

УДК 517.8 : 519.72

В. В. Савченко, А. А. Шкулев

(Нижний Новгород)

**АДАПТИВНАЯ ОЦЕНКА ПРОГНОЗИРОВАНИЯ
СЛУЧАЙНОГО ВРЕМЕННОГО РЯДА
НА ОСНОВЕ АВТОРЕГРЕССИОННОЙ МОДЕЛИ НАБЛЮДЕНИЙ**

Исследованы динамические свойства линейной оценки прогнозирования случайного временного ряда. Выведено аналитическое выражение для кривой адаптации, характеризующее скорость ее сходимости.

При прогнозировании случайных временных рядов одной из наиболее актуальных является проблема малых выборок наблюдений. Ее эффективное решение большинство исследователей связывает с адаптивным подходом на основе новых методов спектрального анализа с высокими динамическими свойствами [1], которые рассчитаны на оценки прогнозирования линейного вида и согласованную с ними по порядку M авторегрессионную (АР) модель наблюдений [2, 3]. Исследованию скорости сходимости таких оценок в зависимости от применяемых методов адаптации и посвящена предлагаемая работа.

Рассмотрим стационарный случайный процесс $X(n)$ в дискретном времени $n = 1, 2, \dots$. В основе линейной оценки его прогнозирования на один шаг в будущее используется следующее выражение:

$$\hat{x}(n) = \sum_{i=1}^M a_i x(n-i), \quad n=1, 2, \dots, \quad (1)$$

где $\{a_i\}$, $i = \overline{1, M}$, – M -вектор постоянных коэффициентов или параметров оценки. На множестве различных алгоритмов настройки параметров $\{a_i\}$ выражение (1) определяет множество различных модификаций линейной оценки прогнозирования M -го порядка. При этом наилучшая из них отвечает очевидному критерию минимума дисперсии ошибки прогнозирования:

$$\sigma_z^2 = \mathbf{M} |z^2(n)| \rightarrow \min, \quad z(n) = x(n) - \hat{x}(n), \quad n = 1, 2, \dots \quad (2)$$

Здесь \mathbf{M} – символ математического ожидания. Из математической статистики известно [1], что оптимальный в смысле критерия (2) вектор (столбец) параметров $\{a_i\}$ = \mathbf{a} отвечает системе нормальных уравнений вида

$$\mathbf{K}_{xx} \mathbf{a} = \mathbf{k}_x. \quad (3)$$

Здесь \mathbf{K}_{xx} – автокорреляционная ($M \times M$)-матрица анализируемого процесса $X(n)$, а \mathbf{k}_x – M -вектор (столбец) коэффициентов его автокорреляции.

При отсутствии достоверных сведений о корреляционных свойствах анализируемого процесса вместо неизвестных в общем случае корреляционных матрицы \mathbf{K}_{xx} и вектора \mathbf{k}_x в выражение (3) подставляют их статистические оценки, получаемые по выборке конечного объема L . При этом предпочтение отдают методам корреляционного анализа с высокими динамическими свойствами. К их числу относится, например, известный метод Берга [1], который сводится к системе рекуррентных выражений:

$$\left. \begin{aligned} S_{m-1}^2 &= \frac{1}{2(L-m)} \sum_{t=m}^{L-1} [x_{m-1}^2(t) + y_{m-1}^2(t-1)]; \\ \rho_m &= \frac{1}{S_{m-1}^2} \sum_{t=m}^{L-1} x_{m-1}(t)y_{m-1}(t-1)/(L-m); \\ a_m(i) &= a_{m-1}(i) + \rho_m a_{m-1}(m-i), \quad i = \overline{1, M}; \\ \sigma_m^2 &= (1 - \rho_m^2) \sigma_{m-1}^2, \quad \sigma_0^2 = S_0^2; \\ \left. \begin{aligned} x_m(t) &= x_{m-1}(t) + \rho_m y_{m-1}(t-1); \\ y_m(t) &= y_{m-1}(t-1) + \rho_m^* x_{m-1}(t), \quad m = \overline{1, M}, \end{aligned} \right\} \quad (4') \end{aligned} \right\} \quad (4)$$

