

УДК 519.22./25

Т.А. РыжкинаДальневосточный государственный технический рыбохозяйственный университет,
690087, г. Владивосток, ул. Луговая, 52б**ПОСТРОЕНИЕ АДАПТИВНОЙ МОДЕЛИ КОНТРОЛЯ
СТОХАСТИЧЕСКОГО ПРОЦЕССА**

Рассматривается постановочный пример природного стохастического процесса, генерирующего последовательность случайных чисел (ПСЧ). Процесс представляется в виде временного ряда (ВР), стационарного или нестационарного. Генерация ПСЧ осуществляется с помощью равномерного и нормального распределений случайных величин (СВ). Моделирование ВР выполняется на основе авторегрессии как линейного фильтра ПСЧ. Для краткосрочных прогнозов результата процесса параметры регрессии считаются переменными и оцениваются на каждом шаге с учетом длины (временного лага) входящего потока. Параметры регрессии корректируются (адаптируются) по методу антиградиента квадратической ошибки прогноза, что означает минимизацию ошибки (иначе, по методу «наискорейшего спуска»). Значение работы определяется конкретной процедурой адаптации модели и подбором параметра адаптации, в частности, для стационарного процесса.

Ключевые слова: последовательность случайных чисел, стационарность, авторегрессия, минимизация ошибки прогноза, метод «наискорейшего спуска», автокорреляция, характеристический многочлен, теорема Руше.

Т.А. Ryzhkina**CONSTRUCTION OF ADAPTIVE MODEL OF CONTROL
OF THE STOCHASTIC PROCESS**

The paper deals with a staged example of a natural stochastic process generating a sequence of random numbers (PSP). The process is represented as a time series (TS), stationary or non-stationary. Generation of PSP is carried out using uniform and normal distributions of random variables (RV). TS simulation is performed on the basis of autoregression as a linear PSP filter. For short-term forecasts of the process result, the regression parameters are considered variable and are estimated at each step, taking into account the length of the incoming flow (time lag). The regression parameters are adjusted (adapted) by the anti-gradient method of the square error of the forecast, which means minimization of the error (otherwise, by the method of “fastest descent”). The value of the work is determined by the specific procedure of adapting the model and the selection of the adaptation parameter, in particular, for the stationary process.

Key words: random number sequence, stationarity, autoregression, minimization of the forecast error, the “fastest descent” method, autocorrelation, characteristic polynomial, Rusche theorem.

Введение

Интерес в данной работе представляет получение моделей, лежащих в основе процедуры прогнозирования. Предполагается, прежде всего, получение информации о развитии процесса в настоящий момент, предсказание показателей процесса в течение кратковременного периода с учетом преемственности данных и степени их влияния на процесс. Объектом исследования служит процесс (ВР), генерирующий ПСЧ. Это может быть контроль наличия случайных включений в продукции в допустимых пределах. Генерируется ПСЧ по формуле интегрального закона нормального распределения $N(a, \sigma)$ [1]:

$$P(X < x) = 0,5 + \Phi(x-a/\sigma), \quad (1)$$

где P – вероятность; $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ – функция Лапласа.

С помощью этой формулы по заданной вероятности P попадания нормально распределенной СВ X в интервал $(-\infty, x)$ возвращается точка x . Для этого используется встроенная в ППП Excel функция «Норм. обр.». Предварительно формируется последовательность для вероятностей с помощью инструмента «Генерация случайных чисел» в «Пакете анализа».

Произвольно сформированное равномерное распределение $R(0,1)$, $n = 30$ упорядочивается для придания процессу x_t прикладного смысла. Каждое число из $R(0,1)$ понимается как вероятность $Rt = P(X < x_t)$, t – текущий момент времени [2].

На основании $R(0,1)$, $n = 30$ функция «Норм. обр.» формирует ПСЧ нормального распределения с заданными параметрами $a = 1$, $\sigma = 0,1$ [2]. Выбор числа наблюдений $n = 30$ сделан с учетом возможных значений лага p , т.е. сдвига по времени, и свойств предполагаемой модели имитации. Полученное случайное нормальное распределение центрируется, т.е. вычисляются отклонения элементов ПСЧ от среднего значения $\tilde{x}_t = x_t - a = x_t - 1$ [2].

