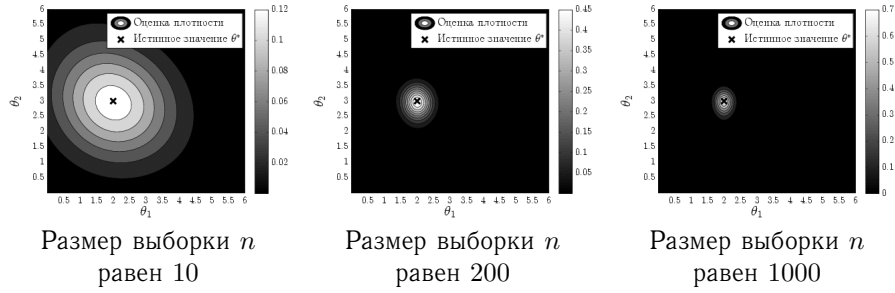


Размер выборки n равен 200



Размер выборки n равен 1000

Рис. 3. Ядерная оценка плотности для $\tilde{\vartheta}$, размерность пространства параметров $p = 1$

Рис. 4. Ядерная оценка плотности для $\bar{\theta}$, размерность пространства параметров $p = 2$ Рис. 5. Ядерная оценка плотности для $\bar{\vartheta}$, размерность пространства параметров $p = 2$

Рассматривался и двумерный случай ($p = 2$), т. е. ковариационная функция имела вид (10). Для $\bar{\theta}$ полученные результаты приведены на рис. 4. Для $\bar{\vartheta}$ полученные результаты приведены на рис. 5. В этом случае с увеличением размера выборки плотности распределений величин $\bar{\theta}$ и $\bar{\vartheta}$ также концентрируются вокруг истинных значений. Кроме того, на рисунках видно, что формы распределений для $\bar{\theta}$ и $\bar{\vartheta}$ похожи.

4.4. Оценки сверху для уклонений апостериорного распределения

Рассмотрим следующий эксперимент для одномерной ($p = 1$) и двумерной ($p = 2$) ковариационных функций. В теореме 1 показано, что нормы $\|D_0(\bar{\vartheta} - \bar{\theta})\|_2^2$ и $\|\mathbb{I}_p - D_0 \mathfrak{S}^2 D_0\|_\infty$ ограничены сверху величинами, которые убывают как $1/\sqrt{n}$ при увеличении размера выборки n . На рис. 6 и 7 показано поведение этих норм для одномерной ковариационной функции (11) и для двумерной ковариационной функции (10) соответственно. Видно, что обе нормы как в одномерном, так и в двумерном случае убывают с увеличением размера выборки n .

Из рис. 8 следует, что величина $\|D_0(\bar{\vartheta} - \bar{\theta})\|_2^2$ также убывает с увеличением размера выборки n , например для размерности пространства параметров $p = 6$.

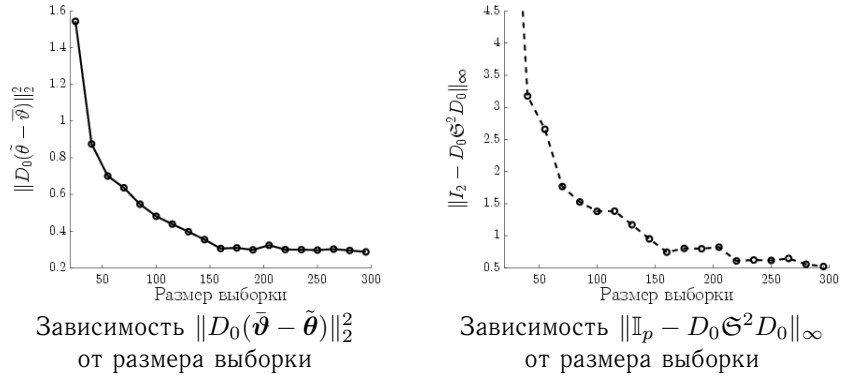


Рис. 6. Зависимость величин (5) и (6) от размера выборки n , размерность пространства параметров $p = 1$

