

УДК: 519.87

Моделирование траекторий временных рядов с помощью уравнения Лиувилля

М. В. Гогуев, А. А. Кислицын^а

ИПМ им. Келдыша РАН,
Россия, 125047, г. Москва, Миусская пл., д. 4

E-mail: ^а alexey.kislitsyn@gmail.com

Получено 21.03.2023, после доработки — 19.04.2024.

Принято к публикации 24.04.2024.

Представлен алгоритм моделирования ансамбля траекторий нестационарных временных рядов. Построена численная схема аппроксимации выборочной плотности функции распределения в задаче с закрепленными концами, когда начальное распределение за заданное количество шагов переходит в определенное конечное распределение, так, что на каждом шаге выполняется полугрупповое свойство решения уравнения Лиувилля. Модель позволяет численно построить эволюционирующие плотности функций распределения при случайном переключении состояний системы, порождающей исходный временной ряд.

Основная проблема, рассматриваемая в работе, связана с тем, что при численной реализации левосторонней разностной производной по времени решение становится неустойчивым, но именно такой подход отвечает моделированию эволюции. При выборе неявных устойчивых схем с «заходом в будущее» используется итерационный процесс, который на каждом своем шаге не отвечает полугрупповому свойству. Если же моделируется некоторый реальный процесс, в котором предположительно имеет место целенаправленность, то желательно использовать схемы, которые порождают модель переходного процесса. Такая модель используется в дальнейшем для того, чтобы построить предиктор разладки, который позволит определить, в какое именно состояние переходит изучаемый процесс до того, как он действительно в него перешел. Описываемая в статье модель может использоваться как инструментальный моделирования реальных нестационарных временных рядов.

Схема моделирования состоит в следующем. Из заданного временного ряда отбираются фрагменты, отвечающие определенным состояниям, например трендам с заданными углами наклона и дисперсиями. Из этих фрагментов составляются эталонные распределения состояний. Затем определяются эмпирические распределения длительностей пребывания системы в указанных состояниях и длительности времени перехода из состояния в состояние. В соответствии с этими эмпирическими распределениями строится вероятностная модель разладки и моделируются соответствующие траектории временного ряда.

Ключевые слова: нестационарный временной ряд, выборочная функция распределения, аппроксимация скорости, кинетическое уравнение, полугруппа

Работа выполнена при поддержке гранта РФФИ, проект № 23-71-10055.

UDC: 519.87

Modeling time series trajectories using the Liouville equation

M. V. Goguev, A. A. Kislitsyn^a

Keldysh Institut of Applied Mathematics RAS,
4 Miusskaya sq., Moscow, 125047, Russia

E-mail: ^a alexey.kislitsyn@gmail.com

Received 21.03.2023, after completion – 19.04.2024.

Accepted for publication 24.04.2024.

This paper presents algorithm for modeling set of trajectories of non-stationary time series, based on a numerical scheme for approximating the sample density of the distribution function in a problem with fixed ends, when the initial distribution for a given number of steps transforms into a certain final distribution, so that at each step the semigroup property of solving the Liouville equation is satisfied. The model makes it possible to numerically construct evolving densities of distribution functions during random switching of states of the system generating the original time series.

The main problem is related to the fact that with the numerical implementation of the left-hand differential derivative in time, the solution becomes unstable, but such approach corresponds to the modeling of evolution. An integrative approach is used while choosing implicit stable schemes with “going into the future”, this does not match the semigroup property at each step. If, on the other hand, some real process is being modeled, in which goal-setting presumably takes place, then it is desirable to use schemes that generate a model of the transition process. Such model is used in the future in order to build a predictor of the disorder, which will allow you to determine exactly what state the process under study is going into, before the process really went into it. The model described in the article can be used as a tool for modeling real non-stationary time series.

Steps of the modeling scheme are described further. Fragments corresponding to certain states are selected from a given time series, for example, trends with specified slope angles and variances. Reference distributions of states are compiled from these fragments. Then the empirical distributions of the duration of the system’s stay in the specified states and the duration of the transition time from state to state are determined. In accordance with these empirical distributions, a probabilistic model of the disorder is constructed and the corresponding trajectories of the time series are modeled.

Keywords: nonstationary time series, sample distribution function, velocity approximation, kinetic equation, semigroup

Citation: *Computer Research and Modeling*, 2024, vol. 16, no. 3, pp. 585–598 (Russian).

The work was supported by the Russian Science Foundation, project No. 23-71-10055.

1. Введение

Во многих практических задачах требуется принять определенное управляющее решение относительно функционала, заданного на случайной траектории некоторых наблюдаемых данных. Функционалом может быть, например, правило действия торговой системы на бирже, определяемое на основе статистического анализа последовательности цен на финансовый инструмент. Большое количество эвристических моделей и индикаторов содержится в обзорах [Бэнсигнор, 2002; Бэстенс, ван дер Берт, Вуд, 1997; Эрлих, 1996]. Также функционалами могут быть так называемые кумулятивные суммы [Ширяев, 1996; Козинов, Мальцев, 2012] и иные индикаторы для анализа разладки [Шугай и др., 2004; Кислицын и др., 2019]. Если изучаемая последовательность представляет собой выборку из стационарного распределения [Корольок и др., 1985], то вероятность реализации заданной целевой функции может быть определена на уровне значимости, который зависит от длины выборки и дисперсии данных. Дальнейший статистический анализ можно провести классическими методами [Кремер, Путко, 2002]. Суть этих методов заключается в построении модели временных рядов как соответствующей динамической системы (например, регрессионной модели или интегрированной регрессионной модели). Если же ряд нестационарный, то ошибка, получаемая при использовании стационарного критерия, зачастую оказывается существенно больше уровня значимости. В этом случае требуется разработка других подходов к принятию решений, основанных на статистическом анализе предыстории. Одним из направлений анализа нестационарных данных является использование кинетических уравнений для описания эволюции плотности функции распределения выборки (SDFD). Такой подход был описан в [Орлов, Осминин, 2008; Босов, Кальметьев, Орлов, 2014; Орлов, 2014].

