УДК 519.24

Д. В. Королюк

Статистичні експерименти з наполегливою лінійною регресією в марковському випадковому середовищі

(Представлено академіком НАН України І. М. Коваленком)

Статистичні експерименти (СЕ) з наполегливою нелінійною регресією розглядаються в дискретно-неперервному часі $k = [Nt]$, $0 \leqslant t \leqslant T$. Напрямні параметри функції регресії приростів залежать від станів вкладеного ланцюга Маркова в однорідному (у часі) рівномірно ергодичному марковському процесі, який описує стани випадкового середовища. СЕ задаються розв'язками різницевих стохастичних рівнянь з двома компонентами: передбачувальної та стохастичної (мартингал-різницями). Одержана апроксимація в схемі серій з параметром серії N (об'єм вибірки), при $N \to \infty$, є дифузійним процесом типу Орнштейна–Уленбека. Параметри зсуву і дифузії визначаються усередненням за стаціонарним розподілом вкладеного ланцюга Маркова.

1. Марковський процес у дискретно-неперервному часі. У попередній роботі [1] *статистичні експерименти* (СЕ) в дискретно-неперервному часі задаються розв'язками *різницевих стохастичних рівнянь* зі сталими параметрами напрямної дії.

В даній роботі марковські СЕ (МСЕ) розглядаються у *марковському випадковому середовищі.* Параметри напрямної дії залежать від станів ланцюга Маркова, що визначає зміну станів марковського випадкового середовища.

Марковське випадкове середовище розглядається у двох припущеннях:

1. Дискретне марковське випадкове середовище задається *ланцюгом Маркова* x_k, $k \geqslant 0$, у стандартному фазовому просторі станів $(E, \mathcal{E})$ з перехідними ймовірностями

$$P(x, B) = \mathcal{P}\{x_{k+1} \in B \mid x_k = x\}, \qquad x \in E, \qquad B \in \mathcal{E}. \tag{1}$$

2. *Неперервне марковське випадкове середовище* задається *марковським процесом* (МП) $x(t)$, $t \geqslant 0$, у стандартному фазовому просторі станів $(E, \mathcal{E})$ з перехідними ймовірностями (1) вкладеного ланцюга Маркова $x_k = x(\tau_k)$, $k \geqslant 0$, та розподілами часів перебування в станах θ_{k+1}, $k \geqslant 0$:

$$\mathcal{P}\{\theta_{k+1} \geqslant t \mid x_k = x\} = \exp[-q(x)t], \qquad t \geqslant 0. \tag{2}$$

Так що моменти відновлення мають вигляд

$$\tau_{k+1} = \tau_k + \theta_{k+1}, \qquad k \geqslant 0. \tag{3}$$

1.1. Дискретне марковське випадкове середовище.

Основне припущення 1. МСЕ у марковському випадковому середовищі з дискретно-неперервним часом $t = k/N$, $k \geqslant 0$, задається розв'язком різницевого стохастичного рівняння

$$\Delta\zeta_N\left(t + \frac{1}{N}\right) = -\frac{1}{N} V(x_k)\zeta_N(t) + \frac{1}{\sqrt{N}}\sigma(x_k)\Delta\mu_N\left(t + \frac{1}{N}\right). \tag{4}$$

Параметри напрямної дії $V(x)$ та $\sigma(x)$, $x \in E$ є обмеженими числовими функціями станів $x \in E$ ланцюга Маркова x_k, $k \geqslant 0$, що визначає марковське випадкове середовище.

