© 2012 г. Г. А. Тимофеева д-р физ.-мат. наук,
Н. А. Тимофеев
(Уральский государственный университет путей сообщения,
Екатеринбург)

ПРОГНОЗИРОВАНИЕ СОСТАВЛЯЮЩИХ КРЕДИТНОГО ПОРТФЕЛЯ НА ОСНОВЕ МОДЕЛИ МАРКОВСКОЙ ЦЕПИ[1]

Рассматривается задача прогнозирования составляющих кредитного портфеля банка, в частности доли проблемных кредитов. Предполагается, что изменение портфеля описывается марковским случайным процессом с дискретным временем и конечным числом состояний. Под состоянием кредита понимается принадлежность его той или иной группе кредитов по наличию и срокам задолженности по платежам. Предполагается, что матрица переходных вероятностей известна неточно и информация о ней поступает по мере функционирования системы.

1. Введение

При описании динамики значительного числа физических, технических и экономических систем используются модели марковских случайных процессов [1]. Одной из наиболее простых динамических моделей этого направления является марковская цепь с дискретным временем и конечным числом состояний. Несмотря на простоту модели, она активно используется при оценке индексов рынка [2, 3], моделировании динамики кредитного портфеля [4,5] и исследовании устойчивости ряда технических систем. Основная проблема, возникающая при моделировании реальной системы с помощью марковских цепей, – это оценивание элементов матрицы переходных вероятностей на основе статистических данных [6]. От точности оценки переходных вероятностей зависит точность прогноза вероятностей состояний системы.

Анализ кредитного портфеля представляет собой систематическое изучение и наблюдение за кредитной деятельностью банка, которое позволяет оценить состав и качество банковских ссуд в динамике. Управление кредитным портфелем построено на системе показателей деятельности банка в области кредитования клиентов. Основными показателями кредитного портфеля являются его доходность и доля

[1]Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 10-01-00672а).

проблемных кредитов. Кредит называется проблемным, если задержка платежей по договору превысила заданное критическое значение (например, 90 дней).

Каждый банк строит анализ показателей на базе своего опыта и аналитических возможностей, используя при этом инструментарий, который накоплен в отечественной и мировой банковской практике [7].

Математическая модель описания динамики портфеля кредитов основана на описании изменений состояния отдельно взятого, "случайного" кредита. Под состоянием кредита понимается его принадлежность той или иной группе кредитов по наличию и сроку задолженности по выплатам. Предполагается, что поведение кредита (с точки зрения выполнения обязательств по платежам) не зависит от предыстории и происходит случайным образом, т. е. процесс переходов из состояния в состояние можно описывать в рамках теории марковских случайных процессов.

В данной статье используется модель с дискретным временем и состояния кредитов фиксируются через одинаковые промежутки времени, например один раз в месяц. Изменение составляющих кредитного портфеля исследуется на временно́м интервале относительно стабильного состояния банковской системы. Предполагается, что вероятности переходов из состояния в состояние существенно не изменяются на данном временно́м интервале, поэтому используется модель с постоянными вероятностями переходов.

Предполагается, что переходные вероятности заранее не известны и оцениваются в ходе функционирования системы. Цель анализа – прогнозирование возможных убытков от кредитования и доли проблемных кредитов через заданный промежуток времени. Отметим, что эти функции линейно зависят от вектора состояния системы. В качестве критерия выбран квантильный критерий [8], т. е. оценивается значение, которое целевая функция не превышает с заданной вероятностью.

Для решения предлагается использовать метод оценивания матрицы переходных вероятностей, основанный на построении доверительных множеств для элементов матрицы, и имитационное моделирование. Доверительный подход позволяет использовать для оценки вектора состояний системы современную технику информационных множеств [9–12]. С другой стороны, имитационное моделирование, основанное на аналитической модели, в сочетании с методами оценивания квантили целевой функции [13] дает в данной задаче более точные результаты.

2. Математическая постановка задачи

При анализе динамики доли проблемных кредитов различными авторами используются разные классификации кредитов по срокам задержки платежей и уровню восстановления [4]. Будем рассматривать кредитный портфель, состоящий из кредитов k различных групп. Предполагаем, что любой кредит может находится в одном из k возможных состояний. Вероятность того, что случайно выбранный кредит находится в i-м состоянии в момент времени t, обозначим через $x_i(t)$, $i = 1, \ldots, k$. Выполняются условия:

$$0 \leqslant x_i(t) \leqslant 1, \quad x_1(t) + \ldots + x_k(t) = 1. \tag{1}$$

Обозначим через $p_{ij}(t)$ вероятность перехода кредита, находившегося в состоянии S_i в момент времени t, в состояние S_j за один шаг. Будем предполагать, что эта вероятность постоянна, т. е. $p_{ij}(t) \equiv p_{ij}$. Получим, что динамика каждого отдельно взятого кредита описывается марковской цепью с дискретным временем:

$$x_j(t+1) = \sum_{i=1}^{k} p_{ij} x_i(t), \quad t = 0, 1, \ldots, T. \tag{2}$$

Через $\{e_1, \ldots, e_k\}$ будем обозначать единичный базис в $\mathbb{R}^k$. Через $\xi(t)$ обозначим вектор состояния системы в момент времени t. Если кредит находится в i-м состоянии, то $\xi(t) = e_i$. Таким образом, вероятности состояний можно записать в виде

$$x_i(t) = \mathcal{P}\{\xi(t) = e_i\}, \quad i = 1, \ldots, k. \tag{3}$$

Здесь и далее через $\mathcal{P}(A)$ будем обозначать вероятность события A.

Обозначим вектор вероятностей состояний как $x(t) = \{x_1(t), \ldots, x_k(t)\}^\top$, $^\top$ – знак транспонирования, а $(k \times k)$-матрицу переходных вероятностей через $P = \{p_{ij}\}$. Динамика системы (2) запишется в виде:

$$x(t+1) = P^\top x(t), \quad t = 0, 1, \ldots, T. \tag{4}$$

Предполагается, что матрица P точно не задана и оценивается в процессе функционирования системы.

Критериями качества портфеля кредитов являются его доходность и риск, определяемый как доля "проблемных" кредитов, поэтому в качестве целевой функции будем рассматривать линейную функцию.

