



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Национальный исследовательский
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ВСЕ ГРАНИ МАТЕМАТИКИ И МЕХАНИКИ

**Сборник статей
Всероссийской молодежной
научной конференции**

Томск, 12–15 мая 2020 г.



ТОМСК
«Издательство НТЛ»
2020

О численном исследовании свойств последовательной оценки параметров авторегрессии

Е.Л. Хабарова, Т.В. Емельянова

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Рассматривается задача оценивания параметров авторегрессионной модели второго порядка с непрерывным временем. Предлагается последовательный план идентификации модели на основе оценок максимального правдоподобия. Для построения плана используется модифицированная выборочная информационная матрица Фишера. Доказывается, что полученные оценки имеют нормальное распределение. Изучены некоторые свойства процедуры оценивания. Численно проиллюстрирован вопрос о длительности процедуры.

Ключевые слова: *авторегрессия с непрерывным временем, метод максимального правдоподобия, нормальное распределение, идентификация параметров, последовательное оценивание, момент остановки, численное моделирование.*

Актуальность и история вопроса

В прикладных исследованиях процессы зачастую описываются стохастическими дифференциальными и стохастическими разностными уравнениями. Эти уравнения в большинстве случаев содержат неизвестные параметры, следовательно, перед их использованием для решения задач необходимо идентифицировать параметры непосредственным оцениванием. Таким образом, оценивание параметров стохастических дифференциальных и разностных уравнений является актуальной задачей современного анализа данных.

Для решения задач оценивания параметров существуют различные эффективные методы (метод максимального правдоподобия, метод моментов, метод наименьших квадратов и другие), однако, в случае зависимых наблюдений эти оценки являются нелинейными функциями, что осложняет процесс исследования их свойств.

В теории идентификации часто предполагают, что процесс имеет неограниченное количество наблюдений. Но в этом случае, для данных небольшого объема предполагается, что выборочные свойства оценок близки к асимптотическим. Такой подход к исследованию оценок для

малого количества данных не всегда является правильным и зачастую приводит к ошибочным выводам.

Одним из методов решения задачи оценивания параметров в неасимптотической постановке является метод последовательного оценивания Липцера и Ширяева [1] для оценки параметра θ процесса, описываемого стохастическим дифференциальным уравнением вида

$$dX_t = \theta f_t(X_t) dt + dW_t$$

по наблюдениям процессов X_t и f_t (здесь W_t – стандартное броуновское движение).

В своей работе они предложили в оценке максимального правдоподобия

$$\theta_T = \left(\int_0^T f_s^2(X) ds \right)^{-1} \int_0^T f_s(X) dX_s$$

заменить детерминированный промежуток наблюдений $[0, T]$ на промежуток случайной длины $[0, \tau(H)]$, где

$$\tau(H) = \inf \left\{ T > 0 : \int_0^T f_s^2 ds = H \right\},$$

и в качестве оценки неизвестного параметра θ использовать последовательную оценку максимального правдоподобия, которая определялась равенством

$$\delta_H(X) = \frac{1}{H} \int_0^{\tau(H)} f_t(X) dX_t.$$

Было доказано, что предложенная в [1] оценка обладает следующими свойствами: она является нормальной, несмещенной и ее среднеквадратическая погрешность не превышает $\frac{1}{H}$.

Однако в том случае, когда количество параметров превышает размерность процесса, возникают сложности с построением такого же красивого аналога. Решению этой проблемы посвящено немало работ. В качестве примера можно привести следующие результаты, сделанные в направлении данных исследований:

- Конев и Пергаменщиков смогли построить двухшаговую последовательную процедуру (без учета априорной информации об области параметров) для моделей авторегрессионного типа [2, 3];

- Конев и Емельянова смогли построить более простую процедуру (одноэтапную), но с учетом априорной информации [4].

В практических исследованиях часто возникает неасимптотическая проблема оценивания, когда необходимо определить длину реализации, при которой оценки достигают заданной точности. Для решения задач в неасимптотической постановке требуются методы, которые позволяют контролировать точность оценок при малых объемах данных.

Поскольку в практических задачах объем доступных данных всегда конечен и стоит задача определения качества оценок, вычисленных по наблюдениям на ограниченном временном интервале, то в таких задачах успешно применяется последовательный анализ. Он характеризуется тем, что длительность наблюдений заранее не фиксируется, а определяется по специальным правилам.