Таблица 1

Центрированное распределение

Table 1

Centered distribution

t	\tilde{x}_t	t	\tilde{x}_t
1	0,2274	16	0,0160
2	0,1259	17	0,0079
3	0,1076	18	0,0050
4	0,0926	19	0,0034
5	0,0873	20	-0,0050
6	0,0851	21	-0,0124
7	0,0783	22	-0,0137
8	0,0664	23	-0,0311
9	0,0555	24	-0,0347
10	0,0421	25	-0,0380
11	0,0273	26	-0,0384
12	0,0272	27	-0,0513
13	0,0258	28	-0,0519
14	0,0215	29	-0,0866
15	0,0179	30	-0,0938

Текущий член ВР, табл. 1, с помощью авторегрессионной модели представляется взвешенной суммой прошлых p значений, $p = 3, 2, 1$. Рассматривается входящий в формулу поток в виде одного, двух или трех первых элементов при условии, что отсутствует постоянный поток. Таким образом, начальный коэффициент авторегрессии принимается равным нулю. Остальные коэффициенты формулы представляют по отдельности «вес» соответствующего входящего элемента \tilde{x}_t .

В дальнейшем предполагается, что текущий член ВР выражается линейной формулой регрессии, коэффициенты которой могут быть переменными.

1. Авторегрессионная модель одномерного процесса

Формула авторегрессии имеет следующий вид:

$$\tilde{x}_t = \alpha + \sum_{i=1}^p b_i \tilde{x}_{t-i} + \varepsilon_t, t = p+1, \dots, n, b_i = \text{const.} \quad (2)$$

Иногда постоянный входящий поток α не включают в формулу для получения прямой зависимости текущего значения от предыдущих значений, т.е.

$$\tilde{x}_t = \sum_{i=1}^p b_i \tilde{x}_{t-i} + \varepsilon_t, t = p+1, \dots, n, b_i = \text{const.} \quad (3)$$

Коэффициенты формулы (3) могут быть получены методом наименьших квадратов (МНК) в простой форме при условии, что остатки ε_t образуют «белый шум». Это есть нормальный стохастический процесс как совместное нормальное распределение двух его разных сечений [1]. Другой вариант расчета коэффициентов регрессии (3) с нулевым свободным членом возможен с помощью встроенной в ППП Excel функцией «Линейн».

Вводится линейный оператор сдвига назад L (оператор лага) как в [3, 4], действующий так: $Lb_p \tilde{x}_{t-p} = b_p L^p$, $p \geq 0$. Авторегрессионный оператор порядка p как функция оператора лага принимает вид $\varphi(L) = 1 - \sum_{i=1}^p b_i L^i$. Формулы (2), (3) преобразуются соответственно:

$$(1 - \alpha - \sum_{i=1}^p b_i L^i)x_t = \varepsilon_t, \quad (1 - \sum_{i=1}^p b_i L^i)x_t = \varepsilon_t. \quad (4)$$

Компактно формулы (2), (3), (4) записываются так:

$$\varphi(L)\tilde{x}_t = \varepsilon_t. \quad (5)$$

С авторегрессией тесно связана автокорреляционная функция, которая выражается в виде линейной комбинации коэффициентов автокорреляции, с такими же коэффициентами, что и в формуле (5),

$$\rho_k = \sum_{i=1}^p b_i \rho_{k-i}, k = 1, \dots, n/4. \quad (6)$$

В объекте рассмотрения (см. табл. 1) число k ограничивается величиной, равной 7, чтобы не ослаблять поле корреляции [4, 5].

Используя автокорреляционный оператор, (6) можно записать так:

$$\varphi(L)\rho_k = 0. \quad (7)$$

Ядро линейного оператора $\varphi(L)$, [6], приводит к рассмотрению уравнения

$$1 - \sum_{i=1}^p b_i z^i = 0. \quad (8)$$

Это уравнение называют характеристическим [4, 5] по отношению к (5). Устойчивость формулы (5) зависит от корней характеристического уравнения (8).