Пусть N – количество кредитов в портфеле, рассмотрим задачу оценки параметров распределения суммы одинаково распределенных случайных векторов. Обозначим

$$\zeta_N(t) = \sum_{n=1}^{N} \xi^{(n)}(t),$$

где для каждого фиксированного $t = 1, 2, \ldots, T$ случайные векторы $\xi^{(n)}(t)$, $n = 1, \ldots, N$ независимы и одинаково распределены, их распределение совпадает с распределением случайного вектора $\xi(t)$.

Будем рассматривать линейную целевую функцию на шаге m:

$$\eta(m,T) = \sum_{t=m}^{T} [l(t)^\top \xi(t)].$$

Здесь $l(t)$ – вектор-функция со значениями в $\mathbb{R}^k$.

При рассмотрении в качестве целевой функции дохода портфеля вектор $l(t)$ можно представить в виде $l(t) = r(t)l$, где $l = \{l_1, \ldots, l_k\}^\top$, l_i – доходность (убыточность, если $l_i < 0$) кредитов i-й группы, $r(t)$ – дисконтирующий множитель.

Другой часто используемый критерий – доля проблемных кредитов в конце периода. В этом случае целевая функция имеет вид $\eta(T) = (l^{(0)})^\top \xi(T)$, где $l^{(0)} = \{0, \ldots, 0, 1\}$.

Через

$$\eta_N(m,T) = \sum_{t=m}^{T} [l(t)^\top \zeta_N(t)]$$

обозначим суммарную линейную функцию, отражающую доход всего портфеля. При оценке доли проблемных кредитов суммарная функция имеет вид

$$\eta_N(T) = \frac{1}{N} (l^{(0)})^\top \zeta_N(T).$$

В статье рассмотрена задача оценивания квантили распределения случайной величины $(-\eta_N(m,T))$ (убытков):

$$q_\alpha = \min\{q : \mathcal{P}\{-\eta_N(m,T) \leqslant q\} \geqslant \alpha\}, \tag{5}$$

Квантиль q_α – это такой уровень потерь, что вероятность того, что потери не превысят q_α, составляет α. Обычно в задаче управления портфелем берут $\alpha = 0,95$ или $\alpha = 0,99$.

Задачи оптимизации вероятности и квантили тесно связаны, их свойства исследовались в [8], а для случая неполной информации о распределении случайных возмущений – в статье [14].

В данном случае задача оценивания квантили связана с задачей нахождения вероятности непревышения случайной величиной $(-\eta_N(T))$ определенного уровня γ, т. е.:

$$d_\gamma = \mathcal{P}\{-\eta_N(m,T) \leqslant \gamma\}.$$

Важным показателем в задаче оценки качества кредитного портфеля является вероятность того, что портфель на интервале $[m;T]$ будет убыточным, т. е. d_0. Ещё один часто используемый показатель d_{-r} – вероятность того, что доход не превысит безрисковую процентную ставку. Значимое отличие от нуля вероятностей $\bar{d}_0$ и $\bar{d}_{-r}$ говорит о плохом качестве кредитного портфеля.

Так как $\eta_N(m,T)$ является суммой независимых одинаково распределенных величин, то она распределена приближенно нормально и

$$q_\alpha = -E(\eta(m,T))N + \tau_\alpha \sigma(\eta(m,T))\sqrt{N},$$

где $E(\eta)$ и $\sigma(\eta)$ – математическое ожидание и среднее квадратичное отклонение случайной величины η, τ_α – квантиль уровня α нормального распределения. При таком подходе задачи оценки квантили и вероятности сводятся к оценке математического ожидания $E(\eta(m,T))$ и дисперсии $\sigma^2(\eta(m,T))$ случайной величины $\eta(m,T)$.

В случае известной матрицы переходных вероятностей P математическое ожидание $E(\eta(m,T))$ и дисперсия $\sigma^2(\eta(m,T))$ являются линейной и квадратичной функциями от $x(m)$ соответственно. От матрицы переходных вероятностей P эти функции зависят мультипликативно:

$$E\eta(m,T) = E\sum_{t=m}^{T}[l(t)^\top \xi(t)] = \sum_{t=m}^{T}(P^{t-m}l(t))^\top x(m),$$

поэтому в случае неизвестной матрицы оценивание квантили (5) является достаточно сложной задачей, которая рассматривается в статье.

3. Оценивание переходных вероятностей

Всюду далее рассматривается случай, когда матрица переходных вероятностей постоянна, но точно не известна.

Для оценивания переходных вероятностей p_{ij} обычно используются статистические данные о переходе из одного состояния в другое. Будем рассматривать далее оценку вероятности перехода из i-го состояния в j-е по данным о количестве переходов на предыдущем шаге m по всем элементам системы (кредитным договорам), т. е. на основании реализовавшихся значений случайных величин $\xi^{(n)}(m-1)$, $\xi^{(n)}(m)$, $n = 1,\ldots,N$.

Обозначим число элементов системы, находившихся в i-м состоянии в момент $t = m-1$ через $n_i = n_i(m-1)$, а через $n_{ij} = n_{ij}(m)$ – количество элементов, перешедших из состояния i в состояние j на шаге m. Значение n_i равно реализации случайной величины

$$\sum_{n=1}^{N} e_i^\top \xi^{(n)}(m-1) = e_i^\top \zeta_N(m-1),$$

а количество переходов n_{ij} определяется как реализация случайной величины

$$\sum_{n=1}^{N} (e_i^\top \xi^{(n)}(m-1))(e_j^\top \xi^{(n)}(m)).$$

Через $w_{ij} = w_{ij}(m)$ обозначим частоту переходов из состояния S_i в состояние S_j на шаге m:

$$w_{ij} = \frac{n_{ij}}{n_i}. \tag{6}$$

Рассмотрим числовые характеристики случайного вектора, отражающего переходы из какого-либо состояния S в состояния S_j, $j = 1, \ldots, k$, за один шаг.

Утверждение 1. *Пусть ζ – случайный вектор, распределение которого определяется вектором вероятностей $p = (p_1, \ldots, p_k)^\top$,*

$$\mathcal{P}\{\zeta = e_j\} = p_j. \tag{7}$$

Здесь $e_1, \ldots, e_k$ – единичный базис в $\mathbb{R}^k$, $p_1 + p_2 + \ldots + p_k = 1$.

Тогда математическое ожидание и ковариационная матрица случайного вектора ζ равны

$$E(\zeta) = p, \tag{8}$$

$$K(\zeta) = E(\zeta\zeta^\top) - E(\zeta)E(\zeta)^\top = \begin{pmatrix} p_1q_1 & -p_1p_2 & \ldots & -p_1p_k \\ -p_2p_1 & p_2q_2 & \ldots & -p_2p_k \\ \ldots & \ldots & \ldots & \ldots \\ -p_kp_1 & -p_kp_2 & \ldots & p_kq_k \end{pmatrix}. \tag{9}$$

Здесь $q_j = 1 - p_j$.

