УДК 519.24

Д. В. Королюк

Дифузійна апроксимація статистичних експериментів з наполегливою нелінійною регресією і еквілібріумом

(Представлено академіком НАН України І. М. Коваленком)

Запропоновано апроксимацію статистичних експериментів з наполегливою нелінійною регресією і еквілібріумом, марковськими процесами в дискретно-неперервному часі: $k = [Nt]$, $0 \leqslant t \leqslant T$, для яких обгрунтовано дифузійну апроксимацію типу Орнштейна–Уленбека з неперервним часом.

У роботі пропонується математична модель статистичних експериментів (СЕ) з наполегливою нелінійною регресією і еквілібріумом в дискретно-неперервному часі. Дифузійна апроксимація моделі здійснюється процесом типу Орнштейна–Уленбека з неперервним часом. Висхідна модель ланцюга Маркова в дискретно-неперервному часі виникає в результаті масштабування дискретного часу, а також параметрів статистичних експериментів.

Початковим об'єктом досліджень є послідовність бінарних статистичних експериментів (вимірювань) відносних частот присутності або відсутності певної ознаки. А саме, розглядається послідовність (за k) вибірок $\{\delta_r(k) = \pm 1, 1 \leqslant r \leqslant N\}$ незалежних у сукупності при кожному $k \geqslant 0$ і однаково розподілених бінарних випадкових величин.

Вибірка містить сукупність елементарних результатів присутності ознаки в системі. Параметр r індексує бінарний результат ± 1 у вибірці; параметр k можна інтерпретувати як порядковий номер експерименту, або дискретний час.

Послідовність (за k) бінарних СЕ задається сумами вибіркових значень елементарних подій

$$S_N(k) := \frac{1}{N}\sum_{r=1}^{N}\delta_r(k), \qquad k \geqslant 0. \tag{1}$$

Фундаментальну роль відіграє умова *наполегливої* регресії: умовні ймовірності елементарних подій

$$P\{\delta_r(k+1) = \pm 1 \mid S_N(k) = s\} = P_{\pm}(s), \qquad |s| \leqslant 1, \qquad k \geqslant 0, \tag{2}$$

залежать від сумарного попереднього значення $S_N(k) = s$.

Функція наполегливої регресії (умовне математичне сподівання СЕ (1))

$$E[S_N(k+1) \mid S_N(k) = s] := C(s) \tag{3}$$

не залежить від дискретного параметра часу k. Очевидно, що

$$C(s) = P_+(s) - P_-(s). \tag{4}$$

Враховуючи очевидне співвідношення

$$P_+(s) + P_-(s) = 1, \qquad \forall\, s \in [-1, +1],$$

ймовірності елементарних подій виражаються через функцію регресії

$$P_{\pm}(s) = \frac{1}{2}[1 \pm C(s)]. \tag{5}$$

Динаміка СЕ (1)–(3) досліджується в моделі ланцюга Маркова з дискретно-неперервним часом для приростів СЕ:

$$\Delta S_N(k) = S_N(k+1) - S_N(k). \tag{6}$$

Відповідна функція регресії приростів

$$E[\Delta S_N(k) \mid S_N(k) = s] := C_0(s) \tag{7}$$

пов'язана з висхідною функцією регресії (3)–(4) співвідношенням

$$C_0(s) = C(s) - s. \tag{8}$$

Функція наполегливої регресії приростів (7) пропонується у вигляді добутка двох співмножників. Лінійний співмножник задається флюктуацією СЕ (аналогічно [1, 2]); нелінійний співмножник визначається функцією, що набуває нульових значень у граничних точках $+1$, -1.

1. Основне припущення. Функція регресії приростів СЕ є кубічною параболою:

$$C_0(s) = -V(1-s^2)(s-\rho), \qquad |s| \leqslant 1, \qquad |\rho| < 1, \tag{9}$$

яка має три дійсних кореня ± 1 і точку рівноваги ρ:

$$C(\rho) = \rho, \qquad C_0(\rho) = 0. \tag{10}$$

Зв'язок між функціями регресії $C_0(s)$, $C(s)$ і точкою рівноваги ρ представлений графіком (рис. 1).