Мы не будем рассматривать ситуации, в которых нестационарные ряды сводятся к стационарным в рамках известных коинтегрированных моделей (см., например, [Кремер, Путко, 2002; Кендалл, Стюарт, 1976]). Цель настоящей работы состоит в создании инструментария для моделирования ансамбля траекторий временного ряда по анализу эволюции его выборочных распределений. Подчеркнем, что речь идет не о прогнозировании нестационарных распределений или функционалов от них, а о построении аппроксимации переходного процесса, когда состояние системы меняется от одного известного распределения к другому, также известному, за определенное число шагов. А именно, требуется построить модель эволюции плотности функции распределения случайной величины на определенном промежутке времени так, чтобы численное решение сохраняло нормировку распределения и обладало полугрупповым свойством. Это позволит сопоставить временному ряду модель в виде динамической системы.

Подчеркнем, что непараметрическая аппроксимация нестационарной функции распределения представляет собой нетривиальную задачу. Необходимо в конечно-разностной форме уравнения эволюции соблюсти как условие сохранения нормировки (что нарушается, например, при регрессионной аппроксимации распределения эмпирических частот), так и полугрупповое свойство, имеющее место быть в силу уравнения Лиувилля.

В данной работе изучаются такие нестационарные случайные процессы, которые допускают следующую интерпретацию. Предполагается, что изучаемая система может находиться в одном из S возможных состояний. Каждое состояние характеризуется функцией распределения его параметров (паттерном состояния), значения которых и наблюдаются в эксперименте. Мы будем считать, что параметры в каждом состоянии описываются стационарной функцией распределения, но в случайный момент времени происходит переключение системы в другое состояние. Поток событий, состоящий в переключении с одного состояния на другое, может быть нестационарным. Переход из одного состояния в другое предполагает, что начиная с некоторого момента времени расстояние в скользящем окне от выборочной функции распределения временного ряда до паттерна текущего состояния монотонно увеличивается, а до паттерна будущего состояния —

монотонно уменьшается. Несмотря на то что такая модель описывает весьма частный случай поведения временного ряда, она вполне адекватна для рядов, имеющих трендовое поведение.

Поскольку система находится в определенном состоянии конечное время, выборочная функция распределения ее параметров будет отличаться от генеральной совокупности для данного состояния. На достаточно большом промежутке времени наблюдения за системой может произойти несколько переключений, так что наблюдаемое распределение будет представлять собой смесь выборок из разных генеральных совокупностей. Это обстоятельство сильно затрудняет статистический анализ конкретного временного ряда, поскольку совокупности априори не известны. В таком случае более продуктивно рассмотреть пучок траекторий, которые отвечают эволюции распределения в соответствии с предположениями о частоте и характере переключений между состояниями. Тогда на таком ансамбле траекторий тестирование параметров функционала управления более корректно, чем на одной достаточно большой, но нестационарной реализации случайного процесса.

Подобный ансамбль траекторий может быть построен для моделирования и анализа поведения рядов длительности телефонных или интернет-соединений, энергопотребления [Wan et al., 2023], эпидемиологических моделей [Nguena Nguefack et al., 2020], последовательности магнитуд землетрясений, выбросов загрязняющих веществ в мегаполисах и других событий, моменты наступления которых, как и их значения, случайны. В частности, прогнозированию фондового рынка с использованием анализа временных рядов посвящено огромное количество исследовательских работ, среди которых можно выделить работу [Bacry et al., 2013; Tsantekidis, Passalis, Tefas, 2021], в которой анализируются модели точечных процессов, учитывающие рыночный шум, а также различные применения теоретических моделей [Bhardwaj, Swanson, 2006; Embrechts, Liniger, Lin, 2011] для описания движения цен финансовых инструментов.

В работе [Kislitsyn, Orlov, 2020] была предложена линейная по времени модель аппроксимации скорости эволюции функции распределения в задаче с закрепленными концами. Построенная модель позволила реализовать численную схему расчета выборочной плотности функции распределения нестационарного временного ряда с использованием уравнения Лиувилля. Для известных начального и конечного распределений, разделенных периодом времени порядка длины выборки, такая схема должна иметь возможность заполнить промежуточные значения возможных траекторий временного ряда. Для генерации этих значений необходимо задать функцию распределения в соответствующий момент времени. Нетривиальность задачи заключалась в том, что образец нестационарен, и обычное стационарное воспроизведение образцов во многих практических случаях не обеспечивает желаемой точности анализа. В данной статье на основе ранее предложенной модели будет построен численный алгоритм генерации траектории со случайными переключениями между состояниями и продемонстрирована его работоспособность.

2. Аппроксимация эволюции функции распределения

Кратко опишем численную схему решения уравнения Лиувилля для моделирования перехода системы от одного состояния к другому. Рассматривается одномерное уравнение Лиувилля относительно плотности $f(x, t)$:

$$\frac{\partial f(x, t)}{\partial t} + \frac{\partial}{\partial x}(u(x, t)f(x, t)) = 0, \quad (1)$$

где $u(x, t)$ есть заданная модельная скорость изменения плотности. В начальный момент времени распределение считается известным. Поскольку значения случайной переменной x в практических задачах конечны, будем считать их равномерно ограниченными по времени. Тогда можно положить, что $x \in [0; 1]$. Граничные условия для плотности $f(x, t)$ выбираем нулевыми.

При численном решении уравнения Лиувилля плотность представляется в виде гистограммы. Соответствующая дискретная схема выглядит следующим образом. Считаем, что область изменения случайной величины разбита на $n + 2$ классовых интервалов по каждой переменной, шаг по времени совпадает с отсчетом нумерации значений случайной величины. Тогда само уравнение эволюции с начальными и граничными условиями принимает вид

$$\begin{aligned} f(i, t + 1) &= f(i, t) + u(i, t)f(i, t) - u(i + 1, t)f(i + 1, t), \\ f(0, t) &= 0, \quad f(n + 1, t) = 0, \quad u(0, t) = 0, \\ f(i, 0) &= f_0(i). \end{aligned} \tag{2}$$