с инициализацией вида $x_0(t) = y_0(t) = x(n-L+t)$ для всех $t = 0, 1, \dots, L-1$. Здесь $x_m(t)$ и $y_m(t)$ – ошибки линейного предсказания «вперед» и «назад» m -го порядка каждая. Финальное значение рекурсии (4) при $m = M$ определяет результирующую, т. е. адаптированную под выборку $\{x(n)\}$ объема $L > 1$ в ретроспективе АР-модель наблюдений [3]:

$$x(n) = \sum_{i=1}^M a_M(i)x(n-i) + \eta(n), \quad n = 1, 2, \dots, \quad (5)$$

с порождающим «белым» шумом $\eta(n)$ на входе. Его дисперсия $\sigma_\eta^2 = \prod_{m=1}^M (1 - \rho_m^2) \sigma_x^2$ определяется средним квадратом

$$S_M^2 = \frac{1}{2(L-M-1)} \sum_{t=M+1}^{L-1} [x_M^2(t) + y_M^2(t-1)]$$

ошибки предсказания M -го порядка и согласована с выборочной оценкой σ_x^2 дисперсии наблюдаемого процесса. Чем точнее модель (5) адаптирована по выборке $\{x(n)\}$ к процессу $X(n)$, тем точнее становится и результирующая оценка прогнозирования (1). В качестве универсального показателя точности такой оценки можно использовать величину информационного рассогла-

сования АР-модели наблюдений (5) с имеющейся выборкой в метрике Кульбака – Лейблера:

$$I_n[f_R | f_K] = \int \dots \int \ln \left[\frac{f_K(\mathbf{x}_n)}{f_R(\mathbf{x}_n)} \right] f_K(\mathbf{x}_n) d\mathbf{x}_n \rightarrow \min_R.$$

Путем стандартных вычислений при предположении о гауссовости анализируемого процесса получим [4]:

$$I_n[f_K | f_k] = \frac{1}{2} \ln \frac{|\mathbf{R}_{n,n}|}{|\mathbf{K}_{n,n}|} - \frac{n}{2} + \frac{1}{2} \text{tr}[\mathbf{R}_{n,n}^{-1} \mathbf{K}_{n,n}], \quad (6)$$

где $\mathbf{R}_{n,n}$ – $(n \times n)$ -матрица автокорреляции АР-процесса (5); $\mathbf{K}_{n,n}$ – $(n \times n)$ -матрица автокорреляции случайного процесса $X(n)$; символом $|\bullet|$ обозначен определитель квадратной матрицы; $\text{tr}(\cdot)$ – ее след (сумма диагональных элементов). Раскроем правую часть последнего выражения в виде [3]:

$$\begin{aligned} I_n[f_R | f_K] &= \frac{n}{2} \ln \frac{\sigma_\eta^2}{\sigma_0^2} - \frac{n}{2} + \frac{1}{2} \text{tr}[\mathbf{M}(\mathbf{x}_n \mathbf{x}_n^T) \mathbf{R}_{n,n}^{-1}] = \\ &= \frac{n}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2} \mathbf{M}[\mathbf{x}_n^T \mathbf{R}_{n,n}^{-1} \mathbf{x}_n] - H_n^*(X) = \\ &= \frac{n}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2\sigma_\eta^2} \mathbf{M} \left[\sum_{i=0}^{n-1} z^2(n-i) \right] - H_n^*(X) = \frac{n}{2} \left[\ln(2\pi\sigma_\eta^2) + \frac{\sigma_M^2}{\sigma_\eta^2} \right] - H_n^*(X). \end{aligned}$$

Здесь $\mathbf{x}_n = \text{col}_n(x(n-i))$ – n -вектор (столбец) отсчетов анализируемого процесса; $z(n-i) = x(n-i) - \sum_{k=1}^M a_M(k)x(n-k-i)$ – отклик фильтра ошибки предсказания M -го порядка [1] на анализируемый процесс в момент $t = n-i$, $i = 0, 1, \dots, n-1$; $\sigma_M^2 = \mathbf{M}[z^2(n-i)]$ – дисперсия ошибки прогнозирования M -го порядка; $H_n^*(X) = \frac{1}{2} \ln[(2\pi e)^n |\mathbf{K}_{n,n}|] = \frac{n}{2} [\ln(2\pi\sigma_0^2)] + \frac{n}{2}$ – энтропия гауссовского закона распределения анализируемого процесса, пропорциональная минимальному собственному числу σ_0^2 автокорреляционной матрицы $\mathbf{K}_{n,n}$. Переходя от этого выражения к удельной (на один отсчет данных) величине информационного рассогласования, получаем

