УДК 519.217.2

К.С. МАТЕЦКИЙ, Ю.С. ХАРИН

О РИСКЕ ПРОГНОЗИРОВАНИЯ ОДНОРОДНЫХ КОНЕЧНЫХ ЦЕПЕЙ МАРКОВА С НЕИЗВЕСТНЫМИ ПАРАМЕТРАМИ

The problem of homogenous finite Markov chains forecasting with unknown parameters is considered. Two asymptotically optimal «plug-in» forecasting statistics are given. For a stationary Markov chain the upper bound of forecasting risk is got. Numerical experiments are conducted to illustrate the theoretical results.

Дискретные временные ряды $x_t \in A$, $t \in \mathbf{N}$, где пространство состояний $A = \{0,1,...,N-1\}$ – конечное множество мощности $2 \le N < +\infty$, широко применяются при математическом моделировании сложных систем и процессов в экономике, технике, медицине, социологии, генетике и других приложениях [1–4]. Известной моделью таких дискретных временных рядов является цепь Маркова. Заметим, что сложная цепь Маркова s-го порядка сводится к рассматриваемой здесь простой цепи Маркова расширением ее пространства состояний. На практике часто возникают задачи прогнозирования будущих значений цепи Маркова по наблюдаемой реализации и оценивания риска прогнозирования [3, 4]. Случай, когда параметры модели известны, является изученным [3]. В приложениях параметры обычно не известны, поэтому актуальной является проблема исследования риска прогнозирования цепи Маркова по оцененным параметрам модели.

Постановка задачи

Рассмотрим ситуацию, когда подлежащий прогнозированию временной ряд x_t является стационарной цепью Маркова с конечным пространством состояний A, вектором стационарных распределений $\pi = (\pi_0, \pi_1, ..., \pi_{N-1})' \in [0,1]^N$ и матрицей вероятностей одношаговых переходов $P = (p_{ij}) \in [0,1]^{N\times N}$:

$$\mathrm{P}\{x_1 = i\} = \pi_i, \ \ \mathrm{P}\{x_{t+1} = j \mid x_t = i\} = p_{ij}, \ \ i, j \in A; \ \ \sum_{i\in A} \pi_i = 1; \ \ \sum_{j\in A} p_{ij} = 1, \ \ i \in A.$$

Временной ряд наблюдается в течение $T \geq 1$ последовательных единиц времени и зарегистрирована реализация $X_1^T = (x_1, x_2, ..., x_T) \in A^T$. Задача заключается в построении прогнозирующей статистики будущего значения $x_{T+\tau} \in A$ на $\tau \geq 1$ шагов вперед и оценивании риска прогнозирования.

Алгоритмы оптимального и асимптотически оптимального прогнозирования

Теорема 1. *Если x_t – стационарная цепь Маркова с известными параметрами π, P, то оптимальная по критерию минимума вероятности ошибки* (*риска для* (0–1)-*функции потерь*) *прогнозирующая статистика имеет вид*

$$\hat{x}_{T+\tau} = \arg\max_{j\in A} p^{(\tau)}_{x_T, j}, \tag{1}$$

где $p^{(\tau)}_{ij}$ – (i, j)-*й элемент матрицы* P^τ. *При этом достигается минимум риска:*

$$r_0(\tau) = \mathrm{P}\{\hat{x}_{T+\tau} \neq x_{T+\tau}\} = 1 - \sum_{i\in A} \pi_i \max_{j\in A} p^{(\tau)}_{ij}. \tag{2}$$

Доказательство теоремы 1 приведено в [3].

Если матрица вероятностей переходов P априорно не известна, то строится ее оценка максимального правдоподобия $\hat{P} = (\hat{p}_{ij})$ по наблюдаемой реализации X_1^T :

$$\hat{p}_{ij} = \begin{cases} \sum_{t=1}^{T-1} \delta_{x_t, i}\delta_{x_{t+1}, j} \Big/ \sum_{t=1}^{T-1} \delta_{x_t, i}, \ \text{если} \ \sum_{t=1}^{T-1} \delta_{x_t, i} > 0, \\ 1/N, \ \text{если} \ \sum_{t=1}^{T-1} \delta_{x_t, i} = 0, \end{cases} \tag{3}$$

