АНАЛИЗ БИНАРНЫХ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, ПОРОЖДЕННЫХ ПРОЦЕССАМИ СКОЛЬЗЯЩЕГО-СРЕДНЕГО

Ю. В. Меленец

Белорусский Государственный Университет
Минск, Беларусь
E-mail: melene@tut.by

Рассматривается задача оценивания параметров и восстановления пропущенных значений процесса скользящего-среднего по наблюдениям над соответствующей бинарной случайной последовательностью. Приводятся результаты имитационных экспериментов.

Ключевые слова: процесс скользящего-среднего, бинарная последовательность, оценивание параметров, восстановление пропущенных значений.

Рассмотрим стационарный случайный процесс u_t, $1 \le t \le n$, $u_t \in R$, который порождает случайную последовательность $(X_t,\ 1 \le t \le n)$ при помощи монотонно неубывающей функции F по следующему правилу:

$$P(X_t = 1 \mid u_t) = F(u_t),\ \ P(X_t = 0 \mid u_t) = 1 - F(u_t). \tag{1}$$

Вопросам исследования свойств последовательности X_t, получаемой при помощи (1), посвящен ряд работ [1-3]. В настоящей статье статистический анализ последовательности X_t проводится для случая, когда F есть функция распределения нормального закона с параметрами $(0,\ b^2)$. Рассматриваются две задачи.

1. В ситуации, когда u_t есть процесс скользящего-среднего $(CC(1))$:

$$u_t = a \cdot \xi_{t-1} + \xi_t, \tag{2}$$

где ξ_t, $t = 1, 2, \ldots$ – последовательность независимых случайных величин, распределенных по нормальному закону с нулевым средним и дисперсией σ^2, необходимо по наблюдениям $x_1, x_2, \ldots, x_n$ оценить неизвестный параметр a из (2) и исследовать свойства получаемых оценок.

2. Построить оценку одного из пропущенных значений в выборке $x_1, x_2, ..., x_n$ по известным значениям процесса X_t до и после пропуска и исследовать свойства этой оценки.

Заметим, что задача 1 обобщает задачу, рассмотренную в [1], где бинарная случайная последовательность X_t порождалась гауссовским случайным процессом u_t при помощи правила:

$$X_t = \begin{cases} 1, & \textit{если} \;\; u_t \geq 0, \\ 0, & \textit{если} \;\; u_t < 0, \end{cases} \tag{3}$$

поскольку (3) получается из (1) при $b \to 0$. Отметим также, что в [2] построены оценки прогноза одного из значений процесса x_t по предшествующему наблюдению. В [3] рассматривалась задача анализа последовательности $(X_t, \, 1 \leq t \leq n)$ в случае, когда u_t, есть процесс авторегрессии первого порядка, Отметим, что в работах [1-2] указаны предметные области, в которых возникают сформулированные математические задачи, и показано, что дают с практической точки зрения полученные решения.

Для решения первой поставленной задачи при известном b^2 оценим неизвестное значение коэффициента a методом максимального правдоподобия. Выбор этого метода обусловлен тем, что для статистического анализа дискретных данных, в нашем случае бинарных, требуется знать совместное распределение всех элементов выборки. В нашей задаче удается получить совместные распределения одного, двух и трех элементов, что достаточно, учитывая свойства последовательности (2).

Последовательность X_t, получаемая при помощи (1) , является стационарной. Определим множества $C(0) = (-\infty, 0)$ и $C(1) = [0, +\infty)$ и воспользуемся теоремой [2].

Пусть бинарный случайный процесс X_t порожден стационарным случайным процессом u_t при помощи соотношения (1). Пусть $Y_t, \, 1 \leq t \leq n$ — последовательность независимых между собой случайных величин с функцией распределения F и Y_t не зависит от X_m, $1 \leq k, m \leq n$. Определим $V_t = u_t - Y_t$. Тогда для $\{x_1, x_2, ..., x_n\} \in \{0,1\}^n$

$$P(X_1 = x_1, ..., X_n = x_n) = P(V_1 \in C(x_1), ..., V_n \in C(x_n)) \tag{4}$$

