

На правах рукописи

Федоров Сергей Леонидович

Моделирование нестационарных временных рядов
и построение оператора эволюции их выборочных распределений
непараметрическими методами

Специальность 05.13.18 — математическое моделирование, численные методы
и комплексы программ

Автореферат диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2017

Работа выполнена в ИПМ им. М.В. Келдыша РАН

- Научный руководитель: Орлов Юрий Николаевич, доктор физико-математических наук, старший научный сотрудник, заведующий отделом № 6 вычислительной физики и кинетических уравнений ИПМ им. М.В. Келдыша РАН
- Официальные оппоненты: Сакбаев Всеволод Жанович, доктор физико-математических наук, доцент кафедры высшей математики Московского физико-технического института (г. Москва)
- Сопин Эдуард Сергеевич, кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов (г. Москва)
- Ведущая организация: Институт теории прогноза землетрясений и математической геофизики РАН (г. Москва)

Защита состоится « 25 » мая 2017 г. в « 15.30 » часов на заседании диссертационного совета Д 002.024.03, при ИПМ им. М.В. Келдыша РАН, 125047, Москва, Миусская пл., д.4

С диссертацией можно ознакомиться в библиотеке и на сайте ИПМ им. М.В. Келдыша РАН <http://keldysh.ru>

Автореферат разослан « ____ » _____ 2017 г.

Учёный секретарь диссертационного совета
к.ф.-м.н.

М.А. Корнилина



ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Проблема моделирования нестационарных временных рядов, возникающих во многих областях человеческой деятельности, в настоящее время приобрела большое практическое значение в связи с развившимися возможностями вычислительной техники и резко возросшей детализацией описания самих процессов. Существует большое число примеров рядов данных, требующих моделирования с учетом нестационарных свойств, которые проявляют выборочные распределения наблюдаемых величин. Таковы биржевые ряды цен сделок на финансовые инструменты, кардиограммы и энцефалограммы в медицине, сейсмограммы, температурные кривые и показатели счетчиков радиоактивности, последовательности символов в текстах и цепочках геномов.

Анализ нестационарных случайных данных является частью проблемы так называемых Больших Данных, когда требуется разработать эффективный инструмент для сокращения описания, позволяющий тем не менее давать содержательные ответы на интересующие исследователя вопросы. Исторически существует важный пример эффективности применения кинетического подхода к анализу Больших Данных, в рамках которого оказалось возможным сведение большого числа уравнений механики к малому числу уравнений гидродинамики. Однако к временным рядам, имеющим не только физическую природу, кинетический подход практически не применялся в силу отсутствия надежного динамического описания соответствующих процессов, а также и по причине вычислительной сложности возникающих статистических задач. Настоящая работа направлена на разработку и применение кинетического метода исследования Больших Данных, а также на создание программного продукта, достаточно универсального с точки зрения конкретной области его применения, для решения определенных задач стохастического управления, таких, например, как оптимизация функционала штрафа, заданного на фрагменте траектории случайного процесса.

Традиционный подход к анализу нестационарных временных рядов состоял в том, что рассматривались только такие ряды, которые с помощью линейных преобразований можно свести к стационарным. Соответствующие модели носят название авторегрессионных интегрированных моделей скользящего среднего. Основы моделей такого типа были заложены в середине прошлого века Боксом и Дженкинсом. Характерно, что эти модели оперируют не с выборочными функциями распределения, а непосредственно с элементами временного ряда. Ряды, не укладывающиеся в рамки регрессионного анализа, изучались разными эвристическими методами, называемыми адаптивными, не имеющими четкого математического обоснования. В них предполагается, что ряды на некотором прогнозном горизонте для выборки некоторой длины могут быть описаны той или иной стационарной моделью типа регрессии или авторегрессии, а потом параметры такой модели должны быть пересчитаны с

учетом новой информации или с учетом сравнения предсказанного значения с фактом. Недостатком этих подходов является то, что они применяются к единственной реализации случайного процесса, тогда как для эволюционирующих распределений методически более корректно изучать ансамбль возможных траекторий. Это требует использования кинетических уравнений – либо для генеральных совокупностей, либо для выборок. К преимуществам кинетического метода следует отнести также и то, что он не предполагает каких-то специальных свойств временных рядов, кроме естественного на практике требования равномерной ограниченности ряда по времени. Последнее нужно для того, чтобы при прогнозировании нестационарных распределений иметь возможность сравнения начального и конечного выборочного распределений на одной шкале значений случайной величины.