где $\delta_{ij} = \{1, \ i = j; \ 0, \ i \neq j\}, \ \ i, j \in A,$ – символ Кронекера. Тогда «подстановочная» прогнозирующая статистика получается из (1):

$$\hat{x}_{T+\tau} = \arg\max_{j\in A} \left(\hat{P}^\tau\right)_{x_T, j}. \tag{4}$$

Поскольку в (1) используется условная вероятность перехода из x_T в j за τ шагов $p^{(\tau)}_{x_T, j}$, то оказывается возможным оценить эту вероятность частотным методом:

$$\tilde{p}_{x_T, j}(\tau) = \begin{cases} \sum_{t=1}^{T-\tau} \delta_{x_t, x_T}\delta_{x_{t+\tau}, j} \Big/ \sum_{t=1}^{T-\tau} \delta_{x_t, x_T}, \ \text{если} \ \sum_{t=1}^{T-\tau} \delta_{x_t, x_T} > 0, \\ 1/N, \ \text{если} \ \sum_{t=1}^{T-\tau} \delta_{x_t, x_T} = 0. \end{cases} \tag{5}$$

Получаем другой вариант «подстановочной» прогнозирующей статистики:

$$\tilde{x}_{T+\tau} = \arg\max_{j\in A} \tilde{p}_{x_T, j}(\tau). \tag{6}$$

Так как оценки (3), (5) сильно состоятельны при $T \to +\infty$ [3], то риски обеих прогнозирующих статистик (4), (6) сходятся к риску $r_0(\tau)$, определяемому (2).

Риск прогнозирования в случае неизвестной матрицы вероятностей одношаговых переходов

Будем рассматривать случай $\tau = 1$. Обозначим: $n_i = \sum_{t=1}^{T-1} \delta_{x_t, i}$, $i \in A$, – частота встречаемости значения i в реализации X_1^{T-1}; $I_N = (\delta_{ij})$ – единичная матрица порядка N; E_N – $(N\times N)$-матрица, все элементы которой равны 1; $I\{A\}$ – индикатор случайного события A.

Лемма 1. *Пусть* Π – $(N\times N)$-*матрица, все строки которой одинаковы и равны* π', $Z = (z_{ij}) = (I_N - P + \Pi)^{-1}$ – *фундаментальная матрица* [5]. *Тогда при* $T \to +\infty$ *справедливы асимптотические разложения:*

$$\mathrm{E}\{n_j \mid x_T = i\} = T\pi_j - \pi_j + \pi_j(z_{ji} - \delta_{ij}) / \pi_i + O(\rho^T), \ \ \rho \in (0,1), \tag{7}$$

$$\mathrm{Cov}\{n_j, n_k \mid x_T = i\} = T(\pi_j z_{jk} + \pi_k z_{kj} - \pi_k \delta_{jk} - \pi_j \pi_k) + O(1), \tag{8}$$

$$\mathrm{D}\{n_j \mid x_T = i\} = T\pi_j (2z_{jj} - \pi_j - 1) + O(1), \ \textit{где}\ i, j, k \in A. \tag{9}$$

Доказательство. Воспользуемся известными асимптотическими разложениями [5]:

$$\sum_{t=1}^{T} P^t = T\Pi + Z - I_N + O(\rho^{T+1}) E_N, \ \ \rho \in (0,1), \ \ \sum_{t=1}^{T} t P^t = \frac{T(T+1)}{2}\Pi + O(1) E_N.$$

При помощи эквивалентных преобразований и марковского свойства получаем

$$\mathrm{E}\{n_j \mid x_T = i\} = \sum_{t=1}^{T-1} \mathrm{E}\{\delta_{x_t, j} \mid x_T = i\} = \frac{\pi_j}{\pi_i}\sum_{t=1}^{T-1} p_{ji}^{(t)} = T\pi_j - \pi_j + \frac{\pi_j}{\pi_i}(z_{ji} - \delta_{ij}) + O(\rho^T), \ \ \rho \in (0,1).$$

$$\mathrm{E}\{n_j n_k \mid x_T = i\} = \sum_{t,r=1}^{T-1} \mathrm{E}\{\delta_{x_t, j}\delta_{x_r, k} \mid x_T = i\} = \delta_{jk}\frac{\pi_j}{\pi_i}\sum_{t=1}^{T-1} p_{ji}^{(t)} + \frac{\pi_j}{\pi_i}\sum_{1 \le t < r \le T-1} p_{jk}^{(r-t)} p_{ki}^{(T-r)} + \frac{\pi_k}{\pi_i}\sum_{1 \le r < t \le T-1} p_{kj}^{(t-r)} p_{ji}^{(T-t)}.$$