Поскольку скорость $u(x, t)$ есть величина случайная и потому не известная в будущем, задачу (2) предлагается переформулировать следующим образом. Пусть через некоторый промежуток времени, условно единичный (это шаг по времени в формуле (2)), плотность стала равной $f(i, 1) = f_1(i)$. Тогда из первых двух уравнений (2) можно определить подходящую скорость $u(i, 0) \equiv v(i)$, которая соответствует такому переходу. Для этого перепишем собственно эволюционное уравнение (2) в виде

$$v(i + 1)f_0(i + 1) = f_0(i) + v(i)f_0(i) - f_1(i) \tag{3}$$

и будем рекуррентно разрешать его относительно скоростей $v(i)$, считая известной скорость в нулевой граничной ячейке: $v(0) = 0$. Тогда получаем для скорости в первой ячейке нулевое значение, а для скорости во второй ячейке имеем

$$v(2) = \frac{f_0(1) \cdot (1 + v(1)) - f_1(1)}{f_0(2)} = \frac{f_0(1) - f_1(1)}{f_0(2)}.$$

Аналогично получаем для $v(i)$ выражение

$$v(i) = \frac{F_0(i - 1) - F_1(i - 1)}{f_0(i)}, \quad 1 \leq i \leq n; \quad F_{0,1}(i) = \sum_{j=1}^i f_{0,1}(j). \tag{4}$$

Из (4) следует, что вычисления корректны, если во всех внутренних ячейках плотность отлична от нуля. Далее будем предполагать, что это условие выполнено.

Найденную скорость $v(i)$ в (4) будем называть скоростью макроэволюции. Нас будет интересовать модель этой эволюции, если рассмотреть ее на микроуровне, то есть в условиях, когда в требуемых на практике единицах времени длина шага макроэволюции составляет N отсчетов. Соответствующую скорость изменения плотности в i -м классовом интервале обозначаем, как и в (2), через $u(i, k)$, $0 \leq k \leq N - 1$. Целью моделирования является представление перехода от $f_0(i)$ к $f_1(i)$ в виде N последовательно выполняемых шагов.

Легко проверить, что скорость макроэволюции в правосторонней разностной схеме первого порядка аппроксимации уравнения Лиувилля в общем случае не может быть представлена как сумма скоростей микроэволюции. Это естественное следствие того, что сумма полугрупп в общем случае не является полугруппой. Поэтому найденное численно на шаге k значение плотности $f(i, k)$ не может служить начальным условием той же задачи, но продолжаемой для моментов времени $k + 1, k + 2, \dots$. Поскольку моделирование временного ряда представляет собой более или менее удачную его аппроксимацию в виде динамической системы, было бы желательно, чтобы при численном воспроизведении выборочного распределения полугрупповое свойство уравнения Лиувилля выполнялось.

Отметим, что проблемы решения уравнения переноса и построение подходящих разностных схем, сохраняющих положительность плотности распределения, обсуждались многими авторами (в качестве примера приведем работу [Самарский, Вабищевич, 1998]). Однако нашей

целью является не решение уравнения переноса с заданным начальным условием, а аппроксимация решения задачи с закрепленными концами, то есть вопросы сходимости и порядка точности схемы не обсуждаются. Основным требованием к схеме является сохранение полугруппового свойства. При этом мы хотим сохранить и «физическое целеполагание» процесса, т. е. не использовать неявные схемы, чтобы не залезать в будущее, не переходить от уравнения первого прядка к уравнению более высокого порядка путем усложнения шаблона дискретизации и т. п.

Далее используем концепцию скорости макроэволюции $v(x)$ для модификации численной схемы решения уравнения Лиувилля. Будем аппроксимировать эволюцию плотности распределения от начального момента времени $t = 0$ до момента $t = k$ за один шаг длиной k . Введем соответствующую скорость для шага длиной k в модели линейной аппроксимации по времени. А именно, скорость в i -й ячейке на шаге k моделируется формулой

$$u(i, k) = v(i) \frac{k}{N}. \quad (5)$$

Из (2) тогда следует, что

$$f(i, k) = f_0(i) + \frac{k}{N}(v(i)f_0(i) - v(i+1)f_0(i+1)). \quad (6)$$

Тем самым плотность распределения становится заданной и на промежуточных этапах эволюции. Покажем, что такой подход позволяет ввести некоторую новую скорость $w(i, k)$ так, что она описывает полугрупповой (то есть в вычислительном плане — итерационный) переход от момента k к моменту $k+1$ согласно (2), т. е.

$$f(i, k+1) = f(i, k)(1 + w(i, k)) - w(i+1, k)f(i+1, k). \quad (7)$$

Действительно, из (2), (5) получаем рекуррентно разрешаемую систему:

$$\begin{aligned} w(0, k) &= 0, \\ w(1, k) &= -\frac{1}{N} \frac{f_1(0) - f_0(0)}{f_0(1) + \frac{k}{N}(f_1(1) - f_0(1))} = 0, \\ w(i, k) &= -\frac{1}{N} \frac{F_1(i-1) - F_0(i-1)}{f_0(i) + \frac{k}{N}(f_1(i) - f_0(i))}, \quad 1 < i < n. \end{aligned} \quad (8)$$

Знаменатель в выражениях (8) для $w(i, k)$ всегда положителен (если только плотность не обращается в какой-либо ячейке в ноль, что исключено по условию задачи), и тогда корректно определена микроскорость в соответствующем уравнении Лиувилля (7). Построенная разностная схема позволяет получить аппроксимацию решения уравнения Лиувилля на заданном временном горизонте в задаче с закрепленными концами, то есть при заданных начальном и конечном распределениях. Это решение дается нелинейной аппроксимацией плотности функции распределения по времени в виде

$$f(i, k+1) = f(i, k) + \frac{f(i, k)}{N} \frac{F_0(i-1) - F_1(i-1)}{\left(1 - \frac{k}{N}\right)f_0(i) + \frac{k}{N}f_1(i)} - \frac{f(i+1, k)}{N} \frac{F_0(i) - F_1(i)}{\left(1 - \frac{k}{N}\right)f_0(i+1) + \frac{k}{N}f_1(i+1)}. \quad (9)$$

Таким образом, плотность функции распределения в промежуточные времена в такой задаче известна, и ее эволюция отвечает некоторой динамической системе. Скорость $w(i, k)$ в (8) можно трактовать как дискретную динамическую систему, непрерывный аналог которой дается уравнением

$$\frac{dx}{dt} = w(x, t) = \frac{F(x) - G(x)}{(N-t)F'(x) + tG'(x)}, \quad (10)$$

где $F(x)$ и $G(x)$ представляют начальную и конечную функции распределения соответственно, N есть горизонт аппроксимации.