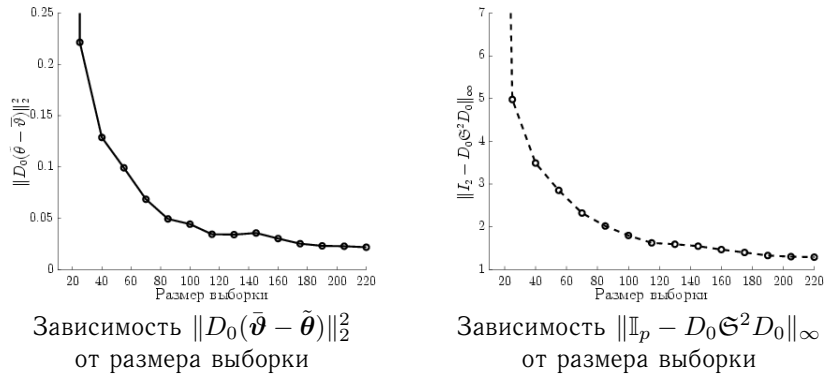


Рис. 7. Зависимость величин (5) и (6) от размера выборки n , размерность пространства параметров $p = 2$

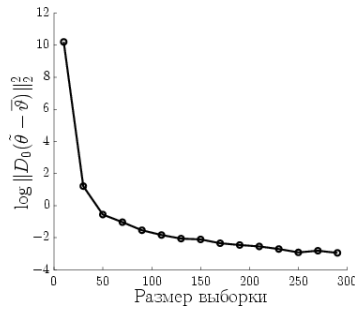


Рис. 8. Зависимость $\|D_0(\hat{\vartheta} - \tilde{\vartheta})\|_2^2$ от размера выборки, размерность пространства параметров $p = 6$

4.5. Близость апостериорного распределения к нормальному распределению

В рамках теоремы Бернштейна—фон Мизеса получено утверждение о том, что апостериорное распределение вектора параметров ϑ близко к нормальному распределению: расстояние по вариации [5] между апостериорным распределением и соответствующим нормальным распределением уменьшается с увеличением объёма выборки n . Математическое ожидание и ковариационная матрица этого нормального распределения берутся равными математическому ожиданию $\bar{\vartheta}$ и ковариационной матрице $\mathbb{S}^2$ апостериорного распределения параметров ϑ .

В данном разделе приведены результаты экспериментов для ковариационной функции (10), которые иллюстрируют данное утверждение теоремы.

На рис. 9, а показана зависимость от объёма выборки n расстояния по вариации между искомым апостериорным распределением и соответствующим нормальным распределением. На рис. 9, б показана зависимость от объёма выборки n расстояния Хеллингера [5] между искомым апостериорным распределением и соответствующим нормальным распределением. Видно, что оба расстояния уменьшаются с увеличением размера выборки, что подтверждает, что оценки (7) и (8) справедливы в рассматриваемом случае.

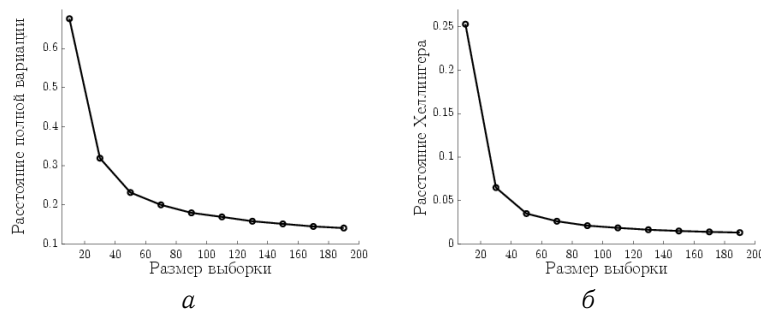


Рис. 9. Зависимость расстояния между апостериорным распределением параметров и нормальным распределением, размерность пространства параметров $p = 2$:
а — расстояние по вариации, б — расстояние Хеллингера

4.6. Соотношение между размером выборки и размерностью

Результат, приведённый в разделе 1 (см. формулу (4)), задаёт гарантированный размер выборки n , начиная с которого выполняется теорема Бернштейна—фон Мизеса. На рис. 10 приведена зависимость риска оценки параметров ковариационной функции от размера выборки n и размерности параметрического пространства p . Видно, что если для каждой размерности пространства параметров p размер выборки n больше некоторой критической величины, то полученные оценки параметров оказываются достаточно точными.

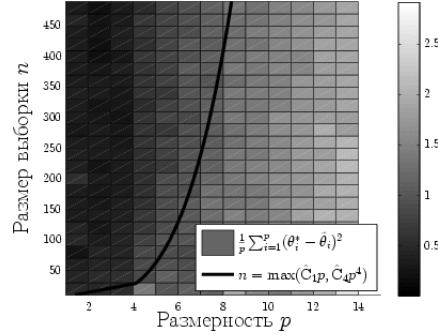


Рис. 10. Зависимость риска оценки параметров ковариационной функции от размера выборки n и размерности параметрического пространства p

5. Заключение

В работе приведены авторские результаты [1, 2], которые обосновывают использование байесовских методов оценивания параметров ковариационной функции при построении регрессии на основе гауссовских процессов:

- теорема Бернштейна—фон Мизеса для неасимптотического случая и возможной ошибочности параметрического предположения о типе ковариационной функции;
- достаточное условие на соотношение объёма выборки и размерности пространства параметров, которое гарантирует выполнение теоремы Бернштейна—фон Мизеса.