Поскольку u_t есть гауссовский случайный процесс с нулевым средним и дисперсией $\dfrac{1+a^2}{\sigma^2}$, то V_t есть случайный процесс, распределенный по гауссовскому закону с нулевым математическим ожиданием и дисперсией $\dfrac{1+a^2}{\sigma^2} + b^2$. Для рассматриваемой модели получено, что

$$P(V_t \in C(x_t)) = \frac{1}{2}, \quad P(V_{t-1} \in C(x_{t-1}), \, V_t \in C(x_t)) = \frac{1}{4} + (-1)^{x_{t-1}+x_t} \times \frac{\arcsin\left(\dfrac{\rho(1)}{1+\omega}\right)}{2\pi}, \tag{5}$$

где $\omega = \dfrac{Du_t}{DY_t}$, $\rho(k) = \dfrac{M(u_t \cdot u_{t-k})}{Du_t}$. Из (4) и (5) следует, что вероятность получения наблюдаемой реализации $\{X_t,\ 1 \le t \le n\}$ равна

$$P(X_1 = x_1, \ldots, X_n = x_n) = 2^{n-1} \times \left(\frac{1}{4} - \frac{\arcsin\left(\dfrac{\rho(1)}{1+\omega} \right)}{2\pi} \right)^5 \times \left(\frac{1}{4} + \frac{\arcsin\left(\dfrac{\rho(1)}{1+\omega} \right)}{2\pi} \right)^{n-s-1},$$

где $s = \sum_{t=1}^{n-1} \left| x_{t+1} - x_1 \right|$. Тогда оценки для a и σ^2 удовлетворяют следующей системе

$$\begin{cases} \sigma^2 = \left(a^2 + \sin^2 \dfrac{s\pi}{n-1} \right) \times \dfrac{b^2}{1+a^2} \times \sin^2 \dfrac{s\pi}{n-1} \\ \dfrac{a(1+a^2)\sigma^2}{(1+a^2)\sigma^2 + b^2} = \sin \dfrac{s\pi}{n-1} \end{cases}. \tag{6}$$

Решение последней системы находится только численными методами. Аналитическое исследование свойств оценок, получаемых из системы (6) наталкивается на трудности. Поэтому алгоритм оценивания неизвестных параметров a и σ^2 в рассматриваемой модели реализован на ПК, и исследование возможности его применения проводилось методами статистического моделирования. Приведем в качестве примера результаты просчетов с объемами выборок N=500 и N=3000. Обработке подвергалось по 50 реализаций. Значения параметров, их оценки и доверительные интервалы с уровнем значимости α=0.05 приведены в таблице 1, анализ данных которой свидетельствует о работоспособности рассматриваемого алгоритма.

Таблица 1

Параметры	$N=500$		$N=3000$	
	Оценки	Доверительные интервалы	Оценки	Доверительные интервалы
$b^2=1$				
$a=0{,}2$ $\sigma^2=1$	0,2721 1,993	[0,1921, 0,3521] [1,9301, 2,0559]	0,1879 1.4556	[0,1543, 0,2215] [1,1333, 1,7779]
$a=0{,}2$ $\sigma^2=9$	0,3729 13,091	[0,2856, 0,4602] [0,7856, 0,7956]	0,1942 7,942	[0,0937, 0,2947] [3,0136, 12,8704]
$a=5{,}0$ $\sigma^2=1$	3,4935 2,006	[3,3071, 3,6799] [1,7406, 2,2714]	4,5075 0,955	[3,5367, 5,4783] [0,7856, 1,1244]
$a=10$ $\sigma^2=9$	-15,1936 14,698	[-18,6627, -11,7245] [10,6017, 18,7443]	-11,2415 10,434	[-12,1777, -10,3053] [8,5106, 12,3574]