Кинетический подход к анализу нестационарных временных рядов развивается в настоящее время группой сотрудников в ИПМ им. М.В. Келдыша РАН под руководством д.ф.-м.н. Ю.Н. Орлова. Этот метод начал разрабатываться относительно недавно. В первых работах была исследована возможность представить эволюцию плотности выборочного распределения уравнением Лиувилля и была введена так называемая эмпирическая скорость переноса вероятности. Затем последовал ряд работ по конструированию новых индикаторов нестационарности. Подход с использованием уравнения Фоккера-Планка для описания эволюции выборочных функций распределения был предложен в 2012 г., однако соответствующий численный алгоритм реализован не был. Также не ставилась задача генерации виртуальных нестационарных траекторий, представляющая собой реализацию нестационарного метода Монте-Карло применительно к данной задаче. Эти вопросы и рассматриваются в представленной диссертационной работе.

Цель работы заключается в создании инструментария, реализованного в виде программного комплекса, для тестирования функционалов, заданных на траектории нестационарного случайного процесса, и для изучения их статистических свойств.

Научная задача – разработка непараметрических индикаторов нестационарности временного ряда и создание численного алгоритма генерации ансамбля нестационарных траекторий, являющихся реализациями решения кинетического уравнения, которым описывается эволюция выборочной функции распределения изучаемого ряда.

Направления исследования. Для достижения поставленной цели в работе определена следующая последовательность исследований. Необходимо:

1. Разработать математическую модель непараметрического индикатора нестационарности выборочных распределений временных рядов в разных нормах и реализовать ее в виде численного алгоритма.
2. Построить математическую модель эволюции выборочных функций распределения, такую, что уравнения эволюции моментов распределений

заданных порядков, следующие из кинетического уравнения, совпадали бы с эмпирически наблюдаемыми их изменениями по элементам выборки.

3. Построить алгоритм численного решения кинетического уравнения относительно эмпирической функции распределения.

4. Предложить модель генерации нестационарной траектории, статистические свойства которой совпадают в пределах точности эксперимента с наблюдаемой выборочной функцией распределения временного ряда, и реализовать ее в виде алгоритма генерации ансамбля траекторий.

5. Разработать метод тестирования функционала, заданного на выборочной траектории нестационарного случайного процесса, с целью его возможной оптимизации и для анализа его статистических свойств.

Методы исследования. В работе использовались кинетические методы анализа временных рядов, аппарат теории диффузионных процессов, а также аналитические и численные методы решения уравнения переноса.

Научная новизна работы заключается в том, что впервые введен и протабулирован индикатор, названный в работе согласованным уровнем значимости, сравнение с которым позволяет определить уровень нестационарности выборочных распределений временного ряда. Также впервые разработан и реализован в виде программного комплекса с интерфейсом нестационарный аналог метода Монте-Карло применительно к анализу и прогнозированию временных рядов.

Теоретическая ценность и практическая значимость работы состоит в том, что в ней выведена математически корректная оценка точности генерации ансамбля траекторий, отвечающего решению определенного кинетического уравнения, и сформулирован новый критерий нестационарности временных рядов, основанный на понятии согласованного уровня стационарности, введенного в диссертации. Результаты исследований были использованы в выполнении проектов ИПМ им. М.В. Келдыша РАН, поддержанных грантами РФФИ № 13-01-00617, № 14-01-00145, № 15-01-07944, № 16-01-00342. Практические задачи, изучаемые в рамках перечисленных проектов, относились к сферам самоорганизации в сложных системах, процессам коммуникации в системах беспроводной связи, а также к анализу эпидемиологической обстановки в мегаполисах.

Достоверность и обоснованность результатов подтверждается использованием строгих математических доказательств и рассуждений и апробированных в научной практике методов численного анализа.

Апробация работы. Материалы диссертации докладывались на научных семинарах в ИПМ им. М.В. Келдыша РАН, в РУДН, в Институте теории прогноза землетрясений и математической геофизики РАН, а также на конференциях:

1. Теоретические и прикладные аспекты современной науки. II научно-практическая международная конференция, август 2014, Белгород, Россия.

2. 14-th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19-25, 2016, Rhodes, Greece.

Публикации. По материалам диссертации опубликовано 10 работ. Из них 2 статьи в рецензируемых журналах, 2 статьи в трудах международных конференций и 6 препринтов ИПМ им. М.В. Келдыша РАН.