$$\gamma_n[f_R | f_K] = \frac{1}{n} I_n[f_R | f_K] \Big|_{n \rightarrow \infty} = \frac{1}{2} \left[\ln(2\pi\sigma_\eta^2) + \frac{\sigma_M^2}{\sigma_\eta^2} \right] - h^*(X)$$

или в зависимости от объема выборки наблюдений

$$\gamma(L) = \frac{1}{2} \left[\ln\sigma_\eta^2(L) + \frac{\sigma_M^2(L)}{\sigma_\eta^2(L)} \right], \quad (7)$$

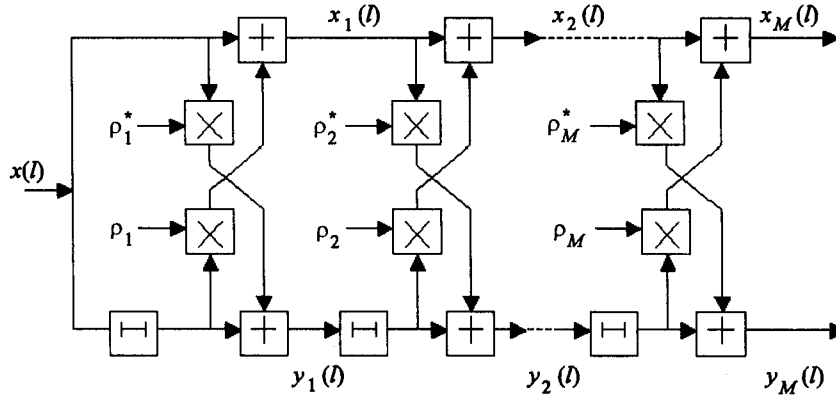


Рис. 1. Структурная схема решетчатого фильтра предсказания

где учтено, что удельная энтропия случайного процесса $X(n)$ никак не связана со способом адаптации его АР-модели.

В зависимости от объема L используемых в (4) наблюдений полученный результат характеризует достигаемую точность прогнозирования в динамике, т. е. может служить в роли кривой ее адаптации (сходимости). При большом объеме наблюдений $L \gg 1$ различия между дисперсиями порождающего шума σ_n^2 и ошибки прогнозирования σ_M^2 практически исчезают. В таком случае выражение (7) сводится к виду

$$\hat{\gamma}(L) = \frac{1}{2} [\ln \sigma_M^2(L) + 1], \quad (8)$$

характерному для общеизвестного критерия минимума выборочной дисперсии ошибки прогнозирования [5].

Раскроем зависимость (7) применительно к методу Берга. Его центральным звеном является решетчатый фильтр предсказания M -го порядка (рис. 1).

Динамика фильтра в дискретном времени описывается системой рекуррентных соотношений (4'). Дисперсия его отклика

$$x_M(n) = x_0(n) + \sum_{m=1}^M \rho_m(L) y_{m-1}(n-1)$$

равна

$$\begin{aligned} \sigma_M^2(L) = \mathbf{M}[x_M^2(n)] = \mathbf{M}[x_0^2(L)] + \\ + 2\mathbf{M}\left[\sum_{m=1}^M \rho_m(L) x_0^*(n) y_{m-1}(n-1)\right] + \mathbf{M}\left\{\left[\sum_{m=1}^M \rho_m(L) y_{m-1}(n-1)\right]^2\right\}. \end{aligned}$$

Учитывая свойство некоррелированности отдельных ступеней решетчатого фильтра, когда выполняются равенства $\mathbf{M}\{y_i(n) y_n^*(n)\} = 0 \quad \forall i \neq n < M$, при

$L \gg 1$ можно записать

$$\sigma_M^2(L) = \sigma_x^2 - \sum_{m=1}^M [\rho_m^2 - \mathbf{D}_m(L)] \sigma_{m-1}^2.$$