Проведено масштабное статистическое моделирование, результаты которого демонстрируют, что все утверждения доказанной авторами теоремы Бернштейна—фон Мизеса имеют место на практике. В частности, показано, что

- в случае почти вырожденной ковариационной матрицы K_ϑ апостериорная плотность распределения вектора параметров далека от соответствующей нормальной плотности распределения;
- плотности распределений оценки максимального (квази)правдоподобия $\tilde{\vartheta}$ и апостериорного среднего значения $\bar{\vartheta}$ концентрируются вокруг истинного значения параметров θ^* ;
- величины $\|D_0(\bar{\vartheta} - \tilde{\vartheta})\|_2^2$ и $\|\mathbb{I}_p - D_0 \mathfrak{S}^2 D_0\|_\infty$ уменьшаются при увеличении объёма выборки n ;
- апостериорное распределение вектора параметров ковариационной функции стремится (в смысле расстояния по вариации) к соответствующему нормальному распределению;
- сформулированное достаточное условие на соотношение объёма выборки n и размерности пространства параметров p действительно влияет на выполнимость теоремы Бернштейна—фон Мизеса.

Литература

- [1] Бурнаев Е. В., Зайцев А. А., Спокойный В. Г. Свойства апостериорного распределения модели зависимости на основе гауссовских случайных полей // Автоматика и телемеханика. — 2013. — Т. 10. — С. 55–67.
- [2] Бурнаев Е. В., Зайцев А. А., Спокойный В. Г. Теорема Бернштейна—фон Мизеса для регрессии на основе гауссовских процессов // Успехи мат. наук. — 2013. — Т. 68, № 5. — С. 269–270.
- [3] Ибрагимов И. А., Хасьяминский Р. З. Асимптотическая теория оценивания. — М.: Наука, 1979.
- [4] Червоненкис А. Я., Чернова С. С., Зыкова Т. В. Применение ядерной гребневой оценки к задаче расчёта аэродинамических характеристик пассажирского самолета (сравнение с результатами, полученными с использованием искусственных нейронных сетей) // Автоматика и телемеханика. — 2011. — Т. 5, № 5. — С. 175–182.
- [5] Ширяев А. Н. Вероятность-1, Вероятность-2. — М.: МЦНМО, 2011.
- [6] Eidsvik J., Finley A. O., Banerjee S., Rue H. Approximate Bayesian inference for large spatial datasets using predictive process models // Comput. Statist. Data Analysis. — 2011. — Vol. 56, no. 6. — P. 1362–1380.
- [7] Forrester A., Sobester A., Keane A. Engineering Design via Surrogate Modelling: A Practical Guide. — Wiley, 2008.
- [8] Kaufman C. G., Schervish M. J., Nychka D. W. Covariance tapering for likelihood-based estimation in large spatial data sets // J. Am. Statist. Assoc. — 2008. — Vol. 103 (484). — P. 1545–1555.
- [9] Kok S. The asymptotic behaviour of the maximum likelihood function of Kriging approximations using the Gaussian correlation function // EngOpt 2012 — Internat. Conf. on Engineering Optimization, Rio de Janeiro, Brazil, 1–5 July 2012.
- [10] Mardia K. V., Marshall R. J. Maximum likelihood estimation of models for residual covariance in spatial regression // Biometrika. — 1984. — Vol. 71, no. 1. — P. 135–146.
- [11] Rasmussen C. E., Williams C. K. I. Gaussian Processes for Machine Learning. Vol. 1. — Cambridge, MA: MIT Press, 2006.
- [12] Shaby B., Ruppert D. Tapered covariance: Bayesian estimation and asymptotics // Journal Comput. Graph. Statist. — 2012. — Vol. 21, no. 2. — P. 433–452.
- [13] Spokoiny V. Parametric estimation. Finite sample theory // Ann. Statist. — 2012. — Vol. 6. — P. 2877–2909.
- [14] Spokoiny V. Bernstein–von Mises theorem for growing parameter dimension. — 2013. — arXiv:1302.3430 [math.ST].
- [15] Wasserman L. All of Statistics. A Concise Course in Statistical Inference. — Berlin: Springer, 2004.