$b^2=10$				
$a=0{,}2$ $\sigma^2=1$	0,3155 2,543	[0,1812, 0,4498] [1,9815, 3,1045]	0,1571 1.936	[0,0936, 0,2206] [1,1844, 2,6876]
$a=0{,}2$ $\sigma^2=9$	0,5729 15,971	[0,3596, 0,7862] [12,2442, 19,6978]	0,2534 8,112	[0,1931, 0,3137] [6,5445, 9,6795]
$a=5{,}0$ $\sigma^2=1$	3,0509 2,667	[1,1971, 4,9049] [1,2388, 4,0952]	5,912 1,989	[4,7855, 7,0385] [1,0013, 2,9767]
$a=-10$ $\sigma^2=9$	-16,6391 18,445	[-20,1252, -13,153] [11,9301, 24,9599]	-13,201 12,105	[-15,3691, -11,0329] [9,0731, 15,1369]

Для решения второй поставленной задачи требуются совместные распределения индикаторов пересечения нулевого уровня гауссовским случайным процессом V_t. Известны одно-, двух- и трехмерные распределения таких величин [1]:

$$P(X_t = x_t) = \frac{1}{2},\; P(X_t = x_t, X_s = x_s) = \frac{1}{4} + (-1)^{x_t + x_s} \frac{\arcsin(\rho(t-s)/(1+\omega))}{2\pi},$$
$$P(X_t = x_t, X_s = x_s, X_\tau = x_\tau) = \frac{1}{8} + (-1)^{x_t + x_s} \frac{\arcsin(\rho(t-s)/(1+\omega))}{2\pi} +$$
$$+ (-1)^{x_t + x_\tau} \frac{\arcsin(\rho(t-\tau)/(1+\omega))}{4\pi} + (-1)^{x_s + x_\tau} \frac{\arcsin(\rho(s-\tau)/(1+\omega))}{4\pi}. \quad (7)$$

Предположим, что в выборке $x_1, x_2, \ldots, x_n$ пропущено значение x_t, $1<t<n$ и необходимо построить оценку Z_t этого значения по элементам x_s, $s<t$ и x_τ, $\tau>t$. Z_t определим при помощи следующего правила:

$$Z_t = \begin{cases} 1, \text{ если } & P(X_t = 1 / X_s = x_s, X_\tau = x_\tau) \ge \dfrac{1}{2} \\ 0, \text{ если } & P(X_t = 1 / X_s = x_s, X_\tau = x_\tau) < \dfrac{1}{2} \end{cases}, \quad (8)$$

и свойства Z_t охарактеризуем вероятностью $P(Z_t \ne X_t)$. Из (7) и (8) получаем, что если выполнено условие $|\rho(t-\tau)| < |\rho(t-s)|$, то $Z_t = 1$ при $x_s = 1,\ x_\tau = 1$ и $x_s = 0,\ x_\tau = 1$. Если выполнено условие $|\rho(t-\tau)| \ge |\rho(t-s)|$, то $Z_t = 1$ при $x_s = 0,\ x_\tau = 0$ и $x_s = 1,\ x_\tau = 0$. В противном случае получаем, что $Z_t = 0$. Вероятность ошибки равна $P(Z_t \ne X_t) = \frac{1}{2} - \frac{\arcsin(\alpha/(1+\omega))}{\pi}$, где $\alpha = \max(|\rho(t-\tau)|, |\rho(t-s)|)$. Но вероятность ошибки оценивания значения X_t при помощи нерандомизированного правила по значению $x_{t-\beta}$ равна $\frac{1}{2} - \frac{\arcsin(\rho(\beta)/(1+\omega))}{\pi}$. Значит, из двух элементов выборки x_s и x_τ для построения оценки Z_t достаточно брать одно значение, имеющее большую корреляцию с X_t. В нашем случае, поскольку u_t есть процесс скользящего-среднего первого порядка, то оценку Z_t достаточно строить по ближайшему к ней значению. Отметим также, что вероятность ошибки уменьшается при уменьшении b^2.

ЛИТЕРАТУРА

1. *Kedem B.* Binary Time Series /. B.Kedem. New York; Dekker. 1990. 243 p

2. *Keenan, D. M.* A time series analysis od binary data / D. M. Keenan // J. of American Statistical Association. 1982. V. 77. № 4. P. 816–821

3. *Меленец, Ю. В.* Анализ бинарных случайных последовательностей, порождаемых авторегрессиями / Ю. В. Меленец // Вестник Белорусского государственного университета, 1987. Серия 1, № 3. С. 72–74.