УДК 519.24

Д. В. Королюк

Диффузионная аппроксимация статистических экспериментов с настойчивой линейной регрессией и эквилибриумом

(Представлено академиком НАН Украины И. Н. Коваленко)

Предложена аппроксимация статистических экспериментов с настойчивой линейной 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.

В настоящей работе предлагается математическая модель статистических экспериментов (СЭ) с настойчивой линейной регрессией и эквилибриумом в дискретно-непрерывном времени. Диффузионная аппроксимация модели осуществляется процессом типа Орнштейна–Уленбека с непрерывным временем. Исходная модель цепи Маркова в дискретно-непрерывном времени возникает в результате масштабирования дискретного времени, а также параметров статистических экспериментов.

1. Постановка задачи. Бинарные повторяющиеся СЭ задаются значениями сумм выборки $\delta(k) = (\delta_r(k),\ 1 \leqslant r \leqslant N)$, $k \geqslant 0$ независимых и одинаково распределенных при каждом фиксированном k случайных величин $\delta_r(k)$, $1 \leqslant r \leqslant N$, принимающих два значения ± 1:

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

Настойчивая линейная регрессия означает, что имеет место соотношение

$$E[S_N(k+1)|S_N(k)] = C(S_N(k)), \qquad k \geqslant 0, \tag{2}$$

в котором функция линейной регрессии не зависит от объема выборки N и от номера $k \geqslant 0$, и задается соотношением

$$C(s) = (1-a)s + a\rho, \qquad |s| \leqslant 1. \tag{3}$$

Параметр ρ служит эквилибриумом функции регрессии

$$C(\rho) = \rho. \tag{4}$$

Направляющий параметр a удовлетворяет условию $0 < a < 1$. При $a = 0$ регрессия исчезает.

Задание СЭ с настойчивой линейной регрессией и эквилибриумом (1)–(3) означает, что вероятности выборочных значений определяются формулами

$$P\{\delta_r(k+1) = \pm 1 \mid S_N(k) = s\} = \frac{1}{2}[1 \pm C(s)]. \tag{5}$$

При этом параметры a и ρ могут быть заданы так, что условие (5) корректно определено.

Функция регрессии (2) преобразуется к следующему виду:

$$\begin{aligned}&C(s) = s + C_0(s) = \rho + C_\rho(s),\\ &C_0(s) = -a(s-\rho), \qquad C_\rho(s) = (1-a)(s-\rho).\end{aligned} \tag{6}$$

Так что дополнительные функции в (6) удовлетворяют условиям

$$C_0(\rho) = C_\rho(\rho) = 0. \tag{7}$$

Кроме того, задание бинарных СЭ (1)–(3) обеспечивает явный вид условной дисперсии

$$\begin{aligned}&D[S_N(k+1)|S_N(k)] = \frac{B(S_N(k))}{N}, \qquad k \geqslant 0,\\ &B(s) = 1 - C^2(s).\end{aligned} \tag{8}$$

2. Эквилибриум и аппроксимация СЭ нормальным процессом авторегрессии. В предыдущей работе [2] установлена сходимость к эквилибриуму

Теорема 1 (ср. [2, теорема 1]). *При сходимости начальных условий (с вероятностью 1)*

$$S_N(0) \xrightarrow{P1} \rho, \qquad N \to \infty,$$

имеет место сходимость СЭ (1)–(3) *(с вероятностью 1):*

$$S_N(k) \xrightarrow{P1} \rho, \qquad N \to \infty,$$

при каждом конечном $k > 0$.

Предложена также аппроксимация СЭ (1)–(3) дискретным нормальным процессом авторегрессии, основанная на следующей теореме.

Теорема 2 (ср. [2, теорема 2]). *При выполнении условия теоремы 1 имеет место сходимость по распределению*

$$\sqrt{N}[S_N(k+1) - C(S_N(k))] \xrightarrow{d} \sigma W(k+1), \qquad N \to \infty,$$

при каждом конечном $k \geqslant 0$.