Тут за означенням

$$\Delta\zeta_N\left(t + \frac{1}{N}\right) := \zeta_N\left(t + \frac{1}{N}\right) - \zeta_N(t), \qquad t \geqslant 0.$$

Стохастична компонента $\Delta\mu_N(t)$, $t \geqslant 0$, задається мартингал-різницями, що характеризуються першими двома моментами:

$$E\left[\Delta\mu_N\left(t + \frac{1}{N}\right)\right] = 0, \qquad E\left[\left[\Delta\mu_N\left(t + \frac{1}{N}\right)\right]^2 \middle| \zeta_N(t)\right] = 1. \tag{5}$$

МСЕ, що задаються розв'язками різницевих стохастичних рівнянь (4), (5), допускають апроксимацію в схемі серій з параметром серії $N \to \infty$ нормальним процесом Орнштейна–Уленбека типу з неперервним часом при додаткових умовах.

Теорема 1. *Нехай виконуються такі умови:*

У1. Ланцюг Маркова з перехідними ймовірностями (1) *є рівномірно ергодичним зі стаціонарним розподілом* $\rho(B)$, $B \in \mathcal{E}$:

$$\rho(B) = \int\limits_E \rho(dx)P(x, B), \qquad \rho(\mathcal{E}) = 1. \tag{6}$$

У2. Мартингал-різниці задовольняють умови центральної граничної теореми [3].

У3. Має місце збіжність початкових умов:

$$\zeta_N(0) \xrightarrow{P} \zeta(0), \qquad N \to \infty, \qquad E|\zeta_N(0)| \leqslant C < +\infty. \tag{7}$$

Тоді має місце збіжність скінченновимірних розподілів МСЕ:

$$\zeta_N(t) \xrightarrow{D} \zeta(t), \qquad N \to \infty, \qquad t \in [0, T]. \tag{8}$$

Граничний процес $\zeta(t)$, $t \geqslant 0$, *задається генератором*

$$\mathbb{L}\varphi(s) = -\widehat{V}\varphi'(s) + \frac{1}{2}\widehat{\sigma}^2\varphi''(s). \tag{9}$$

Параметри напрямної дії $\widehat{V}$ *та дисперсії* $\widehat{\sigma}^2$ *обчислюються за формулами усереднення*

$$\widehat{V} = \int\limits_E \rho(dx)V(x), \qquad \widehat{\sigma}^2 = \int\limits_E \rho(dx)\sigma^2(x). \tag{10}$$

Зауваження 1. Граничний процес $\zeta(t)$, $t \geqslant 0$, з генератором (9), (10) є нормальним дифузійним процесом Орнштейна–Уленбека типу, що визначається стохастичним диференціальним рівнянням

$$d\zeta(t) = -\widehat{V}\zeta(t)dt + \widehat{\sigma}dW(t), \qquad t \geqslant 0. \tag{11}$$

Тут $W(t)$, $t \geqslant 0$, є стандартним процесом броунівського руху з параметрами

$$E[W(t)] = 0, \qquad E[[W(t)]^2] = t.$$

1.2. Неперервне марковське випадкове середовище. На відміну від МСЕ в дискретному марковському випадковому середовищі, в якому динаміка МСЕ визначається в кожний (дискретно-неперервний) момент часу розв'язком різницевого стохастичного рівняння (4), динаміка МСЕ в неперервному марковському випадковому середовищі визначається приростами МСЕ в моменти відновлення МП $x(t)$, $t \geqslant 0$.

Основне припущення 2. МСЕ в неперервному марковському випадковому середовищі задається розв'язками різницевих стохастичних рівнянь[1]

$$\Delta\zeta_N(\tau_{k+1}) = -\frac{1}{N}V(x_k)\zeta_N(\tau_k) + \frac{1}{\sqrt{N}}\sigma(x_k)\Delta\mu_N(\tau_{k+1}). \tag{12}$$

Стохастична компонента $\Delta\mu_N(\tau_k)$, $k \geqslant 0$, характеризується першими двома моментами:

$$E[\Delta\mu_N(\tau_{k+1})|\zeta_N(\tau_k)] = 0, \qquad E[[\Delta\mu_N(\tau_{k+1})]^2|\zeta_N(\tau_k)] = 1. \tag{13}$$