Поскольку выполняется асимптотическое разложение

$$\sum_{1 \le t < r \le T-1} p_{jk}^{(r-t)} p_{ki}^{(T-r)} = T^2 \pi_k \pi_i / 2 + T(-3\pi_k \pi_i / 2 + \pi_k z_{ki} + \pi_i z_{jk} - \pi_k \delta_{ki} - \pi_i \delta_{jk}) + O(1),$$

получаем

$$\mathrm{E}\{n_j n_k \mid x_T = i\} = T^2 \pi_k \pi_j + T\Big(-3\pi_k \pi_j + \frac{\pi_k}{\pi_i}(\pi_j z_{ji} + \pi_i z_{kj}) + \frac{\pi_j}{\pi_i}(\pi_k z_{ki} + \pi_i z_{jk}) - \\ - \pi_k \delta_{jk} - \pi_k \delta_{ij} - \pi_j \delta_{ki}\Big) + O(1).$$

Используя асимптотическое разложение (7), находим

$$\mathrm{Cov}\{n_j, n_k \mid x_T = i\} = \mathrm{E}\{n_j n_k \mid x_T = i\} - \mathrm{E}\{n_j \mid x_T = i\}\mathrm{E}\{n_k \mid x_T = i\} = \\ = T(\pi_j z_{jk} + \pi_k z_{kj} - \pi_k \delta_{jk} - \pi_j \pi_k) + O(1).$$

Разложение (9) получается из (8) при подстановке $k = j$. ■

Лемма 2. *Если* $\pi_j \ne \pi_k$, $j \ne k$, $j, k \in A$, *то справедлива асимптотическая оценка:*

$$\left| I\{\pi_j > \pi_k\} - \mathrm{P}\{n_j > n_k \mid x_T = i\} \right| \le \frac{\nu_{jk}}{T} + O\left(\frac{1}{T^2}\right), \tag{10}$$

где $\nu_{jk} = (2\pi_j (z_{jj} - z_{jk}) + 2\pi_k (z_{kk} - z_{kj}) - (\pi_j + \pi_k) - (\pi_j - \pi_k)^2) / (\pi_j - \pi_k)^2$.

Доказательство. Из леммы 1 следует, что $\mathrm{E}\{n_j \mid x_T = i\} = (T-1)\pi_j + \gamma_{ji}$, где $\gamma_{ji} = \pi_j (z_{ji} - \delta_{ij}) / \pi_i + O(\rho^T)$. Если $\pi_j < \pi_k$, то при достаточно больших T справедлива оценка:

$$\mathrm{P}\{n_j > n_k \mid x_T = i\} = \mathrm{P}\left\{ \frac{n_j}{T-1} - \pi_j - \frac{\gamma_{ji}}{T-1} - \frac{n_k}{T-1} + \pi_k + \frac{\gamma_{ki}}{T-1} > \pi_k - \pi_j + \frac{\gamma_{ki} - \gamma_{ji}}{T-1} \right| x_T = i \Bigg\} \le$$

$$\le \mathrm{P}\left\{ \left| \frac{n_j}{T-1} - \pi_j - \frac{\gamma_{ji}}{T-1} - \frac{n_k}{T-1} + \pi_k + \frac{\gamma_{ki}}{T-1} \right| > \pi_k - \pi_j + \frac{\gamma_{ki} - \gamma_{ji}}{T-1} \right| x_T = i \Bigg\},$$

поскольку при достаточно больших T $\pi_k - \pi_j + \dfrac{\gamma_{ki} - \gamma_{ji}}{T-1} > 0$. Применяя неравенство Чебышева и лемму 1, получаем

$$\mathrm{P}\{n_j > n_k \mid x_T = i\} \le \frac{\mathrm{D}\{(n_j - n_k) / (T-1) \mid x_T = i\}}{(\pi_k - \pi_j + (\gamma_{ki} - \gamma_{ji}) / (T-1))^2} = \frac{\nu_{jk}}{T} + O\left(\frac{1}{T^2}\right),$$

так как $\mathrm{D}\{n_j - n_k \mid x_T = i\} = \mathrm{D}\{n_j \mid x_T = i\} + \mathrm{D}\{n_k \mid x_T = i\} - 2\mathrm{Cov}\{n_j, n_k \mid x_T = i\}$.