Результат теоремы 2 служит основанием задания *процесса нормальной авторегрессии* $\widetilde{S_N}(k)$, $1 \leqslant r \leqslant N$, $k \geqslant 0$ в следующем виде:

Предложение 1 (дискретная аппроксимация). *Процесс нормальной авторегрессии в дискретном времени* $k \geqslant 0$ *задается соотношением*

$$\widetilde{S_N}(k+1) = C(\widetilde{S_N}(k)) + \frac{\sigma}{\sqrt{N}} W(k+1), \qquad k \geqslant 0, \tag{9}$$

в котором предельная дисперсия

$$\sigma^2 = 1 - \rho^2, \tag{10}$$

а $W(k+1)$, $k \geqslant 0$ *— нормально распределенные стандартные случайные величины, независимые при разных* $k \geqslant 1$.

Замечание 1. Процесс нормальной авторегрессии (6) сохраняет свойство настойчивой линейной регрессии

$$E[\widetilde{S_N}(k+1)|\widetilde{S_N}(k)] = C(\widetilde{S_N}(k)), \qquad k \geqslant 0,$$

с той же функцией регрессии (3), что и исходный СЭ (1).

Естественно рассматривать асимптотическое поведение нормированных флюктуаций СЭ относительно эквилибриума ρ:

$$\zeta_N(k) := \sqrt{N}[S_N(k) - \rho], \qquad k \geqslant 0. \tag{11}$$

Функция регрессии (3) для флюктуаций теперь имеет вид

$$E[\zeta_N(k+1)|\zeta_N(k) = s] = (1-a)s, \qquad k \geqslant 0. \tag{12}$$

В дальнейшем будет использована условная регрессия для приращений нормированных флюктуаций СЭ относительно эквилибриума ρ:

$$\Delta\zeta_N(k) := \zeta_N(k+1) - \zeta_N(k), \qquad k \geqslant 0. \tag{13}$$

При этом

$$E[\Delta\zeta_N(k)|\zeta_N(k) = s] = -as, \qquad k \geqslant 0. \tag{14}$$

Нам также понадобится разложение приращений нормированных флюктуаций в следующей форме:

$$\Delta\zeta_N(k) := \mu_N(k+1) - a\zeta_N(k), \qquad k \geqslant 0, \tag{15}$$

в которой мартингал-разность

$$\mu_N(k+1) := \Delta\zeta_N(k) + a\zeta_N(k), \qquad k \geqslant 0, \tag{16}$$

характеризуется следующими свойствами:

$$\begin{aligned} &E[\mu_N(k+1)|\zeta_N(k)] = 0, \qquad k \geqslant 0, \\ &E[\mu_N^2(k+1)|\zeta_N(k)] = B\left(\rho + \frac{\zeta_N(k)}{\sqrt{N}}\right), \qquad k \geqslant 0, \\ &B(s) := 1 - C^2(s). \end{aligned} \tag{17}$$

3. Диффузионная аппроксимация СЭ в дискретно-непрерывном времени. Диффузионная аппроксимация строится для нормированных флюктуаций (11) при масштабировании дискретного времени $k = [Nt]$, $t \geqslant 0$.

Предложение 2. *Марковский процесс с дискретно-непрерывным временем $\zeta_N^0(t)$, $t \geqslant 0$ задается следующим разностным стохастическим уравнением для приращений:*

$$\Delta\zeta_N^0(t) = -\frac{a\zeta_N^0(t)}{N} + \frac{\Delta\mu_N^0(t)}{\sqrt{N}}, \qquad 0 \leqslant t \leqslant T, \tag{18}$$

в котором мартингал-разности $\Delta\mu_N^0(t)$ *характеризуются свойствами*

$$\begin{aligned} &E[\Delta\mu_N^0(t)|\zeta_N^0(t)] = 0, \\ &E[(\Delta\mu_N^0(t))^2|\zeta_N^0(t)] = B\biggl(\rho + \frac{\zeta_N^0(t)}{\sqrt{N}}\biggr). \end{aligned} \tag{19}$$

Основной результат настоящей работы сформулирован в следующей теореме.

Теорема 3. *При сходимости (по вероятности) начальных условий*

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

имеет место сходимость конечномерных распределений процессов

$$\zeta_N^0(t) \xrightarrow{D} \zeta^0(t), \qquad N \to \infty, \tag{20}$$

к предельному диффузионному процессу $\zeta^0(t)$, $t \geqslant 0$, *типа Орнштейна–Уленбека, задаваемого производящим оператором (генератором)*

$$L^0\varphi(s) = -as\varphi'(s) + \frac{1}{2}\sigma^2\varphi''(s), \qquad \sigma^2 = 1 - \rho^2. \tag{21}$$

Аналогичный результат можно сформулировать для масштабированного процесса нормальной авторегрессии аппроксимирующего СЭ (9)–(10).