Уравнение (10) есть линейное дифференциальное уравнение первого порядка относительно функции $t(x)$. Его решение имеет следующий вид:

$$t(x) = N \frac{F(x) - F(x(0))}{F(x) - G(x)}, \tag{11}$$

где $x(0)$ есть начальное значение координаты в момент времени $t = 0$.

Из такой интерпретации следует, что N величин $x(k)$ образуют фрагмент временного ряда с распределением, отвечающим определенной траектории динамической системы, реализующейся с вероятностью $f(x(0))$. Ансамбль траекторий, усредненный по начальному состоянию с функцией $F(x(0))$, дает функцию распределения $G(x)$ конечного состояния.

Если, например, рассматривается процесс перехода от равномерного распределения к треугольному, т. е. когда функция распределения меняется от $F(x) = x$ до $G(x) = 2x^2$, $0 \leq x \leq \frac{1}{2}$, и $G(x) = 4x - 2x^2 - 1$, $\frac{1}{2} < x \leq 1$, то соответствующие моменты времени $t_k = k$ определяют согласно (9) плотности распределения переходного процесса, где $F = F_0$, $G = F_1$. Явный вид уравнения зависит от способа дискретной аппроксимации функции распределения. Если считать, что область определения случайной величины разбита на n ячеек, и положить, что $x(i) = \frac{i}{n}$, $i = 0, 1, \dots, n$, то получаем из (4) выражение для макроскопической скорости в виде

$$v(i) = (i - 1) \left(1 - \frac{2}{n}\right), \quad i \leq \frac{n}{2}; \quad v(i) = \frac{2(i - 1)^2}{n^2} - \frac{3(i - 1)}{n} + 1, \quad i > \frac{n}{2}. \tag{12}$$

Подстановка этого выражения в (6) дает выражение для промежуточной плотности.

3. Генерация траекторий временного ряда в задаче с закрепленными концами

Как известно [Королюк и др., 1985], в основе алгоритмов генерации временного ряда $\{x_k\}$ с заданной непрерывной плотностью функции распределения $f(x)$, для которой $F(x)$ есть ее функция распределения, лежит утверждение о равномерном распределении на отрезке $[0; 1]$ функции распределения как случайной величины. А именно, генерируется произвольная последовательность чисел $\{y_k\}$, равномерно распределенных на $[0; 1]$, после чего по формуле обращения

$$y_k = F(x_k), \quad x_k = F^{-1}(y_k) \tag{13}$$

получается набор требуемых чисел $\{x_k\}$. Этот подход был ранее использован в [Орлов, 2014] для того, чтобы смоделировать траектории нестационарного временного ряда, если известна его функция распределения. В данном случае формула (9) представляет нестационарное решение уравнения Лиувилля, так что на каждом шаге по времени известна и соответствующая функция распределения:

$$F(i, k) = \sum_{j=1}^i f(j, k). \tag{14}$$

Однако такая функция не является непрерывной, поэтому для применения метода (13) аппроксимируем ее кусочно-линейными функциями по гистограмме $f(i, k)$ исходя из того, что в непрерывном представлении $F(x, k) = \int_0^x f(y, k) dy$. Здесь

$$f(x, k) = f_i(k), \quad x \in \left[\frac{i}{n}; \frac{i+1}{n}\right], \quad i = 0 \div n. \tag{15}$$

Тогда соответствующая функция распределения имеет вид

$$F(x, k) = (nx - i) \cdot f_{i+1}(k) + \sum_{j=1}^i f_j(k), \quad x \in \left[\frac{i}{n}; \frac{i+1}{n} \right], \quad i = 0 \div n. \quad (16)$$

В соответствии с (13) генерируется стационарный равномерно распределенный на $[0; 1]$ ряд чисел $\{y_k\}$ длиной N . В момент времени k решается уравнение относительно аргумента x_k :

$$y_k = F(x_k, k). \quad (17)$$

Так образуется одна из возможных траекторий временного ряда $\{x_k\}$. Если теперь взять M таких случайных выборок $\{y_{km}\}$, $k = 1 \div N$, $m = 1 \div M$, то получим M траекторий длиной N для реализации решения задачи с закрепленными концами в виде фрагмента статистического ансамбля.

Напомним: то, что мы получили, есть только один этап перехода от начальной функции распределения к конечной, то есть от одного состояния к другому. Теперь нам надо смоделировать процесс последовательного перехода к разным состояниям, причем моменты переключения являются случайными.

4. Модель потока переключений

Пусть в результате предварительной экспертной кластеризации фрагментов реального временного ряда или в рамках модельных представлений установлены S состояний, в которых может находиться система. Эти состояния определяются своими функциями распределения $F_s(i)$ или, как вариант, плотностями $f_s(i)$. Пусть известны вероятности перехода P_{sr} между состояниями s и r , а также распределения $q_s(N)$ каждого из состояний по длительности и распределения $\rho(\tau)$ длительностей переходов между состояниями, для простоты не зависящими пока от типа состояний, хотя технически эти распределения также можно снабдить индексами s и r .

В простейшем случае переключение мгновенно, так что аппроксимировать переход из одного состояния в другое не требуется. Тогда для построения траектории или ансамбля траекторий на горизонте планирования T предлагается следующий алгоритм действий.

На первом этапе задается начальное состояние — одно из эталонных распределений. Пусть это распределение $F_{s_1}(i)$. Для него из распределения длительности $q_{s_1}(N)$ генерируется длина серии N_1 до переключения в другое состояние. Если оказалось, что $N_1 \geq T$, то моделирование переключений на этом заканчивается. Если же $N_1 < T$, то осуществляем переключение. Выбор другого состояния s_2 производится случайным образом в соответствии с распределением вероятностей перехода $P_{s_1 s_2}$ тем же методом, что и генерация случайного числа в (13). После этого из распределения $q_{s_2}(N)$ генерируется длина N_2 этой второй серии и т. д. В результате построена некоторая траектория временного ряда. Другая траектория строится аналогично, но, в отличие от ансамбля близких траекторий для моделирования перехода между двумя состояниями, она не обязана быть похожей на первую, поскольку случайны не только моменты переключения, но и сами состояния. В результате на горизонте T будет наблюдаться довольно разнородная картина траекторий, статистика которых соответствует выбранной модели.

