

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ МОДЕЛИ УСТОЙЧИВОЙ АВТОРЕГРЕССИИ

Иващенко А.О.

Научный руководитель: Емельянова Т. В.

Томский государственный университет, Механико-математический факультет
anuta_i@bk.ru

Введение

В задачах обработки временных рядов широко используются авторегрессионные модели, описывающие стационарные случайные процессы. Параметры таких моделей в большинстве случаев неизвестны, поэтому перед использованием модели требуется идентифицировать ее параметры непосредственным оцениванием. Проблема оценивания параметров является одной из основных проблем статистического анализа данных. В связи с этим представляет интерес рассмотреть задачу последовательного оценивания случайных процессов при функциях потерь квадратичной формы. Важным является также нахождение оптимальных правил остановки наблюдений в ситуациях, когда оптимальна последовательная процедура.

Таким образом, целью работы являлось проведение сравнения последовательного и байесовского подхода к оцениванию параметра модели устойчивой авторегрессии первого порядка (AR (1)).

Описание алгоритма

Рассматривается процесс устойчивой авторегрессии первого порядка с дискретным временем

$$X_i = \beta X_{i-1} + \varepsilon_i, i = 1, 2, \dots,$$

где $X_0 = 0$, ε_i – независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и конечной дисперсией. В качестве оценки параметра β предлагается построить оценку по методу наименьших квадратов, в которой число наблюдений не фиксируется заранее [1]

$$\hat{\beta}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}.$$

При использовании байесовского подхода к оцениванию параметров вводится некоторая функция потерь, которая характеризуется мерой расхождения между истинным значением параметра и его оценкой. В свою очередь риск – математическое ожидание функции потерь [2]. Рассмотрим функцию потерь вида

$$L_n(\hat{\beta}_n, \beta) = A * \frac{1}{n} \sum_{i=1}^n [\hat{X}_i - E(X_i | X_{i-1})]^2 + n = \frac{1}{n} \sum_{i=1}^n X_i^2 - A n^{-1} I_n (\hat{\beta}_n - \beta)^2 + n,$$

где $I_n = \sum_{i=1}^n X_{i-1}^2$, а $A = c^{-1}$ – обратная величина цены одного наблюдения [2]. Риск функции потерь $\bar{R}_n \approx 2n_0$ [3].

Однако бесполезно использовать n_0 в случае, когда неизвестны значения σ^2 и β . Для такой ситуации предлагается последовательный подход к оцениванию [4]. При этом пользуются правилом остановки t_A , построенным по наблюдаемому процессу

$$t_A = \inf \{n \geq m_A : n \geq A^{\frac{1}{2}} \hat{\sigma}_n\},$$

где $\hat{\sigma}_n = n^{-1} \sum_{i=1}^n (X_i - \hat{\beta}_n X_{i-1})^2$ и m_A – заранее заданный объем выборки.

Таким образом, риск, с учетом заданного правила остановки,

$$R_{t_A} = EL_{t_A}(\hat{\beta}_{t_A}, \beta).$$

Оценка $\bar{X}_{t_A}$ является асимптотически эффективной. Более того, правило остановки также является асимптотически эффективным.

Было сделано предположение о том, что разность рисков $R_{t_A} - \bar{R}_{n_0} = O(1)$ при $A \rightarrow \infty$ [1]. Для подтверждения, проведено численное моделирование в среде Matlab, с помощью которого наглядно показано, что разность рисков при использовании последовательной процедуры оценивания может принимать отрицательные значения даже для нормального распределения ошибок.

Результаты численного моделирования

Рассмотрим результаты проведенного численного моделирования на примере данных, ошибки которых соответствуют стандартному нормальному распределению, т. е. $\varepsilon_i \sim N(0, 1)$. Для определенности значения параметра β приняли равными 0.1, 0.2, ..., 0.9 с шагом 0.1, $X_0 = 0$, $A = 400, 1000, 4000, 10\,000$, $m_A = 5$.

В таблице 1 указаны результаты вычислений при различных значениях $A = c^{-1}$. Здесь n_0 – оптимальное количество наблюдений; $\bar{R}_{n_0}$ – оценка риска при оптимальном количестве наблюдений n_0 ; $\hat{\beta}_{t_A}$ – оценка параметра модели линейной авторегрессии в момент остановки t_A ; $\hat{\sigma}_{t_A}^2$ – оценка дисперсии в момент остановки t_A . Под разностью рисков понимается разность между риском R_{t_A} , вычисленным в момент остановки t_A , и оценкой риска $\bar{R}_{n_0}$ при оптимальном количестве наблюдений n_0 .

Рассмотрим полученные результаты.