Введемо супровідні функціонали МП $x(t)$, $t \geqslant 0$:

$$\nu(t) := \max\{k\colon \tau_k \leqslant t\}, \qquad t \geqslant 0;$$

$$\tau(t) := \tau_{\nu(t)}, \qquad \theta(t) := \tau_{\nu(t)+1} - \tau(t).$$

МСЕ у дискретно-неперервному часі $t = k/N$, $k \geqslant 0$, задається сумою приростів:

$$\zeta_N(t) = \zeta_N(0) + \sum_{k=1}^{\nu(Nt)} \Delta\zeta_N(\tau_k), \qquad t = \frac{k}{N}, \qquad k \geqslant 0. \tag{14}$$

Теорема 2. *Нехай виконуються такі умови:*
У1. МП $x(t)$, $t \geqslant 0$, є рівномірно ергодичним зі стаціонарним розподілом $\pi(B)$, $B \in \mathcal{E}$:

$$\pi(dx)q(x) = q\rho(dx), \qquad q = \int_E \pi(dx)q(x). \tag{15}$$

У2. Мартингал-різниці $\Delta\mu_N(\tau_k)$, $k \geqslant 1$, задовольняють умови центральної граничної теореми [3].

У3. Має місце збіжність початкових умов:

$$\zeta_N(0) \xrightarrow{P} \zeta(0), \qquad N \to \infty, \qquad E|\zeta_N(0)| \leqslant C < +\infty. \tag{16}$$

Тоді має місце збіжність скінченновимірних розподілів МСЕ:

$$\zeta_N(t) \xrightarrow{D} \zeta(t), \qquad N \to \infty, \qquad t \in [0, T]. \tag{17}$$

Граничний процес $\zeta(t)$, $t \geqslant 0$, задається генератором

$$\mathbb{L}\varphi(s) = -\widehat{V}\varphi'(s) + \frac{1}{2}\widehat{\sigma}^2\varphi''(s). \tag{18}$$

[1]Передбачається розширення МСЕ у неперервному марковському випадковому середовищі за схемою

$$\zeta_N(t) = \zeta_N(t_k), \qquad t_k = \frac{k}{N} \leqslant t < t_{k+1} = \frac{k+1}{N}, \qquad k \geqslant 0.$$

Параметри напрямної дії $\widehat{V}$ та дисперсії $\widehat{\sigma}^2$ обчислюються за формулами усереднення:

$$\widehat{V} = q\int\limits_E \rho(dx)V(x), \qquad \widehat{\sigma}^2 = q\int\limits_E \rho(dx)\sigma^2(x). \tag{19}$$

Зауваження 2. Граничний процес $\zeta(t)$, $t \geqslant 0$, з генератором (18), (19) є нормальним дифузійним процесом Орнштейна–Уленбека типу, що визначається стохастичним диференціальним рівнянням (11).

2. Доведення теорем 1 і 2. Основна ідея доведення граничних теорем 1 і 2 для МП полягає в застосуванні граничних теорем для випадкових процесів, зокрема для випадкових еволюцій, що базуються на збіжності генераторів, які породжують відповідні МП [4–7]. Збіжність породжуючих операторів (генераторів) на достатньо багатому класі числових функцій забезпечує збіжність скінченновимірних розподілів [4].

Наявність марковського випадкового середовища спричиняє необхідність використання розв'язків проблеми сингулярного збурення для звідно-зворотного оператора, що визначає рівномірно ергодичний ланцюг Маркова (або МП) [5].

Перш за все використовується характеризація розширеного МП з додатковою компонентою, що задає марковське випадкове середовище.