Доказательство утверждения 1 приведено в Приложении.

Утверждение 2. *Пусть случайные векторы $\zeta^{(l)}$ независимы и одинаково распределены, причем при всех l распределение определяется равенствами*

$$\mathcal{P}\{\zeta^{(l)} = e_j\} = p_j. \tag{10}$$

Тогда математическое ожидание и ковариационная матрица вектора

$$\xi = \frac{1}{n}\sum_{l=1}^{l=n} \zeta^{(l)}$$

равны соответственно

$$E(\xi) = p, \quad K(\xi) = \frac{1}{n}K(\zeta^{(l)}) = \frac{1}{n}\begin{pmatrix} p_1q_1 & -p_1p_2 & \dots & -p_1p_k \\ -p_2p_1 & p_2q_2 & \dots & -p_2p_k \\ \dots & \dots & \dots & \dots \\ -p_kp_1 & -p_kp_2 & \dots & p_kq_k \end{pmatrix}. \tag{11}$$

Утверждение 2 следует из утверждения 1 и свойств моментов суммы независимых случайных величин.

4. Оценивание по агрегированным данным

Предположим, что вместо числа переходов из состояния в состояние будем наблюдать только отношение $y_i(t)$, которое равно доле объектов, находящихся в состоянии S_i в момент t, $t = 0, \dots, T$. Отсутствие данных о числе переходов связано с тем, что получение таких данных требует ведения учета по индивидуальной миграции кредитов, программное обеспечение которого есть не у всех банков. Кроме того, внешние наблюдатели имеют, как правило, дело только с агрегированными данными, т. е. с долями $y_i(t)$.

В [2, 6, 15, 16] рассмотрены методы оценки матрицы переходных вероятностей по агрегированным данным на основе метода наименьших квадратов и его обобщения. Эти методы применяются, если серия наблюдений T достаточно длинная.

Рассмотрим соотношение

$$y_j(t) = \sum_{i=1}^{k} y_i(t-1)p_{ij} + u_j(t-1), \quad t = 1, \dots, T. \tag{12}$$

Здесь $y_j(t)$ – наблюдаемые данные о вероятностях состояний; p_{ij} – неизвестные элементы матрицы; $u_j(t)$ – отклонения, которые нужно минимизировать.

Уравнение (12) запишем в векторной форме

$$Y = GP + U,$$

где

$$U = \{u_1(1), \dots, u_1(T), u_2(1), \dots, u_k(T)\}^\top \in \mathbb{R}^{kT},$$

$$Y = \{y_1(1), \dots, y_1(T), y_2(1), \dots, y_k(T)\}^\top \in \mathbb{R}^{kT},$$

G – $(kT \times kT)$-матрица,

В [6] предложен подход к определению матрицы переходных вероятностей P на основе минимизации суммы квадратов отклонений при линейных ограничениях на вероятности p_{ij}:

$$(Y - GP)^{\top}(Y - GP) \to \min,$$

$$p_{ij} \geqslant 0, \quad \sum_j p_{ij} = 1. \tag{13}$$

Так как дисперсии отклонений U зависят от вектора Y, обычный метод наименьших квадратов неэффективен. В [16] предлагается использовать обобщенный метод наименьших квадратов:

$$(Y - GP)^{\top} Q^{-1} (Y - GP) \to \min,$$

при тех же ограничениях (13). Здесь Q – оценка ковариационной матрицы ошибок U. Алгоритм состоит из поочередно выполняемых шагов:

а) оценки матрицы переходных вероятностей P обобщенным методом наименьших квадратов;

б) оценки матрицы Q – матрицы ковариации ошибок U.

В [2] метод максимального правдоподобия используется для оценки риска кредитования компаний.

5. Доверительные оценки переходных вероятностей

Рассмотрим строки матрицы переходов P. Обозначим $(k-1)$-мерный вектор вероятностей переходов из i-го состояния во все остальные через

$$p^{(i)} = \{p_{ij},\ j = 1, \ldots, k,\ j \neq i\}.$$

Рассмотрим множество возможных значений этих векторов. С учетом условий (13) получаем множество

$$\mathbb{R}_+^{k-1} = \{p_j : \quad 0 \leqslant p_j \leqslant 1, \quad \sum_{j=1}^{k-1} p_j \leqslant 1\} \subset \mathbb{R}^{k-1}.$$

Отметим, что оценки векторов переходных вероятностей из различных состояний $p^{(i_1)}$ и $p^{(i_2)}$, $i_1 \neq i_2$, не коррелированы.

Наблюдаемые частоты w_{ij}, определяемые соотношением (6), являются оценками вероятностей состояний, а разности $w_{ij} - p_{ij}$ имеют асимптотически нормальное распределение [17]. При построении оценок будем учитывать корреляцию между оценками вероятностей переходов из одного состояния во все остальные (см. утверждение 1).

Найдем доверительные множество уровня β для каждой строки матрицы переходов $p^{(i)}$. С учетом моментов его можно задать, в частности, в виде эллипсоида

$$Z_\beta^{(i)} = \{p_{ij} \in \mathbb{R}_+^{k-1} : (p^{(i)} - w^{(i)})^\top G_i (p^{(i)} - w^{(i)}) \leqslant \tau^{(\beta)}\}, \tag{14}$$

где $w^{(i)} = \{w_{ij}, j = 1, \ldots, k\}$, G_i – ковариационная матрица, определяемая равенством

$$G_i = \frac{1}{n_i} \begin{pmatrix} p_{i1}q_{i1} & -p_{i1}p_{i2} & \ldots & -p_{i1}p_{ik} \\ -p_{i2}p_{i1} & p_{i2}q_{i2} & \ldots & -p_{i2}p_{ik} \\ \ldots & \ldots & \ldots & \ldots \\ -p_{ik}p_{i1} & -p_{ik}p_{i2} & \ldots & p_{ik}q_{ik} \end{pmatrix}. \tag{15}$$

Здесь $\tau^{(\beta)}$ – квантиль уровня β распределения χ^2 с $k-1$ степенями свободы. Оценки для p_{ii} получаются из условий

$$p_{i1} + \ldots + p_{ik} = 1, \quad i = 1, \ldots, k. \tag{16}$$

Отметим, что ковариационные матрицы G_i зависят от неизвестных вероятностей, как и при традиционном оценивании вероятности события по его частоте. В связи с этим приходится заменять коэффициенты ковариационной матрицы их оценками.