Личный вклад автора состоит в создании математических моделей временных рядов, разработке функционалов-индикаторов для описания нестационарности и создании программного комплекса, решающего вышеописанные задачи анализа временных рядов. В работах, опубликованных в соавторстве, лично соискателем разработан численный алгоритм и выполнены вычислительные эксперименты.

Структура и объем диссертации. Диссертация «Моделирование нестационарных временных рядов с заданными свойствами эволюции их выборочных распределений» состоит из введения, четырех глав и заключения. Каждая глава разбита на параграфы, имеющие двойную нумерацию, первая цифра которой указывает на соответствующую главу. Формулы внутри каждого параграфа имеют тройную нумерацию, с указанием на главу и параграф. Рисунки и таблицы имеют сквозную нумерацию. Результаты исследования изложены на 108 страницах и содержат 41 рисунок и 2 таблицы. Библиографический список состоит из 66 наименований, расположенных в алфавитном порядке. В автореферате сохранена нумерация формул и рисунков в соответствии с диссертацией.

Основные положения, выносимые на защиту, состоят в следующем.

1. Разработана математическая модель нестационарного временного ряда на основе численного решения эмпирического кинетического уравнения, которым описывается эволюция его выборочной функции распределения, и построена система индикаторов для идентификации уровня нестационарности в задачах статистического анализа нестационарных временных рядов.
2. Построен численный алгоритм генерации ансамбля траекторий нестационарного временного ряда и выборки из него в пределах заданного горизонта прогнозирования на основе решения уравнения Фоккера – Планка относительно выборочной плотности функции распределения, отвечающей данному временному ряду.
3. Разработан метод статистического анализа функционалов, заданных на траектории нестационарного случайного процесса, реализованный в виде программного комплекса с интерфейсом.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационного исследования, дан краткий обзор существующих подходов к обозначенной проблеме, указаны общие сведения о работе и приведен перечень результатов, выносимых на защиту.

В первой главе описывается модель временного ряда на основе кинетического уравнения эволюции его выборочных распределений и формулируется метод генерации траекторий случайного процесса, одной из реализаций которого является изучаемый временной ряд.

Для генерации ансамбля траекторий обоснована следующая схема действий. На первом этапе по имеющейся выборке данных строятся выборочные распределения приростов $x(t)$ значений случайной величины за тот промежуток времени T , который представляет интерес. Тем самым в каждый момент времени t определена выборочная плотность функции распределения (ВПФР) $f_T(x, t)$ и соответствующая ей выборочная функция распределения (ВФР) $F_T(x, t)$. Затем генерируется стационарный равномерно распределенный на $[0; 1]$ ряд чисел $\{y_k\}$ длиной T . Пусть t_0 есть начальный момент времени, в который известна функция распределения за некоторый период наблюдения. Тогда в последующие моменты времени одна из возможных траекторий случайного процесса, для которого ВПФР меняется от $f_T(x, t_0)$ до $f_T(x, t_0 + T)$, строится по формуле обращения соответствующей локальной по времени функции распределения, движущейся в скользящем окне длины T :

$$y_k = F_T(x_k, t_0 + k). \quad (1.5.4)$$

Подчеркнем, что, согласно (1.5.4), в каждый момент времени t из распределения $F_T(x, t)$ генерируется только одно значение ряда. Сама же $F_T(x, t)$ выступает в этот момент времени как генеральная совокупность. Тем самым имитируется процесс наблюдения за динамикой нестационарного временного ряда.

Задавая различные равномерно распределенные ряды $\{y_k\}_j$, где $j = 1, \dots, S$ есть номер генерации, можно получить пучок траекторий, ассоциированных с ВПФР текущих выборок.

Результаты первой главы опубликованы в работах [1, 2, 8].

Во второй главе проводится теоретический анализ модели, в ней также введены индикаторы нестационарности временного ряда: согласованный уровень стационарности и индекс нестационарности. Основным результатом главы II является теоретическое обоснование модели эволюции ВПФР, использующей эмпирическое уравнение Фоккера-Планка.

В главе II вводится согласованный уровень значимости ε по формуле

$$1 - K\left(\sqrt{\frac{N}{2}}\varepsilon\right) = \varepsilon, \quad (2.1.4)$$

где $K(z)$ есть функция Колмогорова. Решение этого уравнения $\varepsilon = \varepsilon^*(N)$ обладает тем свойством, что вероятность превышения значения $\varepsilon^*(N)$ равна значимости используемого для этой цели критерия.