Лінійний член $s - \rho$ у виразі (9) є *флюктуацією СЕ*, тобто відхиленням поточного значення s від точки рівноваги ρ.

Зауваження 1. Функція регресії приростів (9) має таке зображення як функція флюктуацій:

$$C_0(s) = -V\sigma^2(s-\rho) + V(s-\rho)^2(s+\rho), \qquad \sigma^2 = 1 - \rho^2, \tag{11}$$

отже,

$$\widehat{C}_0(\widehat{S}_N(k)) := C_0(S_N(k)), \tag{12}$$

де

$$\begin{aligned} &\widehat{C}_0(\widehat{s}) := -V\sigma^2\widehat{s} + h(\widehat{s})\widehat{s}^2 = -V[\sigma^2 - \widehat{s}(2\rho + \widehat{s})]\widehat{s}, \\ &h(\widehat{s}) = V(\widehat{s} + 2\rho), \qquad \widehat{s} := s - \rho. \end{aligned} \tag{13}$$

2. Марковські процеси, що апроксимують СЕ. Наполеглива регресія (9) породжує різницеве стохастичне рівняння для флюктуацій $\widehat{S}_N(k) := S_N(k) - \rho$ (див. [1])

$$\Delta\widehat{S}_N(k+1) = \frac{\widehat{C}_0(\widehat{S}_N(k)) + \widehat{\mu}_N(k+1)}{\sqrt{N}}, \qquad k \geqslant 0, \tag{14}$$

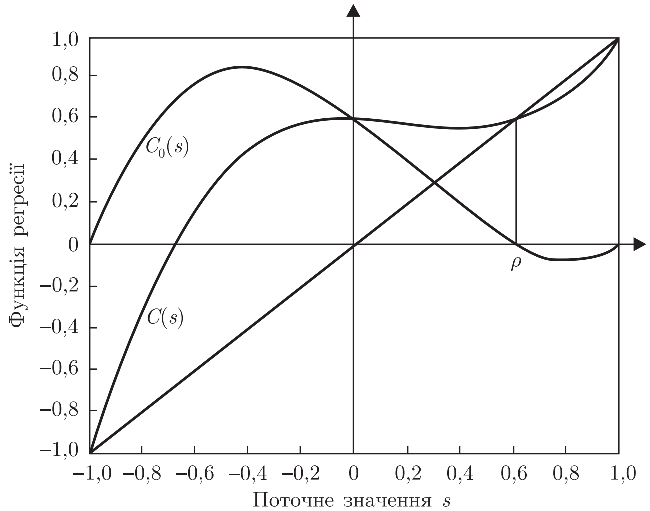

Рис. 1. $C_0(s)$ — кубічна парабола, $C(s)$ — нахилена кубічна парабола

де $\widehat{C}_0(\widehat{s})$ визначено у формулі (13), а мартингал-різниці

$$\widehat{\mu}_N(k+1) := \sqrt{N}[\Delta\widehat{S}_N(k+1) - \widehat{C}_0(\widehat{S}_N(k))], \qquad k \geqslant 0, \tag{15}$$

характеризуються умовними першими двома моментами [4, 6]

$$\begin{aligned} &E[\widehat{\mu}_N(k+1) \mid \widehat{S}_N(k)] = 0, \qquad k \geqslant 0, \\ &E[\widehat{\mu}^2_N(k+1) \mid \widehat{S}_N(k)] = B[\rho + \widehat{S}_N(k)], \qquad B(s) = 1 - C^2(s). \end{aligned} \tag{16}$$

Різницеве стохастичне рівняння (14) служить основою апроксимації СЕ (14) марковським процесом у дискретно-неперервному часі $k = [Nt]$, $t \geqslant 0$.