Здесь $\sigma_m^2 = \sigma_m^2(L) \Big|_{L \rightarrow \infty} = \mathbf{M}\{x_m^2(n)\}$; $\rho_m = \mathbf{M}\{x_{m-1}(n)y_{m-1}^*(n-1)/\sigma_{m-1}^2\}$;

$\mathbf{D}_m(L) = \mathbf{M}\{\rho_m^2(L) - \rho_m^2\} = \mathbf{D}\{\rho_m(L)\}$ – дисперсия текущей оценки m -го коэффициента отражения ($\mathbf{D}\{\cdot\}$ – символ вычисления дисперсии). После несложных преобразований будем иметь [6]

$$\sigma_M^2(L) = \prod_{m=1}^M (1 - \rho_m^2) \sigma_{m-1}^2 + \sum_{m=1}^M \mathbf{D}_m(L) \sigma_{m-1}^2 = \sigma_{M \min}^2 + \sum_{m=1}^M \mathbf{D}_m(L) \sigma_{m-1}^2,$$

где $\sigma_{M \min}^2 = \sigma_M^2(L) \Big|_{L \rightarrow \infty}$ – предельно достижимое значение (минимум) дисперсии ошибки прогнозирования. Учитывая известное соотношение для дисперсий

$$\mathbf{D}\{\rho_{m-1}(L)\} = \mathbf{D}\{x_{m-1}(n)y_{m-1}^*(n-1)/\sigma_{m-1}^2\}/L,$$

после ряда вычислений окончательно получаем

$$\sigma_M^2(L) = \sigma_{M \min}^2 + \sum_{m=1}^M \frac{\sigma_{m-1}^2 (1 + \rho_m^2)}{L}.$$

В свою очередь, дисперсия порождающего шума в АР-модели наблюдений (5) равна:

$$\begin{aligned} \sigma_\eta^2(L) &= \left[\prod_{m=1}^M (1 - \rho_m^2(L)) \right] \sigma_x^2 = \\ &= \left[\prod_{m=1}^M (1 - (\rho_m^2 + \mathbf{D}_m(L))) \right] \sigma_x^2 = \left[\prod_{m=1}^M \left(1 - \rho_m^2 - \frac{1 + \rho_m^2}{L} \right) \right] \sigma_x^2. \end{aligned}$$

Преобразовывая это выражение, получим

$$\sigma_\eta^2(L) = \left[\prod_{m=1}^M \left(\frac{L-1}{L} (1 - \rho_m^2) - \frac{2\rho_m^2}{L} \right) \right] \sigma_x^2.$$

Для случая слабокоррелированного процесса $X(n)$, когда $\rho_m^2 \ll 1 \forall m > 1$, в первом приближении выполняются равенства:

$$\left. \begin{aligned} \sigma_M^2(L) &= \sigma_{M \min}^2 \left(1 + \frac{MK_\Pi}{L} \right), \\ \sigma_\eta^2(L) &= \sigma_{M \min}^2 \left[1 - \frac{1}{L} \right]^M. \end{aligned} \right\}$$

Здесь $K_{\pi} = \sigma_x^2 / \sigma_{M \min}^2$ – коэффициент подавления анализируемого процесса в решетчатом фильтре (см. рис. 1) установленного порядка M . Подставляя эти равенства в выражение (7), получаем искомую величину информационного рассогласования

$$\gamma(L) = \frac{1}{2} \left(\ln \sigma_{M \min}^2 \left[1 - \frac{1}{L} \right]^M + \frac{1 + \frac{MK_{\pi}}{L}}{\left[1 - \frac{1}{L} \right]^M} \right) \quad (9)$$

в зависимости от параметров АР-модели (5). По своей сути эта зависимость характеризует кривую адаптации исследуемой оценки прогнозирования. Ее график для двух значений порядка $M = 5$ (кривая 1) и $M = 10$ (кривая 2) при $\sigma_{M \min}^2 = 1$ и $K_{\pi} = 10$ показан на рис. 2. Видно, что при увеличении объема наблюдений L кривая адаптации монотонно спадает до своего минимального значения $0,5 \ln \sigma_{M \min}^2 + 0,5$, точно соответствуя асимптотике выражения (7). Чем больше порядок оценки прогнозирования M , тем медленнее эта оценка сходится к своему оптимальному значению.