Лема 1. *A. Двокомпонентний ланцюг Маркова (у дискретному марковському випадковому середовищі)*

$$\zeta_N^d(t),\ x_N(t) := x_{Nt}, \qquad t = \frac{k}{N}, \qquad k \geqslant 0, \tag{20}$$

задається генератором

$$\mathbb{L}_N^d(x)\varphi(s,x) = N[\mathbb{C}_N^d(x)\mathbb{P} - \mathbb{I}]\varphi(s,x), \tag{21}$$

де оператори $\mathbb{P}$ і $\mathbb{C}$ визначаються відповідно як

$$\mathbb{P}\varphi(x) := \int\limits_E P(x,dy)\varphi(y); \tag{22}$$

$$\mathbb{C}_N^d(x)\varphi(s) := E\left[\varphi\left(s + \Delta\zeta_N^d\left(t + \frac{1}{N}\right)\right)\middle|\zeta_N^d(t) = s, x_N(t) = x\right], \qquad t \geqslant 0. \tag{23}$$

B. Двокомпонентний МП (неперервне марковське випадкове середовище)

$$\zeta_N^c(t),\ x_N(t) = x(Nt), \qquad t = \frac{k}{N}, \qquad t \geqslant 0, \tag{24}$$

задається генератором

$$\mathbb{L}_N^c(x)\varphi(s,x) = q(x)\mathbb{L}_N^d(x)\varphi(s,x). \tag{25}$$

Доведення леми 1. Твердження **A** випливає з означення генератора ланцюга Маркова у дискретно-неперервному часі з інтервалом $h = 1/N$ (див. [8, 2 : 1]).

Твердження **B** випливає з відомого означення генератора МП з неперервним часом з урахуванням перехідних імовірностей вкладеного ланцюга Маркова.

Істотний крок доведення теорем 1 і 2 реалізується у нижчеподаній лемі.

Лема 2. *Генератор* (21) *двокомпонентного ланцюга Маркова* (20) *допускає асимптотичний розклад на числових тест-функціях* $\varphi(s,x)$, *що мають три обмежені похідні за* s: $\varphi(s,\cdot) \in C^3(R)$:

$$\mathbb{L}_N^d(x)\varphi(s,x) = [\mathbb{Q}_N + \mathbb{C}^0(x)]\varphi(s,x) + \mathbb{R}_N(x)\varphi(s,x). \tag{26}$$

Тут за означенням

$$\mathbb{Q}_N\varphi(\cdot,x) := N[\mathbb{P} - \mathbb{I}]\varphi(\cdot,x), \tag{27}$$

$$\mathbb{C}^0(x)\varphi(s,\cdot) := -V(x)\varphi'(s,\cdot) + \frac{1}{2}\sigma^2\varphi''(s,\cdot) \tag{28}$$

та залишковий член

$$|\mathbb{R}_N(x)\varphi(s,x)| \to 0, \qquad N \to \infty, \qquad \varphi(s,x) \in C^3(R). \tag{29}$$

Доведення леми 2 базується на такому перетворенні генератора (21):

$$\mathbb{L}_N^d(x)\varphi(s,x) = [Q_N + C_N^0(x)]\varphi(s,x). \tag{30}$$

Тут за означенням

$$\mathbb{C}_N^0(x)\varphi(s,\cdot) := [\mathbb{C}_N(x) - \mathbb{I}]\varphi(s,\cdot). \tag{31}$$

Далі використовується формула Тейлора до тест-функцій $\varphi(s)$ до третього порядку включно:

$$\mathbb{C}_N^0(x)\varphi(s) = \frac{1}{N}\mathbb{C}^0(x)\varphi(s) + \frac{1}{N}\mathbb{R}_N(x)\varphi(s) \tag{32}$$

із залишковим членом, що задовольняє умову (29).