Под устойчивостью результатов имитации понимается степень их нечувствительности к изменению входных условий. Оценка устойчивости (которая иногда понимается как ста-

ционарность) может быть выполнена разными способами. Чаще всего контролируют дисперсию результатов модели в зависимости от интервала моделирования. Если увеличения дисперсии не происходит, результаты применения модели считаются устойчивыми. Эффективным способом контроля устойчивости формул (3–5) является установление того факта, что все корни характеристического уравнения (8) (действительные и/или комплексные сопряженные) находятся вне единичного круга комплексной плоскости [2, 4, 5].

Решение вопроса о количестве и величинах корней уравнения (8) может быть достигнуто на основании теоремы Руше [7].

Формулы вида (3) с лагом $p = 3, 2, 1$ для ВР, табл. 1, составленные с помощью инструмента «Регрессия» в надстройке «Пакет анализа», имеют соответственно вид [2]:

$$\tilde{x}_t = 1,30664 \tilde{x}_{t-1} - 0,12935 \tilde{x}_{t-2} - 0,01727 \tilde{x}_{t-3} + \varepsilon_t, \sigma_{ocm} = 0,009199. \quad (9)$$

$$\tilde{x}_t = 1,20446 \tilde{x}_{t-1} - 0,22414 \tilde{x}_{t-2} + \varepsilon_t, \sigma_{ocm} = 0,00968. \quad (10)$$

$$\tilde{x}_t = 0,797481 \tilde{x}_{t-1} + \varepsilon_t, \sigma_{ocm} = 0,0167685. \quad (11)$$

В формулах (9–11) приводится оценка случайной ошибки σ_{ocm} .

Модели (10), (11) можно считать устойчивыми (стационарными). Модель (9) свойством устойчивости не обладает [7].

2. Адаптация коэффициентов модели авторегрессии

Итак, для получения более подходящей модели имитации ВР коэффициенты формулы (3) рассматриваются в зависимости от входящего потока и ошибки прогноза в момент $(t + 1)$.

Анализ ошибки прогноза с квадратом

$$e^2_{t+1} = (\tilde{x}_{t+1} - \sum_{i=1}^p b_i \tilde{x}_{t-i})^2 \quad (12)$$

с точки зрения достижения ею минимума в p -мерном пространстве [4] приводит к рассмотрению классической задачи минимизации квадратической функции нескольких переменных градиентным способом. Уменьшение ошибки может быть достигнуто в направлении антиградиента, т.е. по методу «наискорейшего спуска».

Корректировка вектора коэффициентов B может быть выполнена по правилу

$$B_n = B_s - k \text{grad}(e^2_{t+1}), \quad (13)$$

где B_n, B_s, k – векторы новых, старых коэффициентов и скорость движения ($k > 0$) в направлении антиградиента соответственно.

Если теперь в момент $(t+1)$ использовать новые коэффициенты B_n , имея в виду, что ошибка по абсолютной величине должна стать меньше, получится неравенство

$$e^2_n = (\tilde{x}_{t+1} - \sum_{i=1}^p (b_i)_n \tilde{x}_{t-i})^2 = e^2_s (1 - 2k \sum_{i=1}^p x^2_{t-i})^2 < e^2_s, \quad \left| 1 - 2k \sum_{i=1}^p x^2_{t-i} \right| < 1, \quad (14)$$

где e^2_n, e^2_s – новая и старая ошибки с квадратом соответственно.

Скорость k регулируется неравенством (14) достаточно неопределенно [2, 4]. Вводится в рассмотрение параметр адаптации $\alpha \in (0,2)$. Тогда k определяется через α как

$$k = \frac{\alpha}{2 \sum_1^p x^2_{t-i}}. \quad (15)$$

Параметр адаптации α на весь период имитации принимается заданным и постоянным, а k – переменной величиной. Оптимизация модели достигается подбором α с учетом (14).

Коэффициенты b_i преобразуются по формуле

$$(b_i)_n = (b_i)_s + \frac{\alpha e_{t+1} \tilde{x}_{t-i}}{\sum_1^p \tilde{x}^2_{t-i}}. \quad (16)$$

Получена рабочая формула для корректировки коэффициентов на каждом шаге в поле наблюдений и на один шаг вперед за пределы поля для ближайшего прогноза. Второе слагаемое в (16) обозначается символом $\Delta(b_i)_s$. Вычисления адаптированных коэффициентов авторегрессии выполняются в ППП MS Excel.