Предложение 3. *Процесс нормальной авторегрессии в дискретно-непрерывном времени задается приращениями*

$$\Delta\widetilde{\zeta_N^0}(t) = -\frac{a\widetilde{\zeta_N^0}(t)}{N} + \biggl(\frac{\sigma}{\sqrt{N}}\biggr)\Delta W(t), \qquad 0 \leqslant t \leqslant T. \tag{22}$$

Мартингальная составляющая в формуле (22) задается приращениями стандартного нормального процесса $W(t)$, $t \geqslant 0$:

$$\Delta W(t) := W(t+1) - W(t),$$

которые характеризуются двумя моментами:

$$E\Delta W(t) = 0, \qquad E[\Delta W(t)]^2 = 1. \tag{23}$$

Теорема 4. *При сходимости (по вероятности) начальных условий*

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

имеет место сходимость конечномерных распределений процесса нормальной авторегрессии, задаваемого соотношениями (22)–(23)*:*

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

к диффузионному процессу $\zeta^0(t)$ *типа Орнштейна–Уленбека, задаваемого генератором* (21), *причем диффузионный процесс* $\zeta^0(t)$, $t \geqslant 0$ *является решением стохастического дифференциального уравнения*

$$d\zeta^0(t) = -a\zeta^0(t)dt + \sigma dW(t). \tag{25}$$

4. Обоснование диффузионной аппроксимации. Основная идея доказательства предельных теорем для марковских случайных процессов состоит в применении операторной характеризации марковского процесса (генератором) на классе числовых функций с аргументом во множестве значений марковского процесса. Сходимость производящих операторов на достаточно богатом классе числовых функций обеспечивает сходимость конечномерных распределений процессов [3, 4].

Следуя монографии [4] (см. также [5]), введем генератор марковских процессов в схеме серий

$$L_N\varphi(s) = NE[\varphi(s + \Delta\zeta_N^0(t)) - \varphi(s)|\zeta_N^0(t) = s]. \tag{26}$$

Существенный этап доказательства теоремы 3 состоит в применении теоремы 1 А. В. Скорохода [4; 2 : 1], из которой следует сходимость (20) конечномерных распределений флюктуаций СЭ относительно эквилибриума.

Существенный этап доказательства теоремы 3 содержится в следующей лемме.

Лемма 1. *Имеет место сходимость генераторов* (26)*:*

$$\lim_{N\to\infty} L_N\varphi(s) = L^0\varphi(s), \qquad \varphi(s) \in C^3(\mathbb{R}) \tag{27}$$

на классе $C^3(\mathbb{R})$ *числовых финитных функций, трижды непрерывно дифференцируемых с ограниченными производными. Предельный генератор*

$$L^0\varphi(s) = -as\varphi'(s) + \frac{1}{2}\sigma^2\varphi''(s), \qquad \sigma^2 = 1 - \rho^2 \tag{28}$$

задает предельный процесс типа Орнштейна–Уленбека (25)*.*

Доказательство. Используя представление приращений (18) марковского процесса флюктуаций $\zeta_N^0(t)$, $t \geqslant 0$, вычислим первые два момента приращений.

С учетом (19) находим

$$E[\Delta\zeta_N^0(t)|\zeta_N^0(t) = s] = -\frac{as}{N}, \tag{29}$$

$$E[(\Delta\zeta_N^0(t))^2|\zeta_N^0(t) = s] = \frac{B(\rho)}{N} + O\left(\frac{1}{N^{3/2}}\right). \tag{30}$$

Здесь (см. формулу (8), а также (3))

$$B(\rho) = 1 - C^2(\rho) = 1 - \rho^2 =: \sigma^2. \tag{31}$$

Теперь применим формулу Тейлора в представлении (26) генератора L_N к тест-функции $\varphi(s) \in C^3(\mathbb{R})$:

$$L_N\varphi(s) = N\left[E[\Delta\zeta_N^0(t)|\zeta_N^0(t) = s]\varphi'(s) + E[(\Delta\zeta_N^0(t))^2|\zeta_N^0(t) = s]\frac{1}{2}\varphi''(s) + R_N\varphi(s)\right]. \tag{32}$$

Здесь остаточный член по условию теоремы 3 имеет оценку

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

Используя представления (29), (30) первых двух моментов приращений, получаем

$$L_N\varphi(s) = L^0\varphi(s) + NR_N\phi(s), \tag{33}$$

в котором остаточный член

$$NR_N\varphi(s) \to 0, \qquad N \to \infty, \qquad \varphi(s) \in C^3(\mathbb{R}). \tag{34}$$

Из представлений (33), (34) следует утверждение леммы 1.