Представлення генератора (26) призводить до проблеми сингулярного збурення для зрізаного оператора (див. 5, гл. 5]):

$$\mathbb{L}_N^0\varphi(s,x) := [Q_N + C^0(x)]\varphi(s,x). \tag{33}$$

Лема 3. *На збурених тест-функціях*

$$\varphi_N(s,x) = \varphi(s) + \frac{1}{N}\varphi_1(s,x) \tag{34}$$

розв'язок проблеми сингулярного збурення для оператора (33) *визначається співвідношеннями*

$$\mathbb{L}_N^0\varphi_N(s,x) = \mathbb{L}\varphi(s) + \mathbb{R}_N(x)\varphi(s). \tag{35}$$

Тут граничний оператор (див. (28)*)*

$$\mathbb{L}\varphi(s) = -\widehat{V}(x)s\varphi'(s) + \frac{1}{2}\widehat{\sigma}^2(x)\varphi''(s) \tag{36}$$

та залишковий член $\mathbb{R}_N(x)$, *що задовольняє умову* (29).

Завершення доведення теореми 1 реалізується застосуванням модельної граничної теореми [5, 6 : 3], що забезпечує збіжність скінченновимірних розподілів (27).

Доведення теореми 2 завершується обчисленнями граничного оператора на збурених тест-функціях (34) при розв'язанні проблеми сингулярного збурення для оператора

$$q(x)\mathbb{L}_N^0\varphi_N(s,x) = \mathbb{L}\varphi(s) + \mathbb{R}_N(x)\varphi(s). \tag{37}$$

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.* Дифузійна апроксимація статистичних експериментів з наполегливою нелінійною регресією і еквілібріумом // Доп. НАН України. – 2014. – № 8. – С. 28–34.
3. *Borovskikh Yu. V., Korolyuk V. S.* Martingale approximation. – Utrecht: VSP, 1997. – 320 p.
4. *Скороход А. В.* Асимптотические методы теории стохастических дифференциальных уравнений. – Киев: Наук. думка, 1987. – 328 с.
5. *Korolyuk V. S., Limnios N.* Stochastic systems in merging phase space. – Singapore: World Scientific, 2005. – 331 p.
6. *Korolyuk V. S., Korolyuk V. V.* Stochastic models of systems. – Dordrecht: Kluwer, 1999. – 185 p.
7. *Ethier S. N., Kurtz T. G.* Markov processes: characterization and convergence. – New York: Wiley, 1986. – 536 p.
8. *Невельсон М. Б., Хасьминский Р. З.* Стохастическая аппроксимация и рекуррентное оценивание. – Москва: Наука, 1972. – 304 с.

Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ

Надійшло до редакції 26.11.2014

Д. В. Королюк

Статистические эксперименты с настойчивой линейной регрессией в марковской случайной среде

Статистические эксперименты (СЭ) с настойчивой линейной регрессией рассматриваются в дискретно-непрерывном времени $k = [Nt]$, $0 \leqslant t \leqslant T$. Направляющие параметры функции регрессии приращений зависят от состояний вложенной цепи Маркова в однородном (во времени) равномерно эргодическом марковском процессе, который описывает состояния случайной среды. СЭ задаются решениями разностных стохастических уравнений с двумя компонентами: предсказательной и стохастической (мартингал-разностью). Полученная аппроксимация в схеме серий с параметром серии N (объем выборки), при $N \to \infty$, является диффузионным процессом типа Орнштейна–Уленбека. Параметры смещения и диффузии определяются усреднением по стационарному распределению вложенной цепи Маркова.

D. V. Koroliouk

Statistical experiments with persistent linear regression in the Markov random medium

The statistical experiments (SE) with persistent non-linear regression are considered in the discrete-continuous time $k = [Nt]$, $0 \leqslant t \leqslant T$. The directing parameters of the regression function increments depend on the state of an embedded Markov chain in the (homogeneous in time) uniformly ergodic Markov process, which describes the states of the random medium. SE are defined by the solutions of stochastic difference equations with two components: predictive and stochastic (martingale-difference). The obtained approximation in the series scheme with series parameter N (size of the sample), as $N \to \infty$, is a diffusion Ornstein–Uhlenbeck-type process. The parameters of drift and diffusion are determined by averaging over the stationary distribution of the embedded Markov chain.