Целью нашей работы является построение последовательной оценки максимального правдоподобия с неасимптотическим с нормальным распределением для любых возможных значений параметров двухпараметрической авторегрессионной модели с непрерывным временем и численное исследование некоторых свойств полученной оценки. Провести имитационное моделирование вопроса о длительности процедуры.

Постановка задачи.

Построение последовательной оценки

Определим задачу оценивания параметров авторегрессионной модели второго порядка по наблюдениям процесса X_t :

$$d\dot{X}_t = (\theta_1 \dot{X}_t + \theta_2 X_t) dt + dW_t \quad (1)$$

или в векторном виде:

$$d\dot{X}_t = (\theta_1 \quad \theta_2) \begin{pmatrix} \dot{X}_t \\ X_t \end{pmatrix} dt + dW_t,$$

где $\theta = (\theta_1, \theta_2)'$ – вектор-столбец неизвестных параметров, W_t – винеровский процесс (здесь штрих обозначает транспонирование).

На промежутке $[0, T]$ оценка максимального правдоподобия вектора θ по наблюдениям процесса (1) представлена в следующем виде:

$$\theta_T^* = M_T^+ \int_0^T (\dot{X}_t, X_t)' dX_t, \quad (2)$$

где

$$M_T = \begin{pmatrix} \int_0^T \dot{X}_s^2 ds & \int_0^T \dot{X}_s X_s ds \\ \int_0^T \dot{X}_s X_s ds & \int_0^T X_s^2 ds \end{pmatrix}$$

– выборочная информационная матрица Фишера, M_T^+ – обратная к M_T матрица, если она необособленная, и $M_T^+ = 0$ – иначе.

Построим последовательную оценку максимального правдоподобия с неасимптотическим нормальным распределением для любых возможных параметров.

Для этого введем систему моментов остановки и произведем определенную модификацию выборочной информационной матрицы Фишера.

Для произвольного порога $h > 0$ введем моменты остановки $\tau_i(h)$, $i = 0, 1, 2$ по следующим формулам:

$$\begin{aligned} \tau_0(h) &= 0, \\ \tau_1(h) &= \inf \left\{ T > 0 : \int_0^T \dot{X}_s^2 ds = h \right\}, \\ \tau_2(h) &= \inf \left\{ T > \tau_1(h) : \int_{\tau_1(h)}^T X_s^2 ds = h \right\}. \end{aligned} \quad (3)$$

Пусть

$$m(h) = \begin{pmatrix} \int_0^{\tau_1(h)} \dot{X}_s^2 ds & \int_0^{\tau_1(h)} \dot{X}_s X_s ds \\ \int_{\tau_1(h)}^{\tau_2(h)} \dot{X}_s X_s ds & \int_{\tau_1(h)}^{\tau_2(h)} X_s^2 ds \end{pmatrix} \quad (4)$$

– модифицированная выборочная информационная матрица Фишера.

Определим вектор $V(h) = (V_1(h), V_2(h))'$ по формулам

$$V_1(h) = \int_0^{\tau_1(h)} \dot{X}_s dX_s, \quad V_2(h) = \int_{\tau_1(h)}^{\tau_2(h)} X_s dX_s, \quad (5)$$

а последовательную оценку вектора неизвестных параметров θ равенством

$$\theta^* = m^{-1}(h)V(h) \quad (6)$$

или подробнее

$$\begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix} = \begin{pmatrix} \int_0^{\tau_1(h)} \dot{X}_s^2 ds & \int_0^{\tau_1(h)} \dot{X}_s X_s ds \\ \int_{\tau_1(h)}^{\tau_2(h)} \dot{X}_s X_s ds & \int_{\tau_1(h)}^{\tau_2(h)} X_s^2 ds \end{pmatrix}^{-1} \begin{pmatrix} \int_0^{\tau_1(h)} \dot{X}_s dX_s \\ \int_{\tau_1(h)}^{\tau_2(h)} X_s dX_s \end{pmatrix}.$$

Основной результат

Теорема 1. [5] Пусть задан авторегрессионный процесс второго порядка

$$d\dot{X}_t = (\theta_1 \dot{X}_t + \theta_2 X_t) dt + dW_t.$$

Пусть оценки вектора неизвестных параметров $\theta = (\theta_1, \theta_2)'$ определяются формулами (3) – (6). Тогда для любого параметра $\theta \in \mathbb{R}^2$ и порога $h > 0$

$$\frac{m(h)}{\sqrt{h}} (\theta^*(h) - \theta) \in N(0, I_2),$$

где I_2 – единичная матрица второго порядка.