При нахождении оценок вероятностей переходов в реальных системах учитывают, что не все переходы возможны.

Возьмем $\beta = \sqrt[k]{\alpha}$ и построим доверительные множества $Z_\beta^{(i)}$ для векторов вероятностей переходов $p^{(i)}$. Множество $Z_\beta^{(i)}$ можно задать в виде (14). Тогда доверительное множество уровня α для вероятностей переходов $\{p_{ij}, j \neq i\}$ запишется в виде

$$Z_\alpha = Z_\beta^{(1)} \times \ldots \times Z_\beta^{(k)}.$$

Будем обозначать через $\bar{Z}_\alpha$ доверительное множество для всех элементов матрицы P, построенное на основе Z_α с учетом условия (16).

Решение задачи оценивания вероятности распадается на три шага.

1. Построение доверительного множества $\bar{Z}_\alpha$ для элементов матрицы переходных вероятностей.

2. Построение информационного множества $X(T, \bar{Z}_\alpha)$ многошаговой системы при известном значении $x(m)$:

$$x(t+1) = P^\top x(t), \quad P \subset \bar{Z}_\alpha, \quad t = m, \ldots, T. \tag{17}$$

3. Нахождение границ изменения линейной целевой функции $y(T) = (l^{(0)})^\top x(T)$ при условии $x(T) \in X(T, \bar{Z}_\alpha)$. Другими словами, нахождение опорной функции $\rho(l^{(0)}|X(T))$ информационного множества системы (17).

6. Нахождение информационного множества системы

Рассмотрим подробнее задачу описания информационного множества системы для случая неточно заданной матрицы переходов.

Информационное множество системы (17) имеет вид

$$X(t, Z_\alpha) = \{x \in R_+^k : x = (P^\top)^{T-m} x(m), \quad P \in \bar{Z}_\alpha\}. \tag{18}$$

В случае, когда матрица P точно не задана, а известно лишь множество её возможных значений $\bar{Z}_\alpha$, для построения используются результаты по описанию информационных множеств мультипликативных систем, полученные в работах А.Б. Куржанского, А.И. Матасова, М.И. Гусева, Е.К. Костоусовой [9–12]. Однако алгоритмы решения такой задачи достаточно трудоемкие.

Рассмотрим случай, когда матрица переходных вероятностей *регулярна* в том смысле, что некоторая положительная степень L матрицы P содержит только неотрицательные элементы, т. е.:

$$p_{ij}^{[L]} > 0 \text{ для всех } i, j = 1, \ldots, k. \tag{19}$$

В этом случае марковская цепь эргодическая и существует стационарное распределение вероятностей, к которому стремятся вероятности состояний [18]:

$$\lim_{t \to +\infty} x_i(t) = x_i^*.$$

Отметим, что условие (19) гарантирует отсутствие поглощающих состояний. При наличии одного поглощающего состояния в системе вектор стационарных вероятностей совпадает с одним из базисных векторов и задача определения стационарного распределения вероятностей становится тривиальной.

Пусть марковская цепь регулярна. Если число шагов $T-m$ достаточно велико, то вектор состояния системы (2) в момент времени T лежит достаточно близко x^*, где x^* – стационарное состояние системы, то есть решение системы линейных уравнений

$$\begin{cases} P^\top x^* = & x^*, \\ l_1^\top x^* = & 1. \end{cases} \tag{20}$$

Здесь вектор $l_1 = (1, 1, \ldots, 1)^\top \in \mathbb{R}^k$.

Разность между $x(T)$ и x^* по какому-либо направлению $y \in \mathbb{R}^k$ можно оценить, используя следующее свойство стационарного решения x^*.

У т в е р ж д е н и е 3. Пусть P – $(k \times k)$-матрица переходных вероятностей с положительными элементами, d – минимальный элемент матрицы P, $d = \min_{i,j} p_{ij}$.

Для любого начального распределения вероятностей x_0 при $n = 1, 2, \dots$ выполняется неравенство

$$\|(P^\top)^n x_0 - x^*\| \leqslant 2(1 - 2d)^n, \tag{21}$$

где x^* – решение системы (20).

Доказательство утверждения 3 приведено в Приложении и основано на следующих леммах.

Л е м м а 1 [19, с. 297]. *Пусть P – $(k \times k)$-матрица переходных вероятностей с положительными элементами, $d > 0$ – минимальный элемент матрицы P.*

Пусть $y \in \mathbb{R}^k$ – произвольный вектор, M_n и m_n – максимальная и минимальная координаты вектора $y^{(n)} = P^n y$:

$$M_n = \max_i y_i^{(n)}, \quad m_n = \min_i y_i^{(n)}.$$

Тогда для всех $n = 0, 1, \dots$ выполняются неравенства:

$$m_{n+1} \geqslant m_n, \quad M_{n+1} \leqslant M_n; \tag{22}$$

$$(M_n - m_n) < (1 - 2d)^n (M_0 - m_0). \tag{23}$$

Л е м м а 2 [19]. *В условиях леммы 1 все координаты вектора $y^{(n)} = P^n y$ стремятся к одному и тому же числу*

$$\lim_{n \to +\infty} y_i^{(n)} = c(y), \tag{24}$$

где $c(y) = y^\top x^*$, x^* – решение системы (20).

Отметим, что неравенство (21) доказано лишь для положительных матриц переходных вероятностей, однако для регулярной марковской цепи некоторая степень L матрицы P неотрицательна и неравенство (21) можно применять для количества шагов, кратного L. Порядок L можно определить по графу состояний.

В рассматриваемом случае матрица P точно не задана. Поэтому систему (20) следует рассматривать как систему линейных уравнений с неточно заданной матрицей коэффициентов. Обозначим её решение, соответствующее какой-либо матрице P, через $x^*(P)$. В данном случае интересуемся множеством всех возможных решений, т. е.

$$X^*(Z) = \bigcup_{P \in Z} x^*(P).$$

Такие системы исследуются в рамках теории интервального анализа [20], где получены алгоритмы вычисления множества $X^*(Z)$ для случая задания неопределенности элементов матрицы коэффициентов системы в виде интервалов.