В качестве примера рассматривается имитация ВР с третьего шага с помощью формулы (10) с входящим потоком из двух элементов (2 фактора, 2 ф.). Корректировка коэффициентов приближения элементов ВР по адаптированной формуле на каждом шаге ($t + 1$), ближайший прогноз на выходе из поля наблюдений, средний уровень приближений приводятся в табл. 2.

На третьем шаге \tilde{x}_3 имитируется с помощью (10) и входного потока \tilde{x}_2, \tilde{x}_1 , вычисляется отклонение e_3 полученного приближения от наблюдения \tilde{x}_3 . Приближение на четвертом шаге выполняется по формуле с адаптированными коэффициентами (16) и входным потоком \tilde{x}_3, \tilde{x}_2 , вычисляется e_4 . Следующая итерация позволяет получить прогноз \tilde{x}_5 . Параметр адаптации выбран равным 0,5. Средний уровень приближений равен 0,022708. Прогноз на шаге 31 равен (-0,10184). Средняя квадратическая погрешность равна 0,000143.

Таблица 2

Адаптация авторегрессии, $p = 2, \alpha = 0,5$

Table 2

Adaptation of autoregression, $p = 2, \alpha = 0.5$

t	\tilde{x}_t	\tilde{x}_{t-1}	\tilde{x}_{t-2}	b_1	b_2	\tilde{x}_t , прогноз	e_{t+1}	$\Delta(b_1)_s$	$\Delta(b_2)_s$
1	0,227427			0	1	0,227427	0	0	0
	0,125888			1	0	0,125888	0	0	0
3	0,107558	0,125888	0,227427	1,204458	-0,22414	0,100651	0,006908	0	0
4	0,092624	0,107558	0,125888	1,218008	-0,20828	0,104787	-0,01216	0,01355	0,015859
5	0,087326	0,092624	0,107558	1,190051	-0,24075	0,084333	0,002993	-0,02796	-0,03246
6	0,085097	0,087326	0,092624	1,198116	-0,23219	0,083121	0,001976	0,008065	0,008554
7	0,078345	0,085097	0,087326	1,203771	-0,22639	0,082667	-0,00432	0,005656	0,005804
8	0,066384	0,078345	0,085097	1,191116	-0,24013	0,072883	-0,0065	-0,01265	-0,01375