Выводы теоретических исследований дополнительно подтверждены результатами проведенного эксперимента над анализируемым процессом $X(n)$ типа авторегрессия. Его порядок M^* был установлен равным 10, а коэффициент подавления $K_{\pi} = 10$. Для расчета оценки прогнозирования (1) использовалась АР-модель порядка $M = 5$. Объем выборки L менялся в пределах от 15 до 250. При этом применялся метод Берга в формулировке (4), а в выражение (7) подставлялись выборочные оценки дисперсий порождающего шума σ_{η}^2 и ошибки прогнозирования σ_M^2 , полученные путем усреднения квадратов S_M^2 и $z^2(n)$ соответствующих величин по 1000 независимых реализаций. Полученные результаты отображены на рис. 2 символами «x». Видно, что в целом экспериментальные данные хорошо согласуются с теоретическими при заданном порядке M АР-модели.

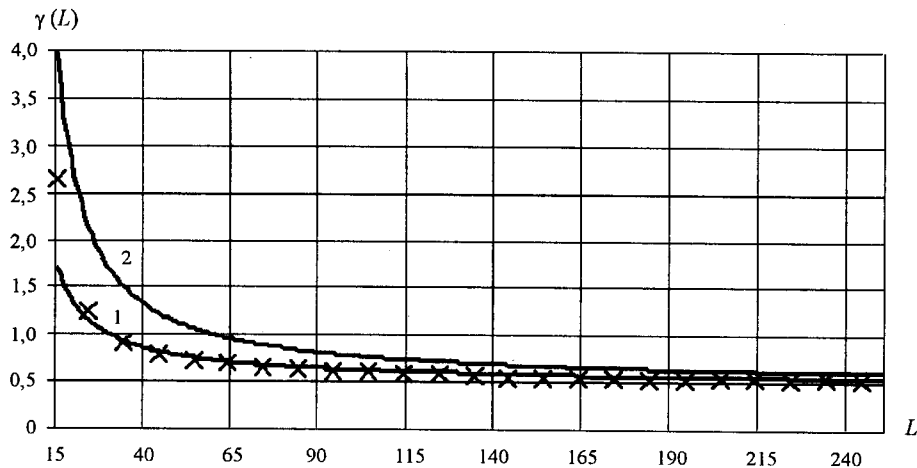


Рис. 2. Кривые адаптации линейной оценки прогнозирования

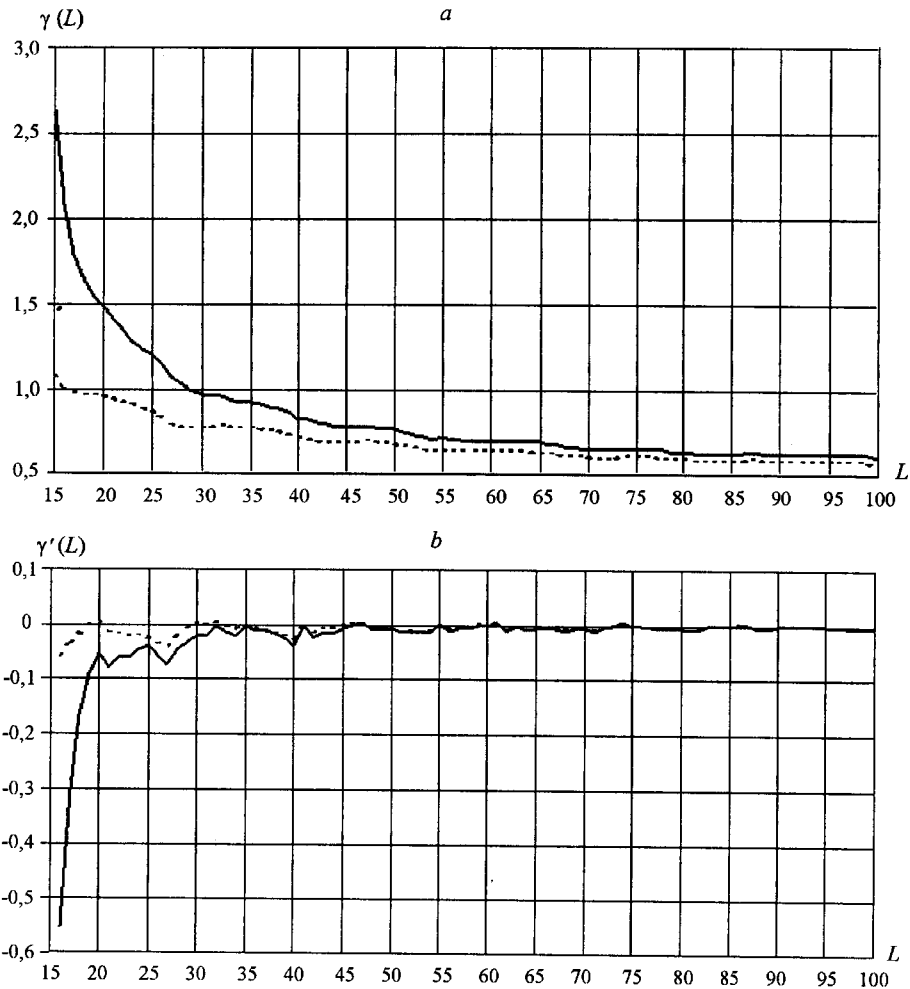


Рис. 3. Экспериментальные зависимости кривых адаптации (а) и их производных (b)

В развитие сделанных выводов на рис. 3 представлены две экспериментальные кривые адаптации (рис. 3, а) и их первые разности, или производные, (рис. 3, б) как характеристика скорости сходимости линейной оценки прогнозирования согласно выражениям (7) (сплошная линия) и (8) (штриховая линия). Сопоставляя эти кривые между собой, приходим к выводу о близости информационного критерия критерию минимума дисперсии ошибки прогнозирования лишь при больших объемах наблюдений, когда параметр адаптации $L/M > 10$. При малых объемах последний критерий дает существенно завышенные оценки скоростных свойств адаптивного алгоритма прогнозирования. Поэтому данным критерием следует пользоваться осторожно: если порядок оценки M сопоставим по своей величине с объемом выборки L , то надо помнить, что стабилизация среднего квадрата ошибки прогнозирования $z^2(n)$ на некотором фиксированном уровне не может служить гарантией того, что все переходные процессы в ней уже завершены в более общем, теоретико-информационном смысле. Напротив, стабилизация информационно-