В случае, когда $\pi_j > \pi_k$, строятся аналогичные оценки для вероятности $1 - \mathrm{P}\{n_j > n_k \mid x_T = i\} = \mathrm{P}\{n_k \ge n_j \mid x_T = i\}$. ■

По наблюдаемой цепи Маркова x_t построим новый дискретный временной ряд $y_t = (x_t, x_{t+1})$, $t \ge 1$; y_t является стационарной цепью Маркова с матрицей вероятностей одношаговых переходов и вектором стационарных распределений:

$$\bar{P} = (\bar{p}_{(ij)(kl)}), \ \ \bar{\pi} = (\bar{\pi}_{(ij)}); \ \ \bar{p}_{(ij)(kl)} = \delta_{jk} p_{kl}, \ \ \bar{\pi}_{(ij)} = \pi_i p_{ij}, \ \ i, j, k, l \in A. \tag{11}$$

Теорема 2. *Если в каждой i-й строке матрицы P существует единственный максимальный элемент с номером* $i^* \in A$, *то справедлива верхняя асимптотическая оценка риска прогнозирования:*

$$\tilde{r}(T) \le r_0 + \left(2^{N-1} - 1\right) \sum_{i \in A} \pi_i p_{ii^*} \max_{k \in A \setminus \{i^*\}} \{\nu_{(ik)(ii^*)}\} \frac{1}{T} + O\left(\frac{1}{T^2}\right), \tag{12}$$

где $\{\nu_{(ik)(ii^*)}\}$ *– величины из леммы 2 для цепи Маркова* y_t.

Доказательство. Для любого $i \in A$ при достаточно большом T верны неравенства

$$1 - \mathrm{P}\{n_{ii^*} > n_{ik}, k \in A \setminus \{i^*\} \mid y_T = (i, i^*)\} = \sum_{l \in A \setminus \{i^*\}} \mathrm{P}\{n_{ii^*} \le n_{il}, n_{ii^*} > n_{ik}, k \in A \setminus \{i^*, l\} \mid y_T = (i, i^*)\} +$$

$$+ \sum_{l, t \in A \setminus \{i^*\}} \mathrm{P}\{n_{ii^*} \le n_{il}, n_{ii^*} \le n_{it}, n_{ii^*} > n_{ik}, k \in A \setminus \{i^*, l, t\} \mid y_T = (i, i^*)\} + \ldots + \mathrm{P}\{n_{ii^*} \le n_{ik}, k \in A \setminus \{i^*\} \mid y_T = (i, i^*)\} \le$$

$$\le \left(C_{N-1}^1 + C_{N-1}^2 + \ldots + C_{N-1}^{N-1}\right) \max_{k \in A \setminus \{i^*\}} \{\mathrm{P}\{n_{ii^*} \le n_{ik} \mid y_T = (i, i^*)\}\} \le$$

$$\le \left(2^{N-1} - 1\right) \max_{k \in A \setminus \{i^*\}} \{\nu_{(ik)(ii^*)}\} \frac{1}{T} + O\left(\frac{1}{T^2}\right).$$

Тогда для риска прогнозирования стационарной цепи Маркова справедлива оценка:

$$\tilde{r}(T) = \mathrm{P}\{\hat{x}_{T+1} \ne x_{T+1}\} = 1 - \sum_{i, j \in A} \pi_i p_{ij} \mathrm{P}\{j = \arg\max_{k \in A} \{\hat{p}_{ik}\} \mid y_T = (i, j)\} \le$$

$$\le 1 - \sum_{i \in A} \pi_i p_{ii^*} \mathrm{P}\{n_{ii^*} > n_{ik}, k \in A \setminus \{i^*\} \mid y_T = (i, i^*)\} \le$$

$$\le r_0 + \left(2^{N-1} - 1\right) \sum_{i \in A} \pi_i p_{ii^*} \max_{k \in A \setminus \{i^*\}} \{\nu_{(ik)(ii^*)}\} \frac{1}{T} + O\left(\frac{1}{T^2}\right). \ \blacksquare$$

Следствие 1. Если $N = 2$ *и в каждой строке матрицы P элементы не равны между собой, то справедлива асимптотическая оценка*:

$$\tilde{r}(T) \le r_0 + \sum_{i \in A} \pi_i p_{ii^*} \nu_{(i, 1-i^*)(ii^*)} \frac{1}{T} + O\left(\frac{1}{T^2}\right).$$

Заметим, что в матрице $\bar{P}$ (11) количество нулевых элементов равно $N^3(N-1)$, т. е. их доля составляет $1 - 1/N$ от количества всех элементов. Учитывая это, построим эффективный алгоритм для нахождения фундаментальной матрицы цепи Маркова y_t.