Далее анализируется статистика расстояний в норме S между так называемыми встык-выборками изучаемого ряда, т.е. между ВФР $F_N(x, t)$ и $F_N(x, t + N)$, сдвинутыми одна относительно другой на величину окна выборки:

$$\rho(N; t) = \|F_N(x, t) - F_N(x, t + N)\| \quad (2.1.6)$$

и строится функция распределения $G(\rho; N)$ расстояний (2.1.6), которая представляет эмпирическую вероятность того, что расстояние между распределениями не больше ρ . Вводится согласованный уровень стационарности (СУС) $\rho^*(N)$ так, что соответствующее расстояние равно значимости критерия, т.е. СУС является решением относительно ρ уравнения

$$G(\rho; N) = 1 - \rho. \quad (2.1.7)$$

В стационарном случае уравнение (2.1.7) переходит в уравнение (2.1.4).

Индексом нестационарности временного ряда будем называть отношение

$$J(N) = \frac{\rho^*(N)}{\varepsilon^*(N)}. \quad (2.1.8)$$

Этот индекс показывает, во сколько раз доля расстояний, больших СУС, превосходит аналогичный показатель для стационарных рядов. Если $J(N) \leq 1$, ряд считается стационарным, а если $J(N) > 1$, то ряд нестационарный. Индекс нестационарности позволяет проанализировать, на каких длинах выборки ряд ведет себя более или менее нестационарным образом, что важно при разработке других индикаторов, основанных на выборочных статистиках.

Пусть $f(x, t)$ – одномерная ВПФР, построенная по выборке некоторого объема, который далее считается постоянным и потому не указывается в аргументах ВПФР. Обозначим также $F(x, v, t)$ двумерную ВПФР значений ряда и его приращений, и введем еще трехмерную ВПФР $\Phi(x, v, w, t)$ значений ряда, его первых и, соответственно, вторых разностей. Тогда в приближении Фоккера-Планка функция $F(x, v, t)$ удовлетворяет двумерному уравнению диффузионного типа. Если матрица диффузионных коэффициентов, определяемая в данном случае коэффициентами λ, χ, μ , не зависит от x, v , а зависит только от времени t , то уравнение Фоккера-Планка для $F(x, v, t)$ имеет вид

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x}(vF) + \frac{\partial}{\partial v}(WF) - \frac{\lambda(t)}{2} \frac{\partial^2 F}{\partial x^2} - \chi(t) \frac{\partial^2 F}{\partial x \partial v} - \frac{\mu(t)}{2} \frac{\partial^2 F}{\partial v^2} = 0, \quad (2.3.5)$$

$$W(x, v, t) = \frac{1}{F(x, v, t)} \int w \Phi(x, v, w, t) dw.$$

Используя связь

$$f(x, t) = \int F(x, v, t) dv$$

и учитывая обращение в ноль ВПФР всех порядков на границе области интегрирования, получаем из (2.3.5) интегрированием по скорости одномерное уравнение Фоккера-Планка:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x}(uf) - \frac{\lambda(t)}{2} \frac{\partial^2 f}{\partial x^2} = 0, \quad u(x,t) = \frac{1}{f(x,t)} \int v F(x,v,t) dv. \quad (2.3.6)$$

Подчеркнем, что в уравнении (2.3.6) коэффициент λ должен определяться по элементам ряда в соответствии с уравнениями эволюции моментов в силу этого самого уравнения. Важно учитывать, что, хотя одномерное уравнение (2.3.6) не содержит в явном виде некоторых коэффициентов, которые входят в двумерное уравнение (2.3.5), эти коэффициенты могут влиять на эволюцию моментов распределения $F(x,v,t)$ по скоростям.

Если мы обрываем цепочку совместных распределений на бинарной ВПФР $F(x,v,t)$, то среднее ускорение $W(x,v,t)$ в (2.3.5) следует задать независимо от трехмерной ВПФР. Это можно сделать в духе наивного прогноза, считая ускорение известным с предыдущего шага по времени (производная по времени трактуется в разностном смысле): $W(x,v,t) = W(x,v,t-1)$. Это же замечание относится и к коэффициентам λ, χ, μ . Если окажется, что такое приближение слишком грубо, то и обрывать на двумерной ВПФР нельзя, а требуется использовать аналогичное уравнение для $\Phi(x,v,w,t)$, и тогда моменты четырехмерного распределения считать известными, и т.д.