Вычисления удается существенно упростить, если в исходной системе возможны переходы только в соседние состояния. В этом случае матрица P записывается в виде:

$$P = \begin{pmatrix} 1-p_{12} & p_{12} & 0 & \dots & 0 & 0 \\ p_{21} & 1-p_{21}-p_{23} & p_{23} & 0 & \dots & 0 \\ 0 & p_{32} & 1-p_{32}-p_{34} & p_{34} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & p_{k,k-1} & 1-p_{k,k-1} \end{pmatrix} \tag{25}$$

и оценке подлежат только $2k-2$ элемента.

Если все вероятности переходов в соседние состояния неотрицательны, т. е. $p_{i,i+1} > 0$ и $p_{i+1,i} > 0$ для $i = 1, \dots, k-1$, то марковская цепь регулярна и существуют финальные вероятности.

Отметим, что матрица P^k содержит только положительные элементы, так как за k шагов можно из любого её состояния попасть в любое другое, поэтому гарантирована следующая оценка скорости сходимости

$$\|(P^\top)^{kn} x_0 - x^*\| \leqslant 2(1-2d)^n,$$

где $0 < d < 0,5$.

Запишем систему для нахождения стационарного решения (20) для марковской цепи с переходами только в соседние состояния:

$$\begin{cases} x_1^* = & (1-p_{12})x_1^* + p_{21}x_2^*, \\ x_2^* = & p_{12}x_1^* + (1-p_{21}-p_{23})x_2^* + p_{32}x_3^*, \\ \dots & \dots \\ x_k^* = & p_{k-1,k}x_{k-1}^* + (1-p_{k,k-1})x_k^*, \\ 1 = & x_1^* + \dots + x_k^*. \end{cases}$$

Непосредственными вычислениями показывается, что стационарное распределение вероятностей находится из соотношений:

$$\begin{gathered} x_1^* = (1 + \lambda_{12} + \lambda_{12}\lambda_{23} + \dots + \lambda_{12}\lambda_{23}\dots\lambda_{k-1,k})^{-1}, \\ x_2^* = \lambda_{12}x_1^*, \quad x_3^* = \lambda_{12}\lambda_{23}x_1^*, \quad x_4^* = \lambda_{12}\lambda_{23}\lambda_{34}x_1^*, \\ \dots \\ x_k^* = \lambda_{12}\lambda_{23}\dots\lambda_{k-1,k}x_1^*, \quad \lambda_{i-1,i} = \tfrac{p_{i-1,i}}{p_{i,i-1}}. \end{gathered} \tag{26}$$

Таким образом, решение системы (20) найдено аналитически, что теоретически позволяет находить множество $X^*(Z)$ всех возможных стационарных состояний системы для заданного множества переходных матриц Z.

7. Оценивание динамики кредитного портфеля для упрощенной схемы

Рассмотрим упрощенную схему классификации кредитов, выделив три основные группы. Разобьем все кредиты на группы в зависимости от задолженности со статусом не меньше определенного количества дней [5]:

1) кредиты без просрочки, в том числе новые (состояние S_1);

2) кредиты с просрочкой 1 - 65 дней и восстановленные – S_2;

3) проблемные кредиты, т. е. кредиты с просрочкой более 65 дней – S_3.

Восстановленным будем называть кредит, который в предшествующий период был проблемным, если по нему заёмщиком было проведено погашение текущей задолженности. Граф переходов рассматриваемой системы приведен на рис. 1.

Весь портфель кредитов в момент времени t разбивается на доли $\{y_1(t), y_2(t), y_3(t)\}$, сумма которых равна единице. Здесь $y_i(t)$ – доля кредитов i-й группы в портфеле в момент t.

При проведении винтажного анализа $[4, 5]$ по формуле (6) находятся отношения w_{ij} по данным о переходе из i-го состояния в j-е. Задача: оценить долю проблемных кредитов через некоторый промежуток времени, т. е. $y_3(T)$.

Отметим, что в схеме переходы $S_1 \to S_3$ и $S_3 \to S_1$ отсутствуют, следовательно, $p_{13} = p_{31} = 0$ и матрица P имеет вид:

$$P = \begin{pmatrix} 1-p_{12} & p_{12} & 0 \\ p_{21} & 1-p_{21}-p_{23} & p_{23} \\ 0 & p_{32} & 1-p_{32} \end{pmatrix}. \tag{27}$$

В данном случае $k = 3$ и следует оценить $3 \cdot 2 - 2 = 4$ переходные вероятности: p_{12}, p_{21},p_{23}, p_{32} по наблюдаемым значениям w_{ij}. Обозначим переходные вероятности через a_s, $s = 1, \ldots, 4$:

$$p_{12} = a_1, \quad p_{23} = a_2, \quad p_{32} = a_3, \quad p_{21} = a_4. \tag{28}$$

Матрица переходных вероятностей запишется в виде

$$P = \begin{pmatrix} 1-a_1 & a_1 & 0 \\ a_4 & 1-a_4-a_2 & a_2 \\ 0 & a_3 & 1-a_3 \end{pmatrix}. \tag{29}$$

Обозначим случайный вектор, описывающий доли кредитов, перешедших из i-го состояния в остальные, через ξ_i $(i = 1, 2, 3)$:

$$\xi_i = \frac{1}{n_i} \sum_{l=1}^{n_i} \xi_i^{(l)},$$

где n_i – количество кредитов, находившихся в состоянии S_i на предыдущем шаге, $\xi_i^{(l)}$ – независимые одинаково распределенные случайные величины, распределение которых задается вектором переходных вероятностей из i-го состояния

$$\mathcal{P}\{\xi_i^{(l)} = e_j\} = p_{ij}, \quad j = 1, \ldots, k.$$

Учитывая утверждение 2, найдем теоретические значения ковариационных матриц K_i случайных векторов ξ_i. Получим

$$K_i = \frac{1}{n_i} \begin{pmatrix} p_{i1}q_{i1} & -p_{i1}p_{i2} & \ldots & -p_{i1}p_{ik} \\ -p_{i2}p_{i1} & p_{i2}q_{i2} & \ldots & -p_{i2}p_{ik} \\ \ldots & \ldots & \ldots & \ldots \\ -p_{ik}p_{i1} & -p_{ik}p_{i2} & \ldots & p_{ik}q_{ik} \end{pmatrix}. \tag{30}$$

В рассматриваемой задаче допустимыми являются только четыре перехода. Обозначим соответствующие случайные величины через ζ_s, $s = 1, \ldots, 4$:

$$\zeta_1 = \xi_{12}, \quad \zeta_2 = \xi_{23}, \quad \zeta_3 = \xi_{32}, \quad \zeta_1 = \xi_{21},$$

здесь через ξ_{ij} обозначена j-я координата случайного вектора ξ_i.