Доказательство. Для доказательства нормальности вектора θ^* достаточно показать нормальность любой линейной комбинации его компонент. [5]

Запишем вектор V поэлементно в следующем виде:

$$V_1(h) = \int_0^{\tau_1(h)} X_s dX_s = \int_0^{\tau_1(h)} X_s \left(\left(\theta_1 X_s + \theta_2 X_s \right) ds + dW_s \right) =$$

$$= \int_0^{\tau_1(h)} \left(\theta_1 X_s^2 + \theta_1 X_s X_s \right) ds + \int_0^{\tau_1(h)} X_s dW_s = m(h)\theta_1 + \eta_1(h),$$

$$V_2(h) = m(h)\theta_2 + \eta_2(h),$$

где $m(h)\theta_i$ — i -я строка матрицы $m(h)$, умноженная на вектор θ ,

$$\eta_1(h) = \int_0^{\tau_1(h)} X_s dW_s, \quad \eta_2(h) = \int_{\tau_1(h)}^{\tau_2(h)} X_s dW_s.$$

Пусть теперь

$$\eta(h) = (\eta_1(h), \eta_2(h))',$$

тогда

$$V(h) = m(h)\theta + \eta(h)$$

и $\eta(h)$ можно представить в следующем виде:

$$\eta(h) = m(h)(\theta^*(h) - \theta).$$

Для доказательства того, что вектор $\frac{\eta(h)}{\sqrt{h}}$ имеет двумерное нормальное распределение с нулевым средним и единичной ковариационной матрицей, достаточно показать, что для любого $\lambda = (\lambda_1 \quad \lambda_2) \in \mathbb{R}^2$ характеристическая функция случайной величины

$$\zeta(h) = \frac{\lambda' \eta(h)}{\sqrt{h}} = \frac{\lambda_1 \eta_1(h)}{\sqrt{h}} + \frac{\lambda_2 \eta_2(h)}{\sqrt{h}}$$

будет иметь вид

$$\Phi_{\zeta(h)}(u) = E e^{iu\zeta(h)} = e^{-\frac{\lambda' \lambda u^2}{2}}.$$

Вычисляя повторные условные математические ожидания и учитывая, что [1]

$$E \left(e^{iu \frac{\lambda_j \eta_j(h)}{\sqrt{h}}} / \mathcal{F}_{\tau_{j-1}} \right) = e^{-\frac{\lambda_j^2 u^2}{2}},$$

будем иметь

$$\begin{aligned}
Ee^{iu\zeta(h)} &= Ee^{iu\left(\frac{\lambda_1\eta_1(h)}{\sqrt{h}} + \frac{\lambda_2\eta_2(h)}{\sqrt{h}}\right)} = E\left(e^{iu\frac{\lambda_1\eta_1(h)}{\sqrt{h}}} \cdot e^{iu\frac{\lambda_2\eta_2(h)}{\sqrt{h}}}\right) = \\
&= E\left(E\left(e^{iu\frac{\lambda_1\eta_1(h)}{\sqrt{h}}}/F_0\right)\right) \cdot E\left(E\left(e^{iu\frac{\lambda_2\eta_2(h)}{\sqrt{h}}}/F_{\tau_1}\right)\right) = \\
&= Ee^{-\frac{\lambda_1^2 u^2}{2}} \cdot Ee^{-\frac{\lambda_2^2 u^2}{2}} = e^{-\left(\frac{\lambda_1^2 u^2}{2} + \frac{\lambda_2^2 u^2}{2}\right)} = e^{-(\lambda_1 \quad \lambda_2)\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \frac{u^2}{2}} = e^{-\frac{\lambda' \lambda u^2}{2}}.
\end{aligned}$$

Таким образом, мы показали, что вектор

$$\frac{m(h)}{\sqrt{h}}(\theta^*(h) - \theta)$$

имеет двумерное нормальное распределение с нулевым средним и единичной ковариационной матрицей. ■

Итак, была построена последовательная оценка максимального правдоподобия с нормальным распределением для любых возможных параметров двухпараметрической авторегрессионной модели с непрерывным временем.

Известное распределение полученной оценки позволяет контролировать ее среднеквадратическую точность за счет выбора порога процедуры, а также строить доверительные области для оцениваемых параметров. Полученный результат, видимо, допускает обобщение на модели регрессионного типа с произвольным количеством параметров.

Численное исследование свойств процедуры

Далее изучим вопрос о длительности этой процедуры оценивания. Теоретических подходов пока не возникло – это одно из направлений дальнейшей работы.

Проиллюстрируем результаты исследований с помощью численного моделирования [6].