Пропозиція 1. *Марковський процес в дискретно-неперервному часі задається різницевим стохастичним рівнянням з нелінійною передбачуваною частиною*

$$\Delta\zeta_N(t) = \frac{\widehat{C}_0(\zeta_N(t))}{N} + \frac{\Delta\mu_N(t)}{\sqrt{N}}, \qquad 0 \leqslant t \leqslant T, \tag{17}$$

де прирости марковських процесів у дискретно-неперервному часі визначаються як

$$\Delta\zeta_N(t) := \zeta_N\left(t + \frac{1}{N}\right) - \zeta_N(t), \tag{18}$$

а мартингальна компонента в рівнянні (17) *характеризується нульовим середнім і сталою дисперсією*

$$E[\Delta\mu_N(t) \mid \zeta_N(t)] = 0, \qquad E[\Delta\mu^2_N(t) \mid \zeta_N(t)] = \sigma^2 := B(\rho) = 1 - \rho^2. \tag{19}$$

Зауваження 2. Нормовані СЕ (14) множником $\sqrt{N}$

$$\widehat{\zeta}_N(k) := \sqrt{N}\widehat{S}_N(k) = \sqrt{N}[S_N(k) - \rho] \tag{20}$$

задовольняють різницеве стохастичне рівняння з лінійною складовою передбачуваної компоненти

$$\Delta\widehat{\zeta}_N(k+1) = -V\sigma^2\widehat{S}_N(k) + h(\widehat{S}_N(k))\widehat{S}_N^2(k)/\sqrt{N} + \widehat{\mu}_N(k+1), \qquad k \geqslant 0. \tag{21}$$

Різницеве стохастичне рівняння (21) служить основою апроксимації СЕ (21) марковським процесом у дискретно-неперервному часі.

Пропозиція 2. *Марковський процес у дискретно-неперервному часі задається різницевим стохастичним рівнянням з лінійною передбачуваною частиною*

$$\Delta\zeta_N^0(t) = -\frac{V\sigma^2\zeta_N^0(t))}{N} + \frac{\Delta\mu_N(t)}{\sqrt{N}}, \qquad 0 \leqslant t \leqslant T. \tag{22}$$

Зауваження 3. Зрозуміло мотивацію побудови різницевих стохастичних рівнянь (22). Нелінійна складова передбачуваної компоненти знехтується, бо прямує до нуля при $N \to \infty$. Мартингальна компонента залишається зі сталою дисперсією (19) з нормуючим множником $1/\sqrt{N}$ [3].

3. Дифузійна апроксимація СЕ в дискретно-неперервному часі. Марковські процеси в дискретно-неперервному часі, що задаються різницевими стохастичнмим рівняннями (17) та (22), допускають дифузійну апроксимацію процесом типу Орнштейна–Уленбека в неперервному часі.

Теорема 1. *За умови збіжності за ймовірністю початкових значень марковських процесів* (17) *та* (22)

$$\zeta_N(0) \xrightarrow{P} \zeta^0, \qquad \zeta_N^0(0) \xrightarrow{P} \zeta^0, \qquad N \to \infty, \tag{23}$$

має місце збіжність скінченновимірних розподілів процесів

$$\zeta_N(t) \xrightarrow{D} \zeta(t), \qquad \zeta_N^0(t) \xrightarrow{D} \zeta^0(t), \qquad N \to \infty, \tag{24}$$

до граничних процесів Орнштейна–Уленбека, що задаються генераторами

$$L\varphi(s) = \widehat{C}_0(s)\varphi'(s) + \frac{1}{2}\sigma^2\varphi''(s), \qquad L^0\varphi(s) = -V\sigma^2 s\varphi'(s) + \frac{1}{2}\sigma^2\varphi''(s) \tag{25}$$

на класі фінітних числових функцій $\varphi(s) \in C^3(R)$ $(R := (-\infty, +\infty))$, *тричі неперервно диференційованих з обмеженими похідними.*