Окончание табл. 2

t	\tilde{x}_t	\tilde{x}_{t-1}	\tilde{x}_{t-2}	b_1	b_2	\tilde{x}_t , прогноз	e_{t+1}	$\Delta(b_1)s$	$\Delta(b_2)s$
9	0,055485	0,066384	0,078345	1,170659	-0,26428	0,057009	-0,00152	-0,02046	-0,02414
10	0,042079	0,055485	0,066384	1,165012	-0,27103	0,046648	-0,00457	-0,00565	-0,00676
11	0,027329	0,042079	0,055485	1,145186	-0,29718	0,031699	-0,00437	-0,01983	-0,02614
12	0,027194	0,027329	0,042079	1,121465	-0,3337	0,016607	0,010587	-0,02372	-0,03652
13	0,025792	0,027194	0,027329	1,218314	-0,23637	0,026671	-0,00088	0,096849	0,09733
14	0,021473	0,025792	0,027194	1,21024	-0,24488	0,024555	-0,00308	-0,00807	-0,00851
15	0,017939	0,021473	0,025792	1,18086	-0,28017	0,01813	-0,00019	-0,02938	-0,03529
16	0,016007	0,017939	0,021473	1,178672	-0,28279	0,015072	0,000935	-0,00219	-0,00262
17	0,007945	0,016007	0,017939	1,191612	-0,26829	0,014261	-0,00632	0,01294	0,014502
18	0,004986	0,007945	0,016007	1,113045	-0,42658	0,002015	0,002971	-0,07857	-0,15829
19	0,003393	0,004986	0,007945	1,197218	-0,29244	0,003646	-0,00025	0,084174	0,134131
20	-0,00505	0,003393	0,004986	1,185445	-0,30974	0,002478	-0,00753	-0,01177	-0,0173
21	-0,01237	-0,00505	0,003393	1,698875	-0,65493	-0,0108	-0,00157	0,51343	-0,34519
22	-0,01366	-0,01237	-0,00505	1,753281	-0,63273	-0,01849	0,004834	0,054406	0,022205
23	-0,03111	-0,01366	-0,01237	1,656042	-0,72079	-0,0137	-0,01741	-0,09724	-0,08806
24	-0,03466	-0,03111	-0,01366	1,890619	-0,61781	-0,05037	0,015713	0,234577	0,102979
25	-0,03796	-0,03466	-0,03111	1,765062	-0,7305	-0,03845	0,000492	-0,12556	-0,11269
26	-0,03845	-0,03796	-0,03466	1,761527	-0,73373	-0,04144	0,002984	-0,00354	-0,00323
27	-0,05129	-0,03845	-0,03796	1,741878	-0,75312	-0,03839	-0,0129	-0,01965	-0,0194
28	-0,05188	-0,05129	-0,03845	1,822391	-0,69277	-0,06684	0,014959	0,080514	0,060358
29	-0,08658	-0,05188	-0,05129	1,749486	-0,76485	-0,05153	-0,03505	-0,07291	-0,07208
30	-0,09381	-0,08658	-0,05188	1,898415	-0,67561	-0,12931	0,035506	0,148928	0,089242
31	-0,12931	-0,09381	-0,08658	1,796218	-0,76993	-0,10184	-0,02748	-0,1022	-0,09432
Среднее	0,022116					0,022708	-0,00156		
e^2_{t+1}							0,000143		

В табл. 3 производится корректировка параметров модели длиной $p = 2$ с разными вариантами параметра адаптации на весь период имитации. Из сравнения достигнутых средних значений приближений, прогноза \tilde{x}_{31} , средних квадратических погрешностей результата в табл. 3 следует, что при $\alpha = 0,5$ прогноз более достоверен.

В случае $p = 3$ авторегрессия неустойчивая, и при том же параметре адаптации $\alpha = 0,5$ прогноз результата и его среднеквадратическая погрешность уступают в сравнении с показателями табл. 2.

Таблица 3

Результаты адаптации авторегрессии при $p = 2, 3$ и разных α

Table 3

The results of the adaptation of the autoregression at $p = 2, 3$ and different α

t	BP	$p=2,$ $\alpha=0$	$p=2,$ $\alpha=0,3$	$p=2,$ $\alpha=0,5$	$p=2,$ $\alpha=1,3$	$p=2,$ $\alpha=1,5$	$p=3,$ $\alpha=0$	$p=3,$ $\alpha=0,5$
1	0,227427	0,227427	0,227427	0,227427	0,227427	0,227427	0,227427	0,227427
2	0,125888	0,125888	0,125888	0,125888	0,125888	0,125888	0,125888	0,125888
3	0,107558	0,100651	0,100651	0,100651	0,100651	0,100651	0,107558	0,107558