Будем моделировать процесс, заданный уравнением

$$d\dot{X}_t = (\theta_1 \dot{X}_t + \theta_2 X_t)dt + dW_t.$$

Для этого перейдем от дифференциального уравнения к разностному.

$$\begin{aligned}
d\dot{X}_t &= (\theta_1 \dot{X}_t + \theta_2 X_t)dt + dW_t = \\
&= \theta_1 \frac{dX_t}{dt}dt + \theta_2 X_t dt + dW_t = \theta_1 dX_t + \theta_2 X_t dt + dW_t.
\end{aligned}$$

Тогда

$$\dot{X}_{t_k} - \dot{X}_{t_{k-1}} = \theta_1 (X_{t_k} - X_{t_{k-1}}) + \theta_2 X_{t_k} \Delta t + W_{t_k} - W_{t_{k-1}},$$

где

$$W_{t_k} - W_{t_{k-1}} \in N(0, \Delta t), \quad \Delta t = t_k - t_{k-1};$$

$$\dot{X}_{t_k} - \dot{X}_{t_{k-1}} = \theta_1 (X_{t_k} - X_{t_{k-1}}) + \theta_2 X_{t_k} \Delta t + \sqrt{\Delta t} \xi_k, \quad \text{где } \xi_k \in N(0, 1).$$

$$\dot{X}_t = \frac{dX_t}{dt}, \quad \dot{X}_{t_k} = \frac{X_{t_k} - X_{t_{k-1}}}{\Delta t}, \quad \dot{X}_{t_{k-1}} = \frac{X_{t_{k-1}} - X_{t_{k-2}}}{\Delta t}.$$

Отсюда получаем

$$\begin{aligned} & X_{t_k} - X_{t_{k-1}} - X_{t_{k-1}} + X_{t_{k-2}} = \\ & = \theta_1 X_{t_k} \Delta t - \theta_1 X_{t_{k-1}} \Delta t + \theta_2 X_{t_k} (\Delta t)^2 + (\Delta t)^{\frac{3}{2}} \xi_k; \\ & X_{t_k} - \theta_1 X_{t_k} \Delta t - \theta_2 X_{t_k} (\Delta t)^2 = 2X_{t_{k-1}} - X_{t_{k-2}} - \theta_1 X_{t_{k-1}} \Delta t + (\Delta t)^{\frac{3}{2}} \xi_k; \\ & X_{t_k} = \frac{1}{1 - \theta_1 \Delta t - \theta_2 (\Delta t)^2} \left[X_{t_{k-1}} (2 - \theta_1 \Delta t) - X_{t_{k-2}} + (\Delta t)^{\frac{3}{2}} \xi_k \right]. \end{aligned}$$

Для моделирования моментов остановки заменим интеграл квадратурной суммой [6].

Тогда для

$$\begin{aligned} \tau_1(h) &= \inf \left\{ T > 0 : \int_0^T \dot{X}_s^2 ds = h \right\}, \\ \tau_2(h) &= \inf \left\{ T > \tau_1(h) : \int_{\tau_1(h)}^T X_s^2 ds = h \right\} \end{aligned}$$

имеем

$$\begin{aligned} \tau_1(h) &= \min \left\{ i > 0 : \Delta t \sum_{k=1}^i \dot{X}_{t_k}^2 \geq h \right\}, \quad \Delta t = \frac{T}{i}; \\ \tau_2(h) &= \min \left\{ j > \tau_1(h) : \Delta t \sum_{k=\tau_1(h)+1}^j X_{t_k}^2 \geq h \right\}, \quad \Delta t = \frac{T - \tau_1(h)}{j - \tau_1(h)}. \end{aligned}$$

Моделирование реализовывалось при следующих значениях: шаг дискретизации $\Delta t = 0.01$, $X_0 = 0$, объем выборки $N = 10000000$, истинные значения оцениваемых параметров $\theta_1 = -0.3$, $\theta_2 = -0.5$. Оценки вычисляются при каждом h из $N_h = 5000$ значений с начальным значением $h_0 = 0.1$ и шагом $\Delta h = 0.1$.

Ниже приведены графики зависимости значений моментов остановки $\tau_1(h)$ и $\tau_2(h)$ от значений порога h , а также графики $\tau_1(h)/h$ и $\tau_2(h)/h$, показывающие стабилизацию длительности процедуры с увеличением порога h .