Зауваження 4. Дифузійний процес, що виражається генератором L (25), є процесом типу Орнштейна–Уленбека, який задається розв'язком стохастичного диференціального рівняння

$$d\zeta(t) = \widehat{C}_0(\zeta(t))dt + \sigma dW(t), \tag{26}$$

в якому $W(t)$, $t \geqslant 0$, є стандартним процесом броунівського руху з характеристиками

$$EW(t) = 0, \qquad EW^2(t) = t. \tag{27}$$

Доведення теореми 1. Збіжність в (24) процесу $\zeta_N(t)$ до граничного процесу, що задається генератором L, проводиться аналогічно випадку лінійної наполегливої регресії, яка досліджена в [1]. Доведення збіжності в (24) процесу $\zeta_N(t)$ базується на застосуванні збіжності генераторів марковського процесу на достатньо широкому класі тест-функцій з аргументом в області значень марковського процесу. Це забезпечує збіжність скінченновимірних розподілів марковських процесів [4, 5].

Введемо генератор марковського процесу в схемі серій

$$L_N\varphi(s) = NE[\varphi(s+\Delta\zeta_N(t)) - \varphi(s) \mid \zeta_N(t) = s]. \tag{28}$$

Лема 1. *Нехай має місце збіжність генераторів* (28)

$$\lim_{N\to\infty} L_N\varphi(s) = L\varphi(s), \qquad \varphi(s) \in C^3(R), \tag{29}$$

на класі $C^3(R)$ числових фінітних функцій, тричі неперервно диференційованих з обмеженими похідними. Тоді граничний генератор

$$L\varphi(s) = \widehat{C}_0(s)\varphi'(s) + \frac{1}{2}\sigma^2\varphi''(s), \qquad \sigma^2 = 1-\rho^2, \tag{30}$$

задає граничний процес типу Орнштейна–Уленбека (26).

Доведення леми 1. Для приростів марковського процесу нормованих флюктуацій $\zeta_N(t)$, $t \geqslant 0$, обчислимо перші два моменти.

Враховуючи (22), маємо

$$E[\Delta\zeta_N(t) \mid \zeta_N(t) = s] = \frac{\widehat{C}_0(s)}{N}, \tag{31}$$

$$E[(\Delta\zeta_N(t))^2 \mid \zeta_N(t) = s] = \frac{\sigma^2}{N} + O\left(\frac{1}{N^2}\right). \tag{32}$$

Тут, як і раніше (див. (19)),

$$\sigma^2 = B(\rho) = 1 - C^2(\rho) = 1 - \rho^2. \tag{33}$$

Тепер застосуємо формулу Тейлора у виразі (28) генератора L_N до тест-функції $\varphi(s) \in C^3(R)$:

$$\begin{aligned} L_N\varphi(s) = N\biggl[&E[\Delta\zeta_N(t) \mid \zeta_N(t) = s]\varphi'(s) + \\ &+ E[(\Delta\zeta_N(t))^2 \mid \zeta_N(t) = s]\frac{1}{2}\varphi''(s) + R_N\varphi(s)\biggr]. \end{aligned} \tag{34}$$

Тут залишковий член, завдяки обмеженості компонентів рівняння (17), за умовою теореми 1, має оцінку

$$R_N\varphi(s) = O\left(\frac{1}{N^{3/2}}\right).$$

Використовуючи вирази (31), (32) перших двох моментів приростів, маємо

$$L_N\varphi(s) = L\varphi(s) + R_N\phi(s), \tag{35}$$

де залишковий член

$$R_N\varphi(s) \to 0, \qquad N \to \infty, \qquad \varphi(s) \in C^3(R). \tag{36}$$

Твердження леми 1 випливає з (35), (36).

За теоремою 1 А. В. Скорохода [4; 2 : 1] випливає збіжність (24) скінченновимірних розподілів нормованих флюктуацій СЕ.

Теорему 1 доведено.