Аналогично доказывается теорема 4. Введем генератор приращений (22) марковского процесса $\widetilde{\zeta_N^0}(t),\ t \geqslant 0$:

$$\widetilde{L_N}\varphi(s) = N[E[\varphi(s + \Delta\widetilde{\zeta_N^0}(t)) - \varphi(s)]|\widetilde{\zeta_N^0}(t) = s]. \tag{35}$$

Вычисляются первые два момента приращений с учетом (23):

$$\begin{aligned} &E[\Delta\widetilde{\zeta_N^0}(t)|\widetilde{\zeta_N^0}(t) = s] = -\frac{as}{N}, \\ &E[(\Delta\widetilde{\zeta_N^0}(t))^2|\widetilde{\zeta_N^0}(t) = s] = \sigma^2 + O\left(\frac{1}{N^{3/2}}\right). \end{aligned} \tag{36}$$

Теперь аналогично доказывается следующая лемма.

Лемма 2. *Имеет место сходимость генераторов* (35), (36)

$$\lim_{N\to\infty} \widetilde{L_N}\varphi(s) = L^0\varphi(s), \qquad \varphi(s) \in C^3(\mathbb{R}), \tag{37}$$

на классе числовых финитных функций $C^3(\mathbb{R})$ *трижды непрерывно дифференцированных с ограниченными производными. Предельный генератор* L^0 *задается представлением* (28).

Утверждение теоремы 4 следует из леммы 2 с применением теоремы 9 [6, с. 415].

Замечание 2. Утверждение теоремы 4 можно получить в виде следствия теоремы 8 [6, с. 406], используя следующее представление решения разностного стохастического уравнения (22):

$$\widetilde{\zeta_N^0}(t) = \widetilde{\zeta_N^0}(0) - a\int\limits_0^{[Nt]/N} \widetilde{\zeta_N^0}(\tau)d\tau + \frac{\sigma}{\sqrt{N}}W([Nt]). \tag{38}$$

При этом следует проверить выполнимость условий теоремы 8 [6, с.406] для функций

$$\begin{aligned} &a_N(t,s) = NEf_N(t,s;W), \\ &B(t,s) = NEf_N^2(t,s;W), \\ &f_N(t,s;W) := -\frac{as}{N} + \frac{\sigma}{\sqrt{N}}W. \end{aligned} \tag{39}$$

Здесь W — нормально распределенная стандартная случайная величина ($EW = 0$, $EW^2 = 1$).

Приведенные в теоремах 3 и 4 аппроксимации флюктуаций СЭ могут быть использованы в статистическом анализе СЭ с применением методов математической статистики.

1. *Королюк Д. В.* Рекуррентные статистические эксперименты с настойчивой линейной регрессией // Укр. мат. вестн. – 2012. – **9**, № 4. – С. 560–567.
2. *Королюк Д. В.* Бинарные повторяющиеся статистические эксперименты с настойчивой линейной регрессией // Там же. – 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 space. – Singapore; London: World Scientific, 2005. – 331 p.
6. *Гихман И. И., Скороход А. В.* Стохастические дифференциальные уравнения и их приложения. – Киев: Наук. думка, 1982. – 611 с.

Институт телекоммуникаций и глобального информационного пространства НАН Украины, Киев

Поступило в редакцию 17.10.2013

Д. В. Королюк

Дифузійна апроксимація статистичних експериментів з наполегливою лінійною регресією та еквілібріумом

Запропоновано апроксимацію статистичних експериментів з наполегливою лінійною регресією та 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.*

D. V. Koroliouk

Diffusion approximation of statistical experiments with persistent linear regression and equilibrium

We propose an approximation of statistical experiments with persistent 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.*