Потоком событий или, в соответствии с задачей, потоком переключений будем называть среднее по ансамблю число переключений в единицу времени. Если n_m есть число переключений для m -й траектории, вероятность реализации которой равна $p_m(T)$, то эмпирическое среднее число событий по ансамблю из M траекторий определяется как

$$\bar{n}(T) = \sum_{m=1}^M n_m p_m(T), \quad (18)$$

а собственно поток переключений тогда равен

$$\mu(T) = \frac{\bar{n}(T)}{T}. \tag{19}$$

Средний поток можно вычислить теоретически, однако при большом горизонте планирования T получающиеся выражения слишком громоздки. Для этого надо определить вероятность $\pi_n(T)$ того, что число переключений на данном горизонте T равно заданному числу n . Тогда

$$\bar{n}(T) = \sum_{k=0}^{\infty} k\pi_k(T). \tag{20}$$

Например, даже для двух состояний вероятность реализации некоторой случайной траектории вычисляется достаточно трудоемко. Вероятность реализации траектории без переключения равна

$$\pi_0(T) = \int_T^{\infty} q_1(t) dt = 1 - Q_1(T). \tag{21}$$

Здесь $Q_1(T)$ есть функция распределения, отвечающая плотности $q_1(t)$. Вероятность траектории с одним переключением равна

$$\pi_1(T) = \int_0^T Q_1(t)(1 - Q_2(T - t)) dt. \tag{22}$$

Для двух переключений получаем

$$\pi_2(T) = \int_0^T Q_1(t) \left(\int_t^T (1 - Q_2(\tau - t))(1 - Q_1(T - \tau)) d\tau \right) dt. \tag{23}$$

С увеличением числа переключений число сомножителей соответственно растет.

Если же состояний не два, а три, то возникает дополнительный учет вероятностей перехода. Например, в этом случае вероятность траектории с одним переключением равна

$$\pi_1(T) = P_{12} \int_0^T Q_1(t)(1 - Q_2(T - t)) dt + P_{13} \int_0^T Q_1(t)(1 - Q_3(T - t)) dt. \tag{24}$$

Следовательно, вместо аналитического исследования общего случая более продуктивно использовать формулу (18) для сгенерированного пучка траекторий при достаточно большом числе траекторий M . Поскольку же поток зависит от горизонта T , то он является нестационарным даже при стационарной матрице переключений.

Однако более интересной и практически важной является ситуация, когда есть некоторый переходный период, в течение которого система перестраивается с одного состояния на другое. Длительность перестройки является случайной величиной, функция распределения которой, вообще говоря, может зависеть от начального и конечного состояний. Мы рассмотрим модель, в которой длительность любого переключения задается распределением с плотностью $\rho(\tau)$. Теперь случайная длина N_1 серии из первого состояния с распределением $q_1(N)$ продолжается случайным промежутком τ_1 , в течение которого надо смоделировать переход от первого распределения ко второму по методике, описанной выше в §§ 2, 3. Длительность второй серии N_2 получается из распределения $q_2(N)$, после чего снова идет переключение в течение случайного промежутка времени τ_2 и т. д.

5. Пример работы алгоритма

Паттерны состояний можно задать теоретически, а можно выделить экспертно для конкретно наблюдаемого фрагмента ряда. Ниже будет рассмотрен пример построения модели потока переключений на тестовых данных. Тестовые данные были сгенерированы на основе равномерного и нормального распределений на $(0; 1)$ (рис. 1). Нормальное распределение было взято со средним $0,5$ и среднеквадратичным отклонением 1 , так что на интервале $(0; 1)$ содержится примерно $0,69$ его нормировки. Соответствующая выборка длиной $T = 1000$ берется из перенормированного распределения.

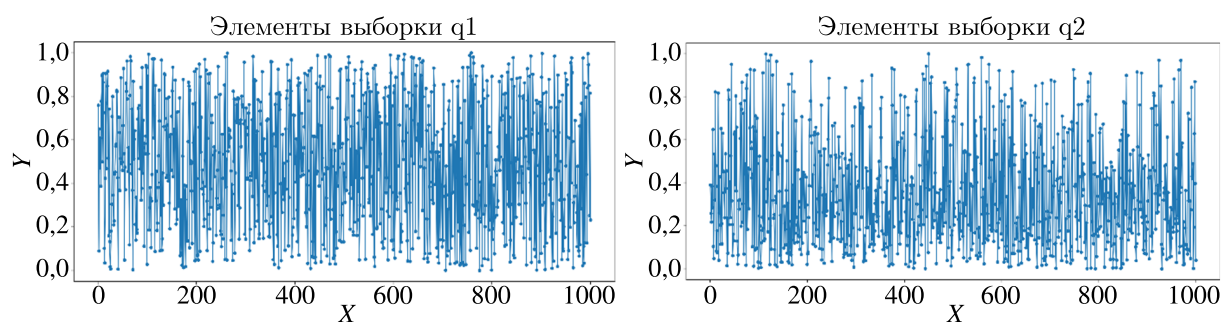


Рис. 1. Тестовые наборы временных рядов $q1$ и $q2$

По тестовым данным строятся эмпирические функции распределения. Организация переключений, как было описано в § 4, задаваемая вероятностью перехода P_{sr} между состояниями s и r , в данном случае будет происходить поочередно с одной функции распределения на другую. В свою очередь, распределение $q_s(N)$ каждого из состояний по длительности задается случайно сгенерированным вектором равномерно распределенных целочисленных значений от 0 до 100 .

В результате новая траектория состоит из фрагментов рядов со случайной длительностью. Визуально исходные функции распределения и процесс переключений в модели, когда время переключения имеет равномерное распределение на промежутке от 0 до 100 , представлены на рис. 2.