Лемма 3. *Пусть* $\bar{Z} = (\bar{z}_{(ij)(kl)})$ *– фундаментальная матрица цепи Маркова* y_t *с параметрами* (11), *тогда ее элементы связаны с элементами фундаментальной матрицы* $Z = (z_{ij})$ *стационарной цепи Маркова* x_t *соотношением*

$$\bar{z}_{(ij)(kl)} = \delta_{ik}\delta_{jk} + (z_{jk} - \pi_k) p_{kl}, \quad i, j, k, l \in A. \tag{13}$$

Доказательство. Пусть P_i, $i \in A$, – ($N \times N$)-матрица, i-я строка которой совпадает с i-й строкой матрицы P, а остальные элементы равны 0, тогда $P = \sum_{i \in A} P_i$ и

$$\bar{P} = \begin{pmatrix} P_0 & P_1 & \ldots & P_{N-1} \\ \ldots & \ldots & \ldots & \ldots \\ P_0 & P_1 & \ldots & P_{N-1} \end{pmatrix} = \begin{pmatrix} I_N \\ \ldots \\ I_N \end{pmatrix} \left(P_0 \quad P_1 \quad \ldots \quad P_{N-1}\right), \ \bar{P}^n = \begin{pmatrix} I_N \\ \ldots \\ I_N \end{pmatrix} P^{n-1} \left(P_0 \quad P_1 \quad \ldots \quad P_{N-1}\right),$$

где $n \ge 1$. Обозначим $\bar{\Pi}$ – ($N^2 \times N^2$)-матрица, все строки которой равны $\bar{\pi}'$, тогда

$$\bar{\Pi} = \begin{pmatrix} I_N \\ \ldots \\ I_N \end{pmatrix} \Pi \left(P_0 \quad P_1 \quad \ldots \quad P_{N-1}\right).$$

Используя представление фундаментальной матрицы из [5], получаем

$$\bar{Z} = I_{N^2} + \sum_{n=1}^{+\infty} \left(\bar{P}^n - \bar{\Pi}\right) = I_{N^2} + \begin{pmatrix} I_N \\ \ldots \\ I_N \end{pmatrix} \sum_{n=1}^{+\infty} \left(P^{n-1} - \Pi\right) \left(P_0 \quad P_1 \quad \ldots \quad P_{N-1}\right) = I_{N^2} + \begin{pmatrix} I_N \\ \ldots \\ I_N \end{pmatrix} Z \left(P_0 \quad P_1 \quad \ldots \quad P_{N-1}\right) - \bar{\Pi}.$$

Записав последнее равенство поэлементно, получаем (13). ■

Следствие 2. Элементы $\nu_{(ik)(ii^*)}$ *из* (12) *равны*

$$\nu_{(ik)(ii^*)} = \left(2\left(p_{ii^*} - p_{ik}\right)\left(p_{ii^*} z_{i^*_i i} - p_{ik} z_{kl}\right) + p_{ii^*} + p_{ik}\right) / \left(\pi_i \left(p_{ii^*} - p_{ik}\right)^2\right) - 3.$$

Численные эксперименты

В ходе численных экспериментов для различных значений параметров N, P, π моделировались цепи Маркова, затем методом Монте-Карло оценивался риск прогнозирования их будущих значений при горизонте прогнозирования $\tau = 1$, вычислялись оценки $r_+(T) = r_0 + \dfrac{A}{T}$ (где константа A определена в правой части (12)) и вычислялись значения статистики $G(T) = T(\tilde{r}(T) - r_0)$. Число прогонов в методе Монте-Карло K=10^5.

Эксперимент 1. Значения параметров: N=2,

$$P = \begin{pmatrix} 0,1 & 0,9 \\ 0,9 & 0,1 \end{pmatrix},\ \ \pi = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix}.$$

Результаты численного эксперимента приведены на рис. 1 и 2.