го показателя (7) одновременно гарантирует сходимость линейной оценки прогнозирования (1) и в отношении дисперсии ее ошибки σ_M^2 .

Таким образом, по результатам проведенного исследования получено аналитическое выражение для кривой адаптации линейной оценки прогнозирования, адаптируемой под конечную выборку наблюдений. Основываясь на кривой адаптации, можно исследовать динамические свойства каждой конкретной реализации оценки прогнозирования (1) в зависимости от применяемого метода адаптации и от ее параметров: порядка M и объема выборки L . Проблема состоит лишь в способе раскрытия зависимостей $\sigma_n^2(L)$ и $\sigma_M^2(L)$. В тех случаях, если эти зависимости не раскрываются в явном виде, можно использовать их статистические или выборочные оценки по конечным реализациям ошибки прогнозирования внутри и вне интервала адаптации: $x_M(n)$ и $z(n)$ соответственно.

СПИСОК ЛИТЕРАТУРЫ

1. Марпл-мл. С. Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990.
2. Савченко В. В. Прогнозирование социально-экономических процессов на основе адаптивных методов спектрального оценивания // Автометрия. 1999. № 3. С. 99.
3. Савченко В. В. Адаптивные методы нелинейного спектрального оценивания на основе принципа минимакса энтропии: Автореф. дис. ... д-ра техн. наук. Н. Новгород: НГТУ, 1993.
4. Кульбак С. Теория информации и статистика. М.: Наука, 1967.
5. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования. М.: Наука, 1984.
6. Савченко В. В. Принципы минимакса энтропии в задачах группового и последовательного спектрального анализа // Изв. вузов. Радиоэлектроника. 1990. № 9. С. 66.

*Нижегородский государственный
лингвистический университет,
E-mail: svv@lunn.sci-nnov.ru*

*Поступила в редакцию
30 марта 2001 г.*