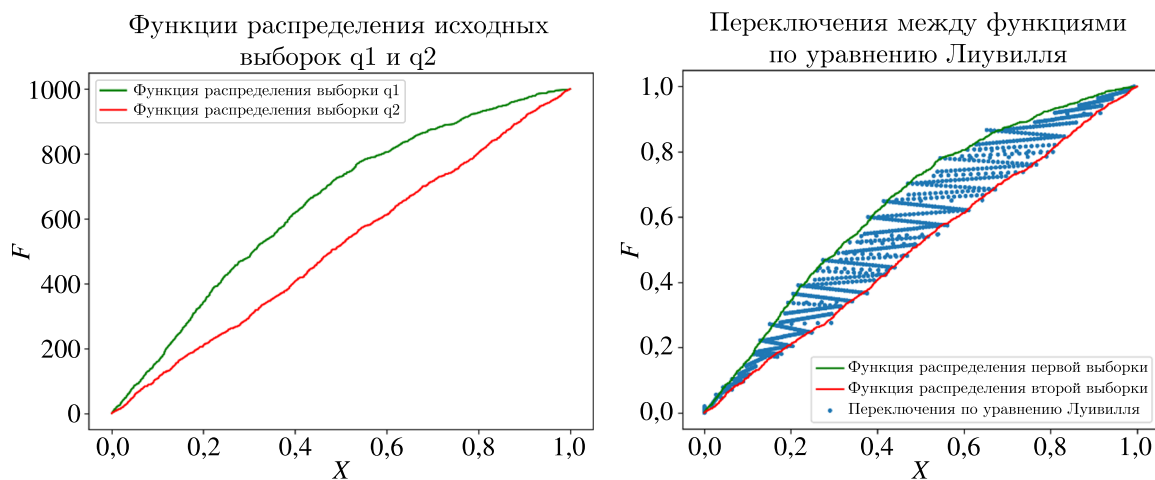


Рис. 2. а) Эмпирические функции распределения набора $q1$ (зеленый) и набора $q2$ (красный); б) визуализация процесса переключения между функциями распределения

Элементы новой траектории генерируются из переключаемых отрезков по формуле (17) с сохранением их порядка в исходной выборке. По окончании генерации новой траектории строится ее эмпирическая функция распределения (рис. 3).

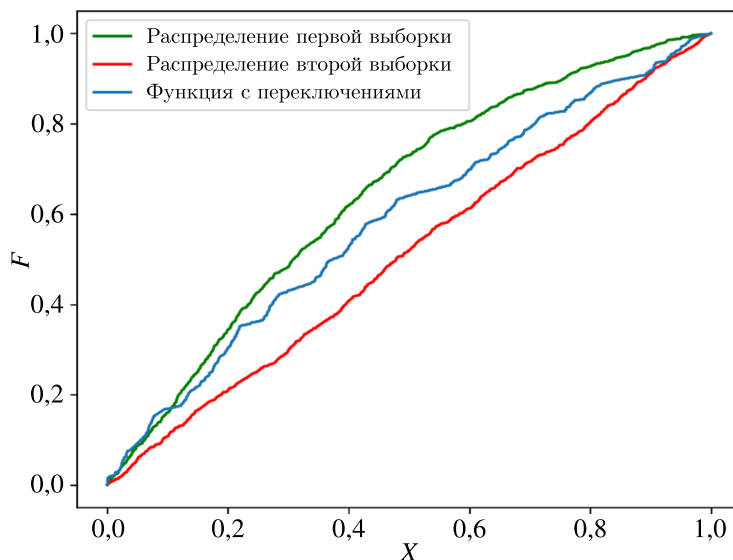


Рис. 3. Эмпирическая функция распределения сгенерированной траектории временного ряда

Таким же методом строятся набор траекторий временного ряда и, соответственно, набор выборочных функций распределения. На рис. 4 приведен пример такого ансамбля.

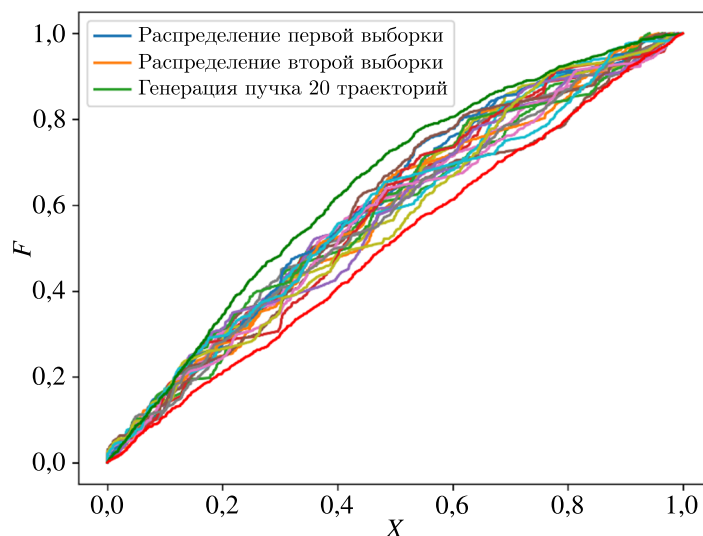


Рис. 4. Фрагмент ансамбля выборочных распределений

Разработанную схему моделирования можно применить на практике для индикации переходного процесса, т. е. построить предиктор разладки, следующим образом. Если предполагается процесс переключения между двумя состояниями с заданными плотностями $f_0(i)$ и $f_1(i)$, то наблюдаемая эмпирическая плотность функции распределения на переходном этапе из N шагов будет иметь вид

$$g_{01}(i) = \frac{1}{2}(f_0(i) + f_1(i)) + \frac{1}{2N}(f_1(i) - f_0(i)). \quad (25)$$

Если $N \gg 1$, то можно взять приближенные выражения для эмпирических плотностей распределений переходных процессов в виде

$$g_{kl}(i) = \frac{1}{2}(f_k(i) + f_l(i)) \quad (26)$$

и включить их в набор базисных состояний. Тогда, тестируя текущее выборочное состояние на близость к одному из эталонов, фиксируем предиктор разладки, если текущее состояние оказывается ближе к одному из эталонов переходного процесса.

Блок-схема алгоритма генерации новых траекторий временных рядов с переключением представлена на рис. 5.