Таблица 1 - Стандартное нормальное распределение: $\varepsilon^2 \sim N(0,1)$

β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$A=400, n_0 \cong 20$									
$\hat{A}_{n_0}$	40	40	40	40	40	40	40	40	40
Разность рисков	9.06	11.02	13.58	16.29	19.57	24.12	28.99	35.62	41.94
$\hat{\beta}_{t_A}$	0.11	0.22	0.31	0.42	0.53	0.62	0.72	0.81	0.89
$\hat{\sigma}_{t_A}^2$	2.63	2.62	2.7	2.68	2.67	2.72	2.69	2.75	2.7
$A=1000, n_0 \cong 31.63$									
$\hat{A}_{n_0}$	63.24	63.24	63.24	63.24	63.24	63.24	63.24	63.24	63.24
Разность рисков	5.95	9.31	14.17	19.83	27.96	37.19	50.14	65.44	78.26
$\hat{\beta}_{t_A}$	0.12	0.23	0.33	0.44	0.52	0.59	0.74	0.82	0.91
$\hat{\sigma}_{t_A}^2$	2.34	2.34	2.39	2.38	2.43	2.42	2.37	2.34	2.4
$A=4000, n_0 \cong 63.24$									
$\hat{A}_{n_0}$	126.49	126.49	126.49	126.49	126.49	126.49	126.49	126.49	126.49
Разность рисков	-14.49	-9.79	-2.89	6.71	19.52	37.54	62.05	92.81	102.22
$\hat{\beta}_{t_A}$	0.11	0.22	0.32	0.43	0.53	0.64	0.74	0.84	0.92
$\hat{\sigma}_{t_A}^2$	1.73	1.75	1.75	1.74	1.74	1.76	1.79	1.79	1.77
$A=10\,000, n_0 \cong 100$									
$\hat{A}_{n_0}$	200	200	200	200	200	200	200	200	200
Разность рисков	-40.2	-35.04	-27.11	-16.72	-2.63	16.79	44.93	79.24	89.46
$\hat{\beta}_{t_A}$	0.08	0.19	0.29	0.39	0.49	0.59	0.7	0.8	0.91
$\hat{\sigma}_{t_A}^2$	1.68	1.7	1.7	1.7	1.71	1.72	1.72	1.73	1.65

По данным, приведенным в таблице 1, построим график, описывающий разность рисков – см. рисунок 1.

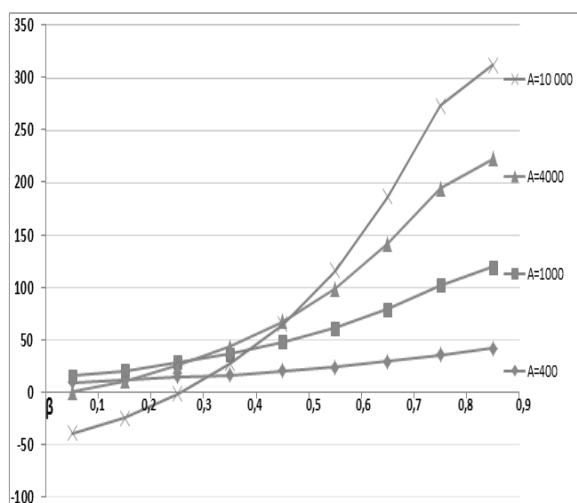


Рисунок 1 - Разность рисков для случая стандартного нормального распределения

Заметим, что разность рисков становится отрицательной при увеличении параметра A . При малых значениях параметра β и цены наблюдений $A = 10\,000$ разность рисков < 0 . При β , лежащих в интервале от $(0.3, 0.5)$, разности рисков примерно одинаковы для всех выбранных значений A , а при $\beta \geq 0.6$ разности положительны.

Следовательно, последовательный подход к оцениванию оказывается эффективным при малых значениях параметра, в то время как байесовский подход эффективен при больших значениях параметра.

Заключение

Численным моделированием продемонстрировано, что использование последовательного оценивания параметра β эффективно минимизирует риск функции потерь, что является подтверждением эффективности последовательного оценивания для параметров модели авторегрессии первого порядка. Последовательный метод оценивания позволяет получить оценки с гарантированным качеством в среднеквадратическом смысле за конечное время.

Список использованных источников

- 1 Sriram T. Sequential Estimation for Time Series Models / T.N.Sriram, R.Iaci // Sequential Analysis: Design Methods and Applications. – 2014. – V. 33. – P. 136-157.
- 2 Sriram T. Sequential Estimation of the autoregressive parameter in a first order autoregressive process / T.N.Sriram// Sequential Analysis: Design Methods and Applications. – 1988. – V. 7(1). – P. 53-74.
- 3 Sriram T. Sequential Estimation of the Mean of a First-Order Stationary Autoregressive Process / T.N.Sriram // The Annals of Statistics. – 1987. – V. 15. – P. 1079-1090.
- 4 Блекуэлл Д. Теория игр и статистических решений / Д. Блекуэлл; пер. с англ. И.В. Соловьева, под ред. Б.А. Севастьянова. – М.: Издательство иностранной литературы, 1958. – 380с