Подчеркнем, что нашей целью является сведение задачи к одномерному уравнению (2.3.6), а двумерное уравнение (2.3.5) будет использоваться для обоснования метода замыкания моментной системы.

Рассмотрим второй центральный выборочный момент распределения:

$$\sigma^2(t) = \int (x - m(t))^2 f(x,t) dx, \quad m(t) = \int x f(x,t) dx.$$

В соответствии с (2.3.6) имеем

$$\begin{aligned} \frac{d\sigma^2}{dt} &= - \int (x - m(t))^2 \frac{\partial(uf)}{\partial x} dx + \frac{\lambda}{2} \int (x - m(t))^2 \frac{\partial^2 f}{\partial x^2} dx = \\ &= \lambda + 2 \int u x f dx - 2mU, \quad U = U(t) = \int u(x,t) f(x,t) dx. \end{aligned}$$

Поскольку ковариация координаты и скорости равна

$$\begin{aligned} R(t) &= \text{cov}_{x,v}(t) = \int x v F(x,v,t) dx dv - \int x F(x,v,t) dx dv \cdot \int v F(x,v,t) dx dv = \\ &= \int x u(x,t) f(x,t) dx - m(t)U(t) \equiv \text{cov}_{x,u}(t), \end{aligned}$$

то эволюция выборочной дисперсии в силу уравнения Фоккера-Планка имеет вид

$$\frac{d\sigma^2}{dt} = \lambda + 2R. \quad (2.3.17)$$

Тем самым выявлен статистический смысл параметра $\lambda(t)$ в уравнении (2.3.6). Это есть разность между производной текущей выборочной дисперсии

ряда и удвоенной выборочной ковариацией исходного ряда и ряда его первых разностей.

Пусть имеется набор равномерно распределенных рядов $\{y_k\}_j, j=1, \dots, S$ длиной N . Тогда из них можно получить пучок из S траекторий, ассоциированных с двумя ВПФР: $f_N(x, t)$ и $f_N(x, t + N)$, согласно наблюдаемой эволюции этих распределений. Каждая j -ая траектория из набора траекторий построенного пучка порождает на отрезке $[t_0 + 1; t_0 + N]$ ВПФР $\tilde{f}_N(\{y\}_j; x, t_0 + N)$, отличную, вообще говоря, от наблюдаемой $f_N(x, t_0 + N)$. Однако по построению все эти выборочные траектории являются реализациями одного и того же нестационарного распределения вероятностей.

По совокупности сгенерированных траекторий можно оценить, насколько значимо отклонение модельного и фактического распределений. Используем для этого расстояние между функциями распределения в норме С:

$$\rho = \left\| \tilde{F}_N(\{y\}; x, t_0 + N) - F_N(x, t_0 + N) \right\|. \quad (2.4.2)$$

Рассмотрим также все попарные расстояния между ВПФР для сгенерированных траекторий

$$\tilde{\rho} = \left\| \tilde{F}_N(\{y\}; x, t_0 + N) - \tilde{F}_N(\{y'\}; x, t_0 + N) \right\|. \quad (2.4.3)$$