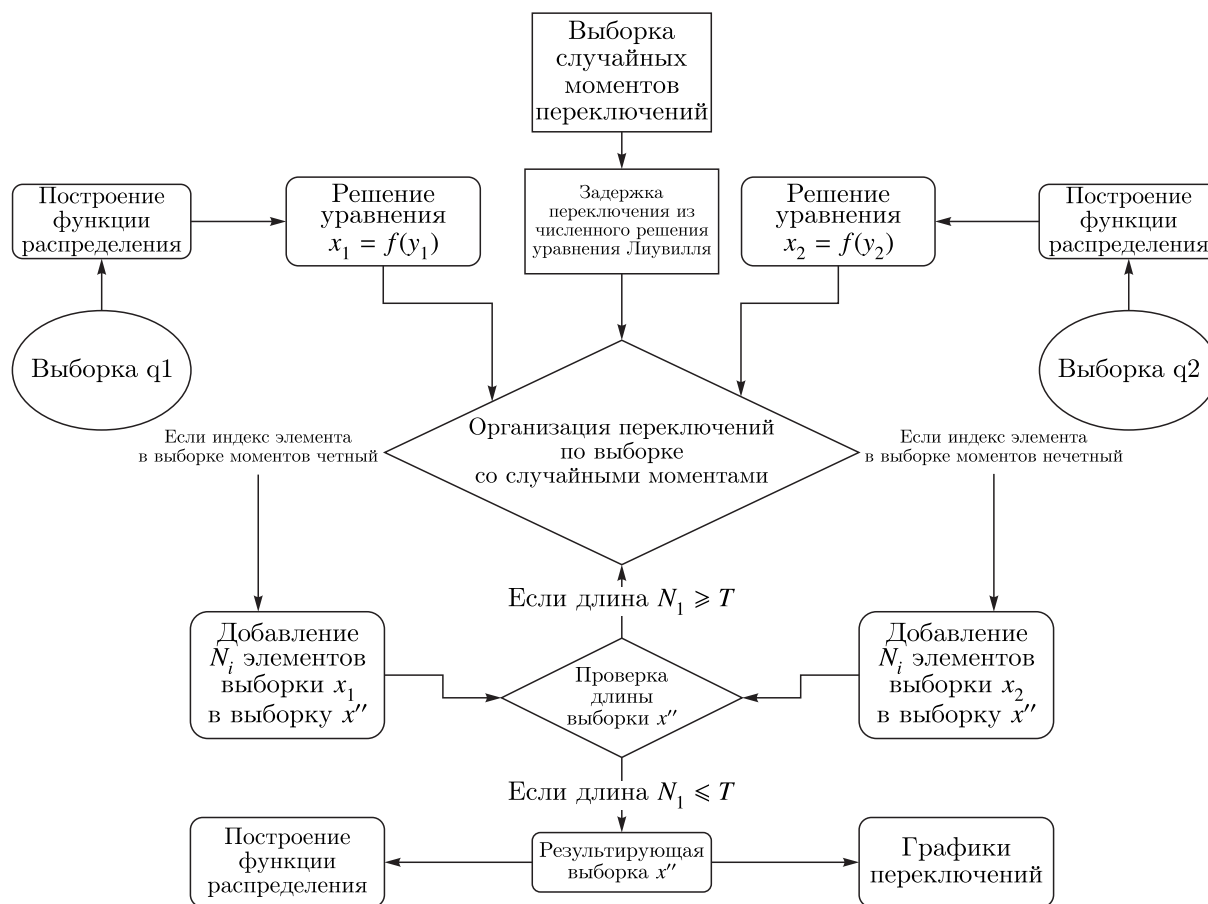


Рис. 5. Блок-схема алгоритма генерации с переключением

Описанный алгоритм позволяет генерировать временные ряды со случайными переключениями выборочных распределений. Важный аспект моделирования состоит в том, что каждая траектория временного ряда представляет собой реализацию некоторой динамической системы, статистические свойства которой отвечают уравнению Лиувилля.

6. Заключение

В работе представлен программный инструментарий, позволяющий проводить оптимизацию стохастического управления на основе анализа эмпирических данных, а также моделировать определенный класс нестационарных временных рядов с заданными статистическими

свойствами. Кроме того, модель переходного процесса позволяет строить предиктор разладки, для которого можно определить оптимальную ширину окна сканирования.

Существенно также и то, что алгоритм позволяет провести анализ чувствительности функционала управления к вариации основных статистик. Например, если распределения времен пребывания системы в том или ином состоянии нестационарны, то можно численно определить минимальную ошибку идентификации текущего состояния по близости выборочного распределения и эталонного паттерна в зависимости от уровня нестационарности указанных распределений.

Определенный интерес представляет также возможность выявления паттернов состояний по выборочной функции распределения, которая содержит смесь паттернов. Отметим, что построение паттернов, кластеризация фрагментов ряда и идентификация текущего состояния — основные элементы системы автоматической обработки данных, представляющих основу функционирования логически прозрачного искусственного интеллекта. Применение построенного алгоритма для решения указанных задач представляет перспективное направление дальнейших исследований.

Параграфы 2, 3, 5 написаны А. А. Кислицыным, § 4 написан М. В. Гогуевым, остальные разделы написаны совместно.

Список литературы (References)

- Босов А. Д., Кальметьев Р. Ш., Орлов Ю. Н.* Моделирование нестационарного временного ряда с заданными свойствами выборочного распределения // Математическое моделирование. — 2014. — Т. 26, № 3. — С. 97–107.
- Bosov A. D., Kal'met'ev R. Sh., Orlov Yu. N.* Modelirovanie nestacionarnogo vremennogo ryada s zadannymi svoystvami vyborochnogo raspredeleniya [Sample distribution function construction for non-stationary time-series forecasting] // Matem. modelirovanie. — 2014. — Vol. 26, No. 3. — P. 97–107 (in Russian).
- Бэнсигнор Р.* Новое мышление в техническом анализе. — М.: Интернет-трейдинг, 2002.
- Bensignor R.* New thinking in technical analysis. — Internet trading, 2000. (Russ. ed.: *Bensignor R. Novoe myshlenie v tekhnicheskom analize.* — Moscow: Internet-trejding, 2002.)
- Бэстенс Д. Э., ван дер Берг В. М., Вуд Д.* Нейронные сети и финансовые рынки: принятие решений в торговых операциях. — М.: ТВП, 1997.
- Bestens D. E., van der Bert V. M., Vud D.* Neural networks and financial markets: decision making in trading operations. — Moscow: TVP, 1997. (Russ. ed.: *Bestens D. E., van der Bert V. M., Vud D. Nejrionnye seti i finansovye rynki: prinyatie reshenij v trgovykh operacijah.* — Moscow: TVP, 1997.)
- Кендалл М., Стюарт А.* Многомерный статистический анализ и временные ряды. — М.: Наука, 1976.
- Kendall M., Styuart A.* Mnogomernyj statisticheskij analiz i vremennye ryady [Multivariate statistical analysis and time series]. — Moscow: Nauka, 1976 (in Russian).
- Кислицын А. А., Орлов Ю. Н., Козлова А. Б., Корсакова М. Б.* Индикатор разладки для нестационарных случайных процессов // Доклады РАН. Сер. Математика. — 2019. — Т. 484, № 4. — С. 393–396.
- Kislitsyn A. A., Orlov Yu. N., Kozlova A. B., Korsakova M. B.* Disorder indicator for nonstationary stochastic processes // Doklady Mathematics. — 2019. — Vol. 99, No. 1. — P. 57–59. (Original Russian paper: *Kislitsyn A. A., Orlov Yu. N., Kozlova A. B., Korsakova M. B. Indikator razladki dlia nestacionarnih sluchainih processov // Doklady Akademii nauk. Ser. Matimatika.* — 2019. — Vol. 484, No. 4. — P. 393–396.)
- Козинов И. А., Мальцев Г. Н.* Модифицированный алгоритм обнаружения разладки случайного процесса и его применение при обработке многоспектральных данных // Обработка информации и управление. — 2012. — № 3. — С. 9–17.
- Kozinov I. A., Maltsev G. N.* Modificirovannyj algoritm obnaruzheniya razladki sluchajnogo processa i ego primenenie pri obrabotke mnogospetral'nyh dannyh [Modified algorithm for detecting disorder of a random process and its application in processing multispectral data] // Information processing and management. — 2012. — No. 3. — P. 9–17 (in Russian).
- Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф.* Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985.
- Korolyuk V. S., Portenko N. I., Skorokhod A. V., Turbin A. F.* Spravochnik po teorii veroyatnostej i matematicheskoj statistike [Handbook on probability theory and mathematical statistics]. — Moscow: Nauka, 1985 (in Russian).