Обозначим через $K(a)$ ковариационную матрицу вектора $\zeta = (\zeta_1, \ldots, \zeta_4)^\top$. С учетом (28) и (30) получим

$$K(a) = \begin{pmatrix} \frac{a_1(1-a_1)}{n_1} & 0 & 0 & 0 \\ 0 & \frac{a_2(1-a_2)}{n_2} & 0 & -\frac{a_2a_4}{n_2} \\ 0 & 0 & \frac{a_3(1-a_3)}{n_3} & 0 \\ 0 & -\frac{a_2a_4}{n_2} & 0 & \frac{a_4(1-a_4)}{n_2} \end{pmatrix}. \tag{31}$$

Проводились исследование статистических данных по относительно небольшой группе однотипных кредитов ($N = 900$). Желаемый период прогноза вероятностей состояний – 12 месяцев, т. е. $T = m + 12$, и оценивалась доля проблемных кредитов $y_3(T)$.

На основании статистических данных $W_{ij} = \{w_{ij}(t) | t = 1, \ldots, m\}$ о числе переходов из одной группы кредитов в другие были получены оценки вероятностей переходов $\hat{p}_{ij} = \hat{a}_s$. На основании этих оценок с учетом вида ковариационной матрицы (31) найдены оценки коэффициентов корреляции и дисперсий.

Расчеты проводились тремя разными способами.

1. Построение доверительных множеств Z_α для переходных вероятностей p_{ij}, $i \neq \neq j$ и нахождение финальных вероятностей с помощью решения системы (20).

2. Построение доверительных множеств Z_α для переходных вероятностей и построение информационного множества системы (17) для $T = m + 12$.

3. Имитационное моделирование динамики системы (17).

7.1. Построение доверительных множеств

Пусть случайный вектор η состоит из независимых случайных величин, $E(\eta_i) = 0$ и $\sigma(\eta_i) = 1$, тогда математическое ожидание и ковариационная матрица вектора $\zeta = \hat{a} + G\eta$ соответственно равны:

$$E(\eta) = \hat{a}, \quad cov(\zeta) = GG^\top.$$

С помощью стандартной процедуры [19] найдем (4×4)-матрицу G, такую что $GG^\top = K(\hat{a})$. Обозначим через B_α доверительное множество уровня α в форме четырехмерного куба для стандартного гауссовского вектора $\zeta \sim \mathcal{N}(0, I)$:

$$B_\alpha = \{b \in \mathbb{R}^4 : |b_s| \leqslant \tau_\beta, \quad s = 1, \ldots, 4\},$$

где τ_β – двух сторонняя квантиль уровня β нормального распределения, $\beta = \sqrt[4]{\alpha}$. Для $\alpha = 0,95$, $\beta = 0,987$, $\tau_\beta = 2,49$.

Доверительные множества для элементов $p_{ij} = a_s$ матрицы переходных вероятностей P вида (29) были построены в форме параллелотопов [12]:

$$Z_\alpha = \hat{a} + GB_\alpha \subset \mathbb{R}^4, \tag{32}$$

где $\hat{a}$– средние выборочные значения для соответствующих элементов матрицы.

Нахождение стационарного режима при неточно заданной матрице переходов. Как было показано выше, финальные вероятности состояний (стационарный режим работы) в случае специального вида матрицы P находятся с помощью уравнений (26).

Для рассматриваемого случая $n = 3$ финальная вероятность последнего состояния (доля проблемных кредитов) определяется по формуле

$$x_3^* = x_3^*(a) = \frac{a_1 a_2}{a_3 a_4}\left(1 + \frac{a_1}{a_4} + \frac{a_1 a_2}{a_3 a_4}\right)^{-1}. \tag{33}$$

Анализ равенства (33) показывает, что при положительных значениях a_s, $s = 1, \ldots, 4$, максимальное значение $x_3^*(a)$ достигается при максимальных значениях параметров a_1 и a_2 и минимальных a_3 и a_4. Это позволяет непосредственно вычислить $\max_{a \in Z_\alpha} x_3^*(a)$. Для рассмотренных данных и Z_α вида (32) получили, что $x_3^* \leqslant 0,35$ при $a \in Z_\alpha$.

Нахождение решения системы с неточно заданной матрицей переходов. Вычислялась оценка информационного множества по направлению $(0,0,1)^\top$ состояния системы (17) при $T = m + 12$ при условиях, что $a \in Z_\alpha = \hat{a} + GB_\alpha$ и матрица P имеет вид (29). При выполнении расчетов использовались результаты [12].

Получили после проведения расчетов $x_3(m + 12) \leqslant 0,24$ при $a \in Z_\alpha$. Таким образом, квантиль уровня $\alpha = 0,95$ доли проблемных кредитов в портфеле к концу периода можно оценить $q_\alpha \leqslant 0,24$.

Разница по сравнению с результатом первого подхода связана с разностью между финальной вероятностью третьего состояния x_3^* и его вероятностью через 12 шагов $x_3(m + 12)$. При средних значениях переходных вероятностей $p_{ij} = \hat{p}_{ij}$ эта разность составляет всего 0,05.

Анализ полученных результатов показывает, что первый достаточно простой способ, опирающийся на замену вероятностей состояний на финальные, дает слишком грубые оценки квантили доли проблемных кредитов, по крайней мере при прогнозировании на 12 шагов вперед.

7.2. Имитационное моделирование

Применялась следующая имитационно-аналитическая модель.

1. На основании статистических данных $W_{ij} = \{w_{ij}(t) |\ t = 1, \ldots, m\}$ о числе переходов из одной группы в другую определялись средние выборочные значения $\hat{p}_{ij} = \hat{a}_s$, $s = 1, \ldots, 4$.

2. При каждом прогоне программы случайным образом моделируется нормально распределеный вектор $a = (a_1, \ldots, a_4)^\top$. Математическое ожидание берется равным $\hat{a}$, а ковариационная матрица $K(\hat{a})$ находится по формуле (31).

3. На каждом прогоне имитационной модели осуществляется расчет по формулам (2) при постоянной матрице переходов. На выходе фиксируется доля проблемных кредитов через 12 шагов $X_3(T) = x_3(m + 12)$.