Враховуючи апроксимацію СЕ нормальним процесом авторегресії, сформульовану в [7, пропозиція 4.1], доцільно ввести конструктивну модель СЕ в дискретно-неперервному часі.

Пропозиція 3. *Марковський процес у дискретно-неперервному часі, що апроксимує СЕ з нелінійною регресією приростів* (9)*, може бути заданий нормальним процесом авторегресії*

$$\Delta\widetilde{\zeta_N^0}(t) = \frac{\widehat{C}_0(\widetilde{\zeta_N^0}(t))}{N} + \sigma\Delta W_N(t). \tag{37}$$

Тут за означенням прирости вінерівського процесу

$$\Delta W_N(t) := W\left(t + \frac{1}{N}\right) - W(t)$$

мають другий момент

$$E[\Delta W_N(t)]^2 = \frac{1}{N}.$$

Для конструктивної моделі апроксимації СЕ марковським процесом (37) має місце аналог теореми 1.

Таким чином, за умови збіжності за ймовірністю початкових значень

$$\widetilde{\zeta}_N^0(0) \xrightarrow{P} \widetilde{\zeta}^0, \qquad N \to \infty, \tag{38}$$

має місце збіжність за розподілом скінченновимірних розподілів процесу нормальної авторегресії (37)

$$\widetilde{\zeta}_N^0(t) \xrightarrow{D} \widetilde{\zeta}^0(t), \qquad N \to \infty, \tag{39}$$

до дифузійного процесу $\widetilde{\zeta}^0(t)$, $t \geqslant 0$, що визначається генератором (див. (25))

$$L\varphi(s) = \widehat{C}_0(s)\varphi'(s) + \frac{1}{2}\sigma^2\varphi''(s) \tag{40}$$

на класі фінітних числових функцій $\varphi(s) \in C^3(R)$ $(R := (-\infty, +\infty))$, тричі неперервно диференційованих з обмеженими похідними.

1. *Королюк Д. В.* Диффузионная аппроксимация статистических экспериментов с настойчивой линейной регрессией и эквилибриумом // Доп. НАН України. – 2014. – № 3. – С. 18–24.
2. *Королюк Д. В.* Бинарные повторяющиеся статистические эксперименты с настойчивой линейной регрессией // Укр. мат. вестн. – 2013. – **10**, № 4. – С. 497–506.

3. *Ethier S. N., Kurtz T. G.* Markov processes: characterization and convergence. – New York: Wiley, 1986. – 534 p.
4. *Скороход А. В.* Асимптотические методы теории стохастических дифференциальных уравнений. – Киев: Наук. думка, 1987. – 328 с.
5. *Koroliuk V. S., Limnios N.* Stochastic systems in merging phase. – Singapore; London: Space, World Scientific, 2005. – 331 p.
6. *Гихман И. И., Скороход А. В.* Стохастические дифференциальные уравнения и их приложения. – Киев: Наук. думка, 1982. – 611 с.
7. *Королюк Д. В.* Бінарні статистичні експерименти з наполегливою нелінійною регресією // Теорія ймовірностей і матем. статистика. – 2014. – № 91.

Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ

Надійшло до редакції 20.02.2014

Д. В. Королюк

Диффузионная аппроксимация статистических экспериментов с настойчивой нелинейной регрессией и эквилибриумом

Предложена аппроксимация статистических экспериментов с настойчивой нелинейной регрессией и эквилибриумом, марковскими процессами в дискретно-непрерывном времени: $k = [Nt]$, $0 \leqslant t \leqslant T$, для которых обоснована диффузионная аппроксимация типа Орнштейна–Уленбека с непрерывным временем.

D. V. Koroliouk

Diffusion approximation of statistical experiments with persistent non-linear regression and equilibrium

We propose an approximation of statistical experiments with persistent non-linear regression and equilibrium by Markov processes in the discrete-continuous time: $k = [Nt]$, $0 \leqslant t \leqslant T$, for which the diffusion approximation of an Ornstein–Uhlenbeck type process with continuous time is constructed.