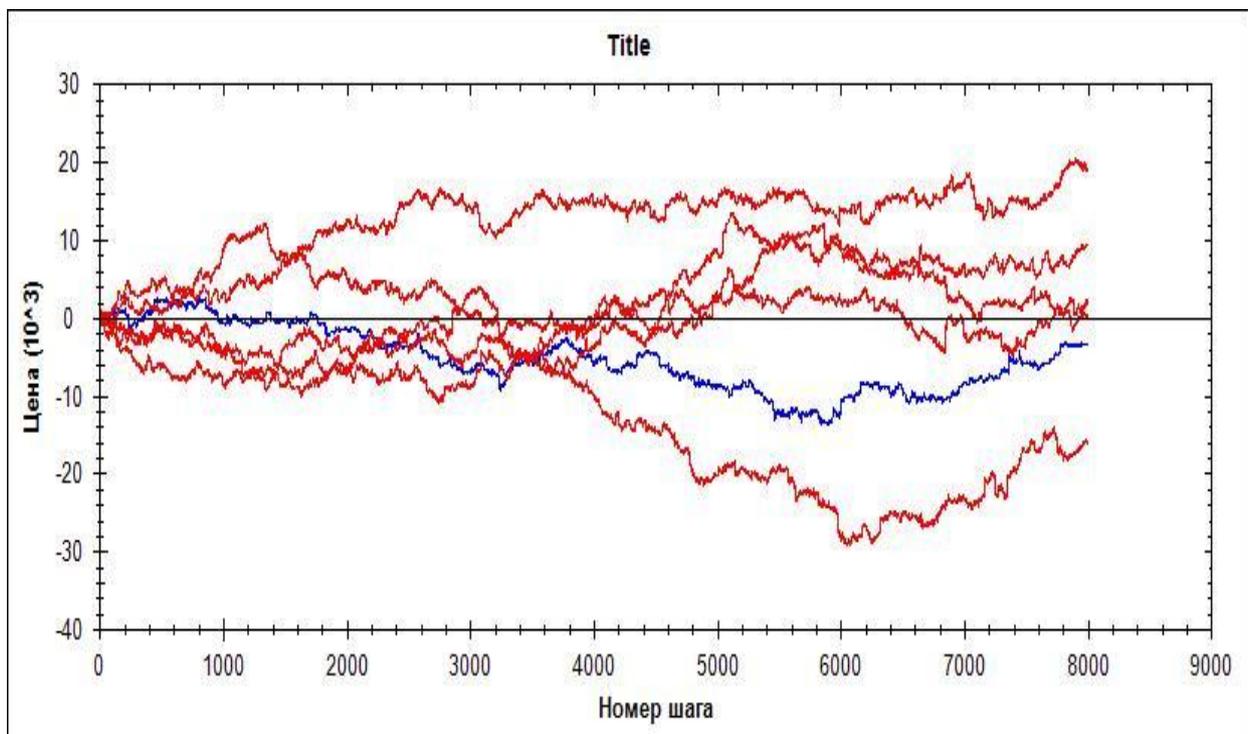


Рис. 13. Ансамбль траекторий в окне $N = 8000$ для решений уравнения Фоккера-Планка

В результате моделирования должен получаться ансамбль траекторий временного ряда, который обладает следующими двумя свойствами. Во-первых, СУС расстояний (2.4.2) приближенно равен СУС расстояний (2.4.3).

Это необходимо для того, чтобы отклонение сгенерированной траектории от фактической находилось на уровне статистической неопределенности, характерной для данного ряда. Во-вторых, если сравнить сгенерированные выборки в окне $[t_0 + 1; t_0 + N]$ с исходной ВФР $F_N(x, t_0)$, то соответствующий СУС должен быть приблизительно равным $\rho^*(N)$ в соответствии с формулами (2.1.6), (2.1.7) для встык-выборок. Пример такого ансамбля как решения уравнения Фоккера-Планка для эволюции выборочного распределения приростов биржевого индекса РТС приведен на рис. 13.

Подчеркнем, что использование уравнения Фоккера-Планка позволило, во-первых, сгенерировать ряд с теми статистическими свойствами его выборочного распределения, которые для него характерны, и, во-вторых, спрогнозировать выборочное распределение в среднем заметно точнее, чем наивный прогноз, в рамках которого распределение считается стационарным.

Результаты второй главы опубликованы в [1, 2, 5, 6].

В третьей главе приведены вычислительные алгоритмы, применяемые для нахождения индикаторов нестационарности, решения уравнения Фоккера-Планка, генерации ансамбля траекторий и тестирования заданных на этих траекториях функционалов. Описана также структура программного комплекса, реализующего разработанные алгоритмы.

В каждой ячейке разбиения, где существуют значения временного ряда, эмпирическая скорость определяется по формуле

$$u(i+1, t) f(i+1, t-1) = \sum_{k=1}^i (f(k, t) - f(k, t-1)) = F(i, t) - F(i, t-1). \quad (3.3.3)$$

Таким образом, правая часть (3.3.3) представляет собой изменение по времени выборочной функции распределения в ячейке, предшествующей той, в которой вычисляется значение эмпирической скорости.

Если же в $i+1$ -ой ячейке плотность распределения $f(i+1, t-1)$ оказалась равной нулю, то формально скорость в такой ячейке не определена. В фиктивных граничных ячейках скорость естественно положить равной нулю. Во внутренних ячейках с нулевой плотностью скорость определяем по линейной интерполяции между окаймляющими ячейками, плотность в которых отлична от нуля.