Окончание табл. 3

t	BP	$p=2,$ $\alpha=0$	$p=2,$ $\alpha=0,3$	$p=2,$ $\alpha=0,5$	$p=2,$ $\alpha=1,3$	$p=2,$ $\alpha=1,5$	$p=3,$ $\alpha=0$	$p=3,$ $\alpha=0,5$
4	0,092624	0,101333	0,103405	0,104787	0,110313	0,111695	0,085532	0,120329
5	0,087326	0,087454	0,085996	0,084333	0,072156	0,067731	0,085679	0,091087
6	0,085097	0,08442	0,083513	0,083121	0,090435	0,096142	0,083809	0,086228
7	0,078345	0,082922	0,082529	0,082667	0,081752	0,077592	0,084124	0,08487
8	0,066384	0,07529	0,07366	0,072883	0,069738	0,07135	0,076492	0,073878
9	0,055485	0,062397	0,058767	0,057009	0,053097	0,05144	0,062117	0,056666
10	0,042079	0,05195	0,047907	0,046648	0,047218	0,048785	0,050572	0,045813
11	0,027329	0,038246	0,033248	0,031699	0,027757	0,025638	0,036501	0,031021
12	0,027194	0,023485	0,018127	0,016607	0,015397	0,016963	0,020819	0,015426
13	0,025792	0,026628	0,025258	0,026671	0,035783	0,036961	0,024833	0,02654
14	0,021473	0,02497	0,023785	0,024555	0,020859	0,018245	0,025529	0,025386
15	0,017939	0,020082	0,0183	0,01813	0,017149	0,018872	0,02009	0,018259
16	0,016007	0,016794	0,015201	0,015072	0,015377	0,014386	0,016271	0,014806
17	0,007945	0,015258	0,014141	0,014261	0,014875	0,015627	0,014938	0,014239
18	0,004986	0,005982	0,003259	0,002015	-0,0034	-0,00533	0,005256	0,001769
19	0,003393	0,004225	0,003321	0,003646	0,010243	0,013802	0,002762	0,002484
20	-0,00505	0,00297	0,002414	0,002478	-0,00205	-0,00647	0,002436	0,00255
21	-0,01237	-0,00684	-0,00919	-0,0108	-0,01097	-0,00489	-0,00788	-0,0112
22	-0,01366	-0,01376	-0,01657	-0,01849	-0,0114	-0,00523	-0,01608	-0,01888
23	-0,03111	-0,01368	-0,0141	-0,0137	-0,00707	-0,01143	-0,01538	-0,01199
24	-0,03466	-0,0344	-0,04479	-0,05037	-0,06385	-0,06378	-0,03677	-0,04393
25	-0,03796	-0,03477	-0,03954	-0,03845	-0,02101	-0,02393	-0,03894	-0,03781
26	-0,03845	-0,03795	-0,04253	-0,04144	-0,04427	-0,04676	-0,03982	-0,03546
27	-0,05129	-0,03781	-0,04041	-0,03839	-0,03437	-0,03329	-0,03943	-0,03495
28	-0,05188	-0,05316	-0,06358	-0,06684	-0,0815	-0,0803	-0,05559	-0,06322
29	-0,08658	-0,05099	-0,05464	-0,05153	-0,03258	-0,03267	-0,05461	-0,05327
30	-0,09381	-0,09265	-0,11865	-0,12931	-0,16793	-0,16533	-0,09768	-0,12465
31	Прогноз		-0,10648	-0,10184	-0,0574	-0,05738		-0,10911
e^2_{t+1}		8,7E-05	0,000101	0,000143	0,000444	0,00044	7,52E-05	0,000153
Среднее	0,022116	0,026745	0,023426	0,022708	0,021857	0,021993	0,025215	0,024562

Заключение

Имеет место оригинальность в постановке исследуемого процесса и в самой процедуре корректировки параметров авторегрессионной модели известным методом. Последовательные итерации при подходящем подборе параметра адаптации на весь период имитации позволяют выйти из поля наблюдений с ближайшим достоверным прогнозом, среднеквадратическая погрешность которого имеет порядок 10^{-5} .

Для физических случайных процессов установлена оптимальная длина устойчивой (стационарной) модели авторегрессии $p \leq 2$ [2, 4]. При построении адаптивной модели за основу лучше брать стационарную модель.

Научные методы и подходы в данной работе могут быть полезны при формировании профессионально ориентированных творческих заданий для студентов.

Список литературы

1. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высш. шк., 2003. 479 с.
2. Рыжкина Т.А., Старовойтова З.П. Моделирование устойчивых фильтров для стохастических процессов // Науч. тр. Дальрыбвтуза. Владивосток: Дальрыбвтуз, 2015. Вып. 34. С. 25–38.
3. Эконометрика: учебник / под ред. В.Б. Уткина. М.: Дашков и К⁰, 2009. 564 с.
4. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003. 416 с.
5. Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. М.: Финансы и статистика, 2007. 576 с.
6. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука, 1972. 496 с.
7. Лаврентьев М.А., Шабат Б.В. Методы теории функций комплексного переменного. М.: Наука, 1965. 716 с.

Сведения об авторе: Рыжкина Тамара Александровна, кандидат физико-математических наук, доцент.