- Кремер Н. Ш., Путко Б. А.* Эконометрика. — М.: ЮНИТИ-ДАНА, 2002.
Kremer N. Sh., Putko B. A. Ekonometrika [Econometrics]. — Moscow: YUNITI-DANA, 2002 (in Russian).
- Орлов Ю. Н.* Кинетические методы исследования нестационарных временных рядов. — М.: МФТИ, 2014.
Orlov Yu. N. Kineticheskie metody issledovaniya nestacionarnyh vremennyh ryadov [Kinetic methods for studying non-stationary time series]. — Moscow: MFTI, 2014 (in Russian).
- Орлов Ю. Н., Осминин К. П.* Построение выборочной функции распределения для прогнозирования нестационарного временного ряда // Математическое моделирование. — 2008. — Т. 20, № 9. — С. 23–33.
Orlov Yu. N., Osminin K. P. Postroenie vyborochnoj funkicii raspredeleniya dlya prognozirovaniya nestacionarnogo vremennogo ryada [Sample distribution function construction for non-stationary time-series forecasting] // Mathematical Modeling. — 2008. — Vol. 20, No. 9. — P. 23–33 (in Russian).
- Самарский А. А., Вабищевич П. Н.* Нелинейные монотонные схемы для уравнения переноса // Доклады РАН. — 1998. — Т. 361, № 1. — С. 21–23.
Samarsky A. A., Vabishchevich P. N. Nelinejnyye monotonnnyye skhemy dlya uravneniya perenosa [Nonlinear monotonic schemes for the transport equation] // Doklady RAN. — 1998. — Vol. 361, No. 1. — P. 21–23 (in Russian).
- Ширяев А. Н.* Минимаксная оптимальность метода кумулятивных сумм в случае непрерывного времени // Успехи математических наук. — 1996. — Т. 51, № 4. — С. 173–174.
Shiryayev A. N. Minimaksnaya optimal'nost' metoda kumulyativnyh summ v sluchae nepreryvnogo vremeni [Minimax optimality of the method of cumulative sums in the case of continuous time] // Russian Math. Surveys. — 1996. — Vol. 51, No. 4. — P. 750–751 (in Russian).
- Шугай Ю. С., Доленко С. А., Персианцев И. Г., Орлов Ю. В.* Нейросетевые алгоритмы прогнозирования событий и поиска предвестников в многомерных временных рядах // Искусственный интеллект. — 2004. — № 2. — С. 211–215.
Shugaj Yu. S., Dolenko S. A., Persiancev I. G., Orlov Yu. V. Nejrosetevye algoritmy prognozirovaniya sobytij i poiska predvestnikov v mnogomernyh vremennyh ryadah [Neural network algorithms for forecasting events and searching for precursors in multidimensional time series] // Iskusstvennyj intellekt. — 2004. — No. 2. — P. 211–215 (in Russian).
- Эрлих А.* Технический анализ товарных и финансовых рынков. — М.: ИНФРА-М, 1996.
Erlih A. Tehnicheskij analiz tovarnih i finansovih rinkov [Technical analysis handbook for the commodity and financial markets]. — Moscow: INFRA-M, 1996 (in Russian).
- Bacry E., Delattre S., Hoffmann M., Muzy J. F.* Modeling microstructure noise with mutually exciting point processes // Quantitative Finance. — 2013. — Vol. 13, No. 1. — P. 65–77.
- Bhardwaj G., Swanson N. R.* An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series // Journal of Econometrics. — 2006. — Vol. 131. — P. 539–578.
- Embrechts P., Liniger T., Lin L.* Multivariate Hawkes processes: an application to financial data // Journal of Applied Probability. — 2011. — Vol. 48A. — P. 367–378.
- Kislitsyn A. A., Orlov Yu. N.* Dynamical system model with the use of Liouville equation for empirical distribution function densities // Discontinuity, Nonlinearity and Complexity. — 2020. — Vol. 9, No. 4. — P. 529–540.
- Nguena Nguetack H. L., Pagé M. G., Katz J., Choinière M., Vanasse A., Dorais M., Lacasse A.* Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches // Clinical Epidemiology. — 2020. — Vol. 12. — P. 1205–1222. — <https://doi.org/10.2147/CLEP.S265287>
- Tsantekidis A., Passalis N., Tefas A.* Diversity-driven knowledge distillation for financial trading using Deep Reinforcement Learning // Neural Networks. — 2021. — Vol. 140. — P. 193–202. — DOI: 10.1016/j.neunet.2021.02.026
- Wan A., Chang Q., Al-Bukhaiti K., He J.* Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism // Energy. — 2023. — Vol. 282. — P. 128. — DOI: 10.1016/j.energy.2023.128274