Оценка величины $\lambda(t)$ по выборке в соответствии с дискретным аналогом дисперсии имеет вид

$$\lambda(t) = \frac{1}{T} \sum_{k=t-T+1}^t (x(k) - x(k+1))^2 - \frac{1}{T^2} (x(t+1) - x(t-T+1))^2 \quad (3.3.4)$$

и при $T > 1$ эта величина строго положительна, т.е. модельный диффузионный процесс является регулярным.

Для повышения устойчивости решения уравнения Фоккера-Планка далее используется численная схема, когда аппроксимация второй производной делается в лево-разностном шаблоне, в котором значение функции в ячейке x

берется со следующего шага по времени. Описанная процедура приводит к разностному уравнению:

$$f(x, t+1) = f(x, t) + \frac{f(x, t)u(x, t-1) - f(x+1, t)u(x+1, t-1)}{h} + \\ + \frac{\lambda(t-1)}{2h^2} (2f(x-1, t) - f(x, t+1) - f(x-2, t)).$$

Разрешая его относительно $f(x, t+1)$, получаем схему расчета:

$$f(x, t+1) = f_T(x, t) \frac{f(x, t)}{1 + \lambda(t-1)/2h^2} + \frac{f(x, t)u(x, t) - f_T(x+1, t)u(x+1, t)}{h + \lambda(t-1)/2h} + \\ + \frac{\lambda(t-1)}{\lambda(t-1) + 2h^2} (2f(x-1, t) - f(x, t+1) - f(x-2, t)). \quad (3.3.5)$$

Решение уравнения Фоккера-Планка по формуле (3.3.5) осуществляется на такой горизонт τ , внутри которого отклонение прогнозной ВПФР от опорной (известной в начальный момент времени) не превосходит величины $\tau \rho^* / T$. Если предельное отклонение достигнуто, то в этот момент вычисленная ВПФР принимается за новое начальное распределение, для которого по выборке длины T находится новая скорость $u(x, t)$ и параметр диффузии $\lambda(t)$, вычисляемые уже не по элементам ряда, а непосредственно по имеющимся функциям распределения, включая фрагмент прогнозной ВПФР в скользящем окне ширины T .

Далее приводится алгоритм генерации ансамбля траекторий в соответствии с решением эмпирического кинетического уравнения. Из решения уравнения Фоккера-Планка относительно ВПФР находится прогнозная ВФР на заданном промежутке T , известная в каждый момент времени от t_0 (исходная ВПФР) до $t_0 + T$. Она имеет вид

$$F(x, t) = (nx - j) \cdot f_{j+1}(t) + \sum_{k=1}^j f_k(t), \quad x \in [(j-1)/n; j/n], \quad j = 1 \div n. \quad (3.4.1)$$

Затем генерируются S выборок длины T из равномерного распределения на $[0;1]$. Обозначим их $\{y_i^s\}$, $i = 1, \dots, T$; $s = 1, \dots, S$. После этого в каждый момент времени $t_k = t_0 + k$, $0 < k \leq T$ решается уравнение относительно случайной величины x :

$$F(x, t_k) = y_k^s. \quad (3.4.2)$$

В силу строгой монотонности ВФР $F(x, t_k)$ классовый интервал Δx_j , содержащий соответствующее решение для x_k^s , единственный. Он определяется из условий

$$F\left(\frac{j-1}{n}, t_k\right) \leq y_k^s, \quad F\left(\frac{j}{n}, t_k\right) > y_k^s. \quad (3.4.3)$$

Для найденного интервала Δx_j решение уравнения (3.4.2) имеет вид

$$x_k^s = \frac{1}{n} \left(\frac{y_k^s - \sum_{i=1}^j f_i(t_k)}{f_{j+1}(t_k)} + j \right), \quad (3.4.4)$$

если $f_{j+1}(t_k) > 0$. Если же $f_{j+1}(t_k) = 0$, то в качестве x_k^s выбирается центр классового интервала Δx_j .

Совокупность описанных алгоритмов объединена в программный комплекс, блок-схема которого представлена на рис. 14.