4. После M прогонов программы получаем выборку значений $\{X_3^{(l)}(T), l = 1, \ldots, M\}$. По выборке $\{X_3^{(l)}\}$ вычисляется выборочная квантиль $\hat{q}_\alpha(M)$ уровня α [8].

Так так вычисления происходят достаточно быстро, то при проведении серии вычислительных экспериментов M выбирали достаточно большим $M = 1000, 2000, 10000$.

При увеличении количества наблюдений M выборочная квантиль $\hat{q}_\alpha(M)$ приближается к точному значению квантили распределения q_α. Можно найти оценку дисперсии выборочной квантили для заданного уровня вероятности, используя результаты [13]. Получили оценку квантили распределения $q_\alpha = 0,134$, с вероятностью 0,999 ошибка оценки не превышает $0,002$.

На рис. 2 представлено сравнение эмпирической плотности распределения с нормальным распределением с теми же первыми моментами. Эмпирическое распределение достаточно близко к нормальному, но у него несколько более тяжелый правый "хвост" , есть небольшие положительные асимметрия и эксцесс. Эти особенности графика эмпирической плотности распределения наблюдались и при расчетах для других статистических данных. При замене эмпирического распределения нормальным с теми же моментами и нахождении квантили получается несколько меньшее значение $q_\alpha^{\mathcal{N}} = 0,130$, что говорит о предпочтительном использовании выборочной квантили распределения.

На рис. 2. приведены результаты 10^4 прогонов имитационной программы, выполненной в MathCad.

По горизонтальной оси отложено прогнозируемое значение доли проблемных кредитов $x_3(T)$ через 12 месяцев, по вертикальной – плотность распределения. Сплошная линия – эмпирическое распределение, пунктирная – нормальное.

Существенное различие между результатами оценивания квантили с помощью имитационного моделирования и с помощью построения доверительных множеств показывает более высокую точность имитационных оценок. Получение завышенной оценки квантили целевой функции при использовании доверительных множеств связано с тем, что множества Z_α выбирались стандартной формы (паралеллотопы), в то время как для получения точного значения оптимальной оценки квантили следует выбирать доверительные множества в виде множеств уровня целевой функции [8]. Учитывая наличие четырех случайных параметров и мультипликативную зависимость от них целевой функции, это вряд ли возможно в рассматриваемой задаче.

8. Заключение

В статье исследуются методы прогнозирования составляющих кредитного портфеля на основе модели простой марковской цепи с дискретным временем. Предполагается, что переходные вероятности точно не известны и оцениваются в ходе функционирования системы. Рассмотрены метод доверительного оценивания элементов матрицы переходных вероятностей и метод имитационного моделирования.

Метод с использованием доверительных оценок требует довольно громоздких вычислений и пока его удается реализовать только для задач с небольшим числом групп кредитов. Использование имитационного моделирования дает достаточно точные результаты, однако этот метод не позволяет проводить аналитические исследования. Методы проиллюстрированы на модельном примере с тремя группами кредитов.

ПРИЛОЖЕНИЕ

Доказательство утверждения 1. По определению математического ожидания для дискретного случайного вектора

$$E(\zeta) = \sum_{j=1}^{j=k} p_j e_j = p.$$

Матрица $\zeta\zeta^\top$ в соответствии с законом распределения ζ принимает следующие значения:

$$\mathcal{P}\{\zeta\zeta^\top = e_j e_j^\top\} = p_j, \tag{П.1}$$

где $e_j e_j^\top$ – $(k \times k)$-матрица, у которой на j-м месте по главной диагонали единицы, остальные элементы равны нулю.

Вычислим ковариацию

$$K(\zeta) = E(\zeta\zeta^\top) - E(\zeta)E(\zeta)^\top = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_k \end{pmatrix} - \begin{pmatrix} p_1^2 & p_1p_2 & \dots & p_1p_k \\ p_2p_1 & p_2^2 & \dots & p_2p_k \\ \dots & \dots & \dots & \dots \\ p_kp_1 & p_kp_2 & \dots & p_k^2 \end{pmatrix}.$$

Отсюда получаем равенство (9).

Доказательство утверждения 3. Возьмем произвольный вектор единичной длины $y \in R^k$, $\|y\| = 1$. Из лемм 1 и 2 следует, что для координат вектора $y^{(n)}$ выполняется неравенство

$$\max_i \|y_i^{(n)} - c(y)\| \leqslant (M_0 - m_0)(1 - 2d)^n. \tag{П.2}$$

Из того, что $\|y\| = 1$, следует, что

$$M_0 - m_0 = \max_i y_i - \min_i y_i \leqslant 2.$$

Умножим вектор $y^{(n)} - c(y)l_1$ на начальное распределение вероятностей x_0 и оценим модуль полученного числа. Получим с учетом неравенства (П.2):

$$|x_0^\top (y^{(n)} - c(y)l_1)| = |\sum_i x_{0,i}(y_i^{(n)} - c(y))| \leqslant \max_i \|y_i^{(n)} - c(y)\| \leqslant 2(1 - 2d)^n, \tag{П.3}$$

так как компоненты вектора x_0 неотрицательны и в сумме равны единице.

С другой стороны, с учетом того, что $c(y) = y^\top x^*$ и $x_0^\top l_1 = 1$, получаем

$$x_0^\top (y^{(n)} - c(y) l_1) = x_0^\top (P^n) y - c(y) x_0^\top l_1 = y^\top (P^\top)^n x_0 - y^\top x^* = y^\top ((P^\top)^n x_0 - x^*).$$

Отсюда, учитывая (П.3), получаем неравенство (21).

СПИСОК ЛИТЕРАТУРЫ

1. *Марков А.А.* Распространение закона больших чисел на величины, зависящие друг от друга // Изв. Физико-математического общества при Казанском университете. 1906. Сер. 2. Т. 15. С. 135–156.

2. *Jones M.T.* Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk. IMF Working Paper. 2005. WP-05-219.

3. *Thyagarajan V., Saiful M.* Retail Banking Loan Portfolio Equilibrium Mix: A Markov Chain Model Analysis // Amer. J. of Applied Sci. 2005. V. 2(1). P. 410–419.

4. *Журавель Ю.Ю.* Актуальные вопросы резервирования розничного кредитного портфеля // Банковский ритейл. 2007. № 4. С. 21–36.

5. *Тимофеев Н.А.* Математическая модель винтажного анализа кредитного портфеля банка // Вестн. УрГУПС. 2011. Т. 9. № 1. С. 86–92.