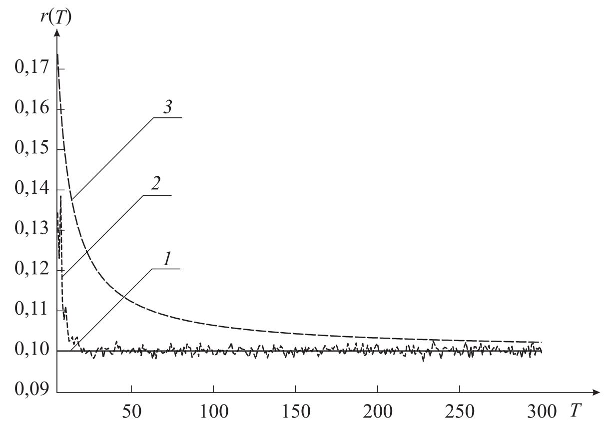

Рис. 1. Графики оценок риска: *1* – r, *2* – r_0, *3* – r_+

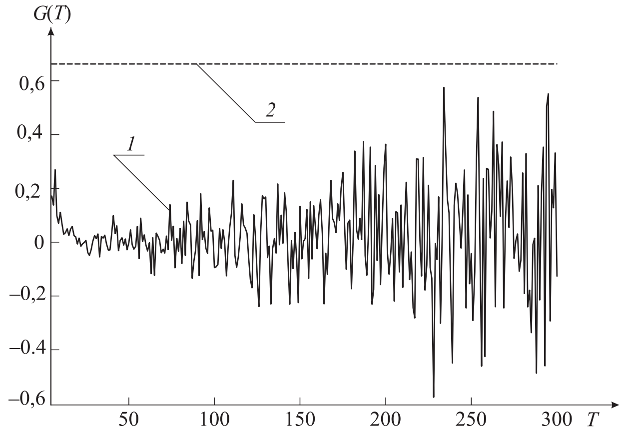

Рис. 2. Графики статистики G (*1*) и константы A (*2*)

Эксперимент 2. Значения параметров: N=3,

$$P = \begin{pmatrix} 0,1 & 0,8 & 0,1 \\ 0,1 & 0,1 & 0,8 \\ 0,8 & 0,1 & 0,1 \end{pmatrix},\ \ \pi = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}.$$

Результаты численного эксперимента даны на рис. 3.

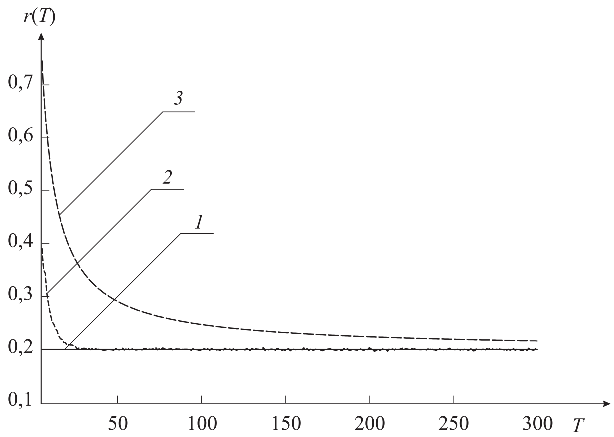

Рис. 3. Графики оценок риска прогнозирования: *1* – r, *2* – r_0, 3 – r_+

Из полученных результатов видно, что точность главного члена оценки (12) существенно зависит от матрицы P. Эта точность ниже, если матрица вероятностей переходов P имеет доминирующие диагональные элементы по сравнению со случаем, когда диагональные элементы не являются доминирующими.

1. Basawa I. V., Rao B. L. C. Statistical interference for stochastic processes. New York, 1980.
2. Дуб Дж. Вероятностные процессы. М., 1956.
3. Харин Ю. С. Оптимальность и робастность в статистическом прогнозировании. Мн., 2008.
4. Харин Ю. С. Вероятностно-статистический анализ цепей Маркова высокого порядка // Вестн. БГУ. Сер. 1. 2006. № 3. С. 80.
5. Кемени Дж., Снелл Дж. Конечные цепи Маркова. М., 1970.

Поступила в редакцию 09.06.10.

Константин Станиславович Матецкий – студент 5-го курса факультета прикладной математики и информатики.

Юрий Семенович Харин – член-корреспондент НАН Беларуси, доктор физико-математических наук, профессор, заведующий кафедрой математического моделирования и анализа данных.