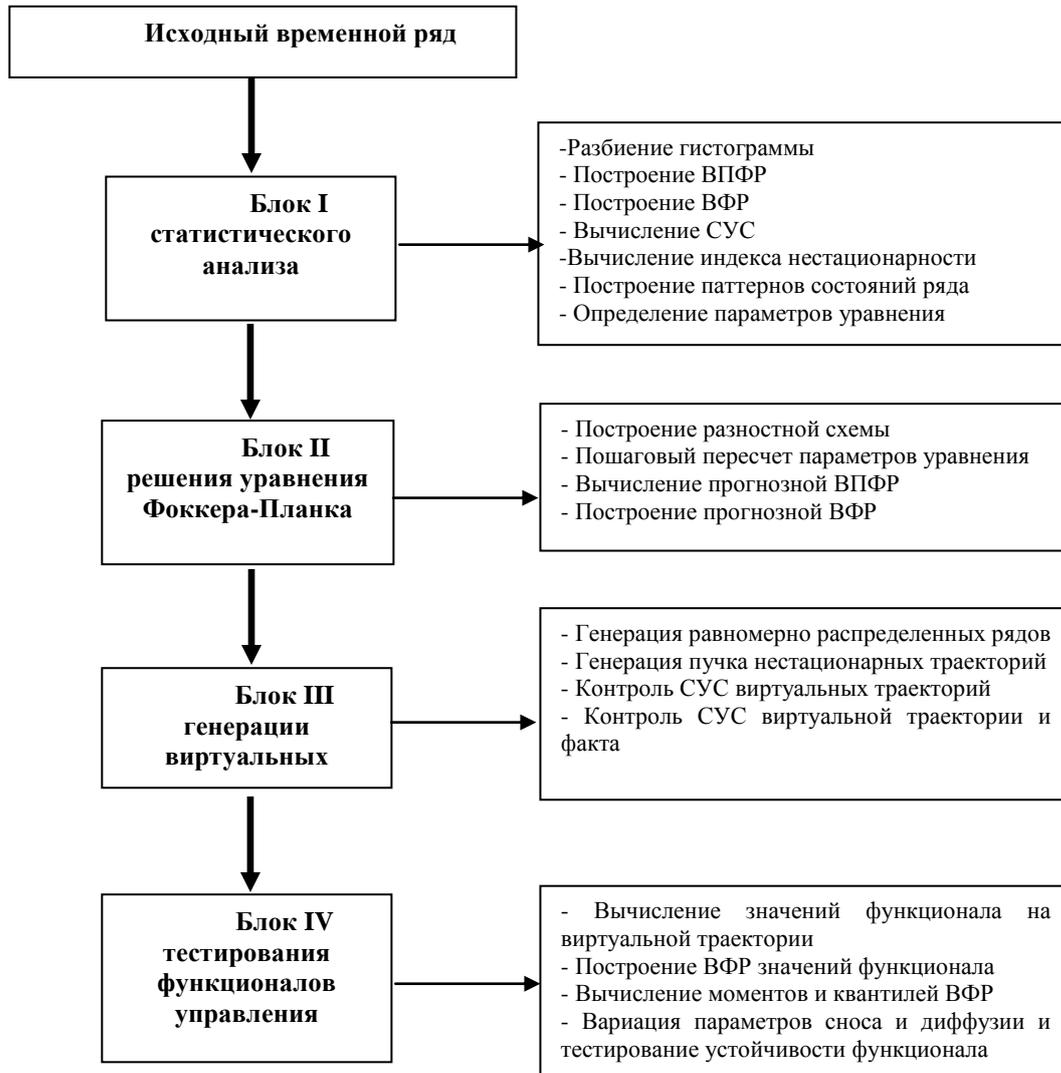


Рис. 14 – Блок-схема программного комплекса

Архитектура программного комплекса следующая. Главным объектом для работы является абстрактный класс `Developer`. Каждый потомок класса `Developer` реализует ряд методов для визуализации и записи информации. Для обращения к объектам `Developer` используется шаблон `Singleton`. Класс `OS` содержит в себе список всех объектов `Developer`, доступных пользователю. Возможность визуализации обеспечивается переопределением абстрактных функций класса `Developer`: `ToGraph()`, `Write()`, `Show()`.

Вычислительный алгоритм написан на языке `C#`.

Для работы программы необходима операционная система не ниже `Windows XP`, установка программы требует `30 Mb` свободного места на диске.

Для написания собственных функционалов необходима среда разработки `Visual Studio 2010` или более новая версия.

Для целей практического использования разработан интерфейс, позволяющий управлять решением блоков задач. Скрин-шот рабочего окна программы представлен на рис. 21. Кратко опишем назначение соответствующих окон меню.