6. *Lee T.C., Judge G.G., Zellner A.* Estimating the Parameters of the Markov Probability Model From Aggregate Time Series Data. Amsterdam: North Holland, 1970.

7. Международная конвергенция измерения капитала и стандартов капитала. Уточненные рамочные подходы. Базельский комитет по банковскому надзору. М.: Банк международных расчетов, 2004.

8. *Кан Ю.С., Кибзун А.И.* Задачи стохастического программирования с вероятностными критериями. М.: Физматлит, 2009.

9. *Varaiya P., Kurzhanski A.B.* Ellipsoudal Methods for Dynamics and Control. Part I// J. of Mathematical Sci. 2006. V. 139(5). P. 6863–6901.

10. *Матасов А.И.* Введение в теорию гарантирующего оценивания. М.: Изд-во МАИ, 1999.

11. *Гусев М.И.* О внешних оценках множеств достижимости нелинейных управляемых систем// Тр. ин-та матем. и механ. УрО РАН. 2011. Т. 17. № 1. С. 60–69.

12. *Костоусова Е.К.* О внешних полиэдральных оценках для множеств достижимости систем с билинейной неопределенностью // Прикл. матем. и механ. 2002. Т. 66. Вып. 4. С. 559–571.

13. *Вишняков Б.В, Кибзун А.И.* Применение метода бутстрипа для оценивания функции квантили// А и Т. 2007. № 11. С. 46–60.

14. *Тимофеева Г.А.* Оптимальные и субоптимальные решения стохастически неопределенной задачи квантильной оптимизации// А и Т. 2007. № 7. С. 31–43.

15. *Kalbfleisch J. D., Lawless J. F.* Least-Squares Estimation of Transition Probabilities From Aggregate Data// Canadian J. of Statistics. 1984. V. 12(3). P. 169–182.

16. *MacRae E. Ch.* Estimation of Time-Varying Markov Processes with Aggregate Data// Econometrica, 1977. V. 45(1). P. 183—198.

17. *Anderson T. W., Goodman L. A.* Statistical Inference About Markov Chains// Annals of Mathematical Statistics. 1957. V. 28. P. 89–110.

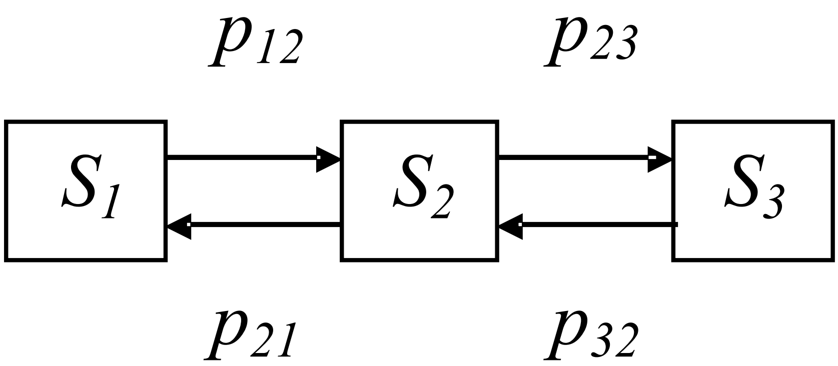
18. *Вентцель Е.С., Овчаров А.А.* Теория случайных процессов и её инжереные приложения. М.: Наука, 1991.

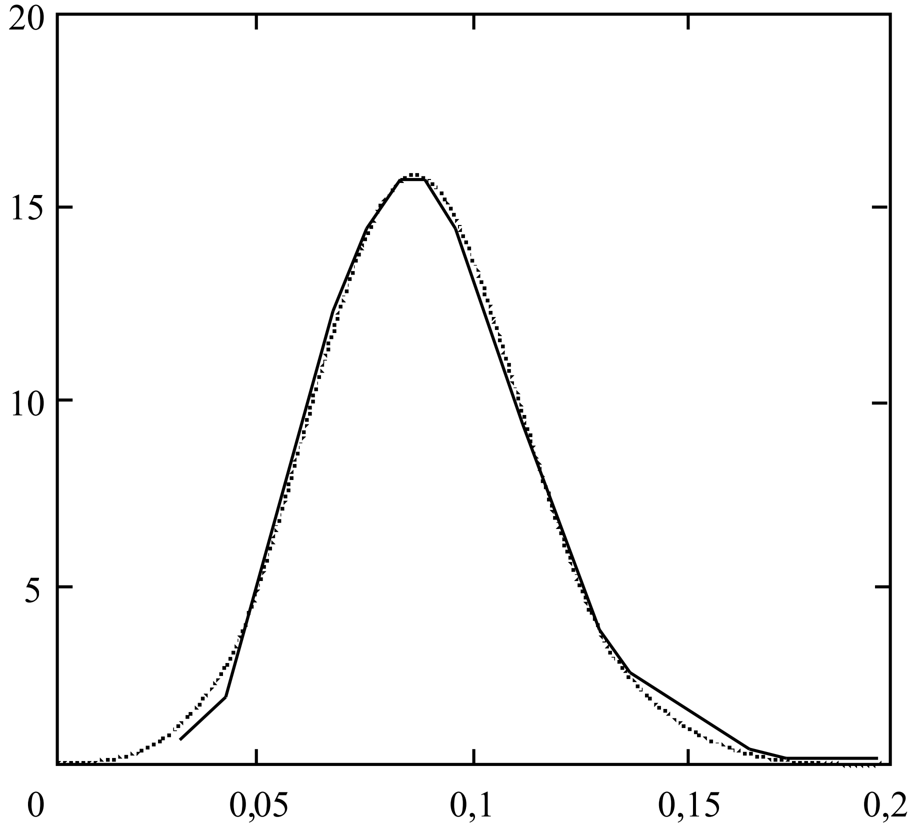
19. *Беллман Р.* Введение в теорию матриц. М.: Наука, 1968.

20. *Фидлер М., Недома Й., Рамик Я., Рон И. и др.* Задачи линейной оптимизации с неточными данными. М.– Ижевск: НИЦ Регулярная и хаотическая динамика, 2008.

Рис. 1. Граф состояний системы с тремя группами кредитов.

Рис. 2. Сравнение эмпирического и нормального распределений


p_{12}
p_{23}
S_1
S_2
S_3
p_{21}
p_{32}


20
15
10
5
0
0,05
0,1
0,15
0,2