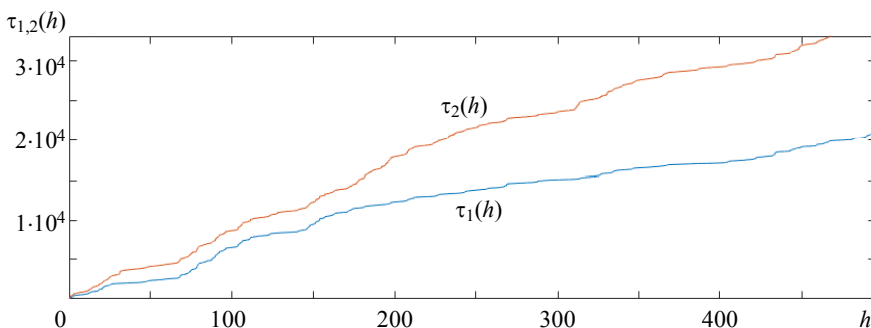


Рис. 1. Моменты остановки $\tau_1(h)$ и $\tau_2(h)$ для каждого значения порога h

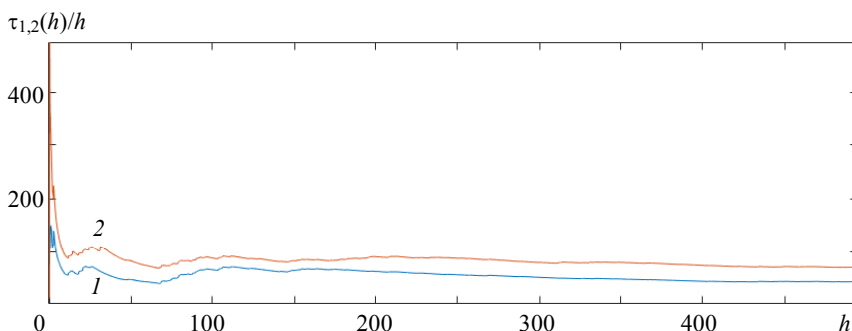


Рис. 2. Стабилизация значений $\frac{\tau_1(h)}{h}$ (кр. 1) и $\frac{\tau_2(h)}{h}$ (кр. 2)

Видно, что с увеличением порога h значения $\tau_i(h)/h$, $i=1,2$, стабилизируются. Численное исследование показывает, что зависимость длительности процедуры от порога h , скорее всего, линейна. Теоретические исследования этого вопроса продолжаются.

Заключение

Описана новая последовательная процедура, позволяющая получить оценки параметров авторегрессионной модели второго порядка, имеющие нормальное распределение. Это позволяет строить доверительные области с достаточно большой точностью. Процедура предполагает некоторую модификацию выборочной информационной матрицы Фишера, а также введение двух моментов остановки, построенных по этой матрице. Изучено одно из свойств оценок неизвестных параметров авторегрессионной модели второго порядка. Представлено численное подтверждение изученного свойства.

СПИСОК ЛИТЕРАТУРЫ

1. *Липцер Р.Ш., Ширяев А.Н.* Статистика случайных процессов. М.: Наука, 1974. 696 с.
2. *Конев В.В., Пергаменчиков С.М.* Гарантированное оценивание периодического сигнала на фоне авторегрессионных помех с неизвестными параметрами // Проблемы передачи информации. 1997. Т. 33. Вып. 4. С. 26–44.
3. *Конев В.В., Пергаменчиков С.М.* О последовательном оценивании параметров случайных процессов диффузионного типа // Проблемы передачи информации. 1985. Т. 21. Вып. 1. С. 48–61.
4. *Емельянова Т.В., Конев В.В.* О последовательном оценивании параметров непрерывной авторегрессии // Вестник Томского госуниверситета. Математика и механика. 2013. № 5(25). С. 12–25.
5. *Хабарова Е.Л., Емельянова Т.В.* О последовательном оценивании параметров авторегрессионной модели с непрерывным временем // Всероссийская молодежная научная конференция студентов, аспирантов и молодых ученых «Все грани математики и механики» (23–27 апреля 2019 г.): сб. статей. Томск, 2019. С. 189–197.
6. *Seydel R.* Tools for computational finance. Springer, 2017.
7. *Ширяев А.Н.* Вероятность. М.: Наука, 1980. 576 с.
8. *Булдинский А.В., Ширяев А.Н.* Теория случайных процессов. М.: Физмалит; Лаборатория Базовых Знаний, 2003. 400 с.

Хабарова Елена Леонидовна, студентка ММФ ТГУ; el_khabarova@mail.ru

Емельянова Татьяна Вениаминовна, к.ф.-м.н., доцент кафедры математического анализа и теории функций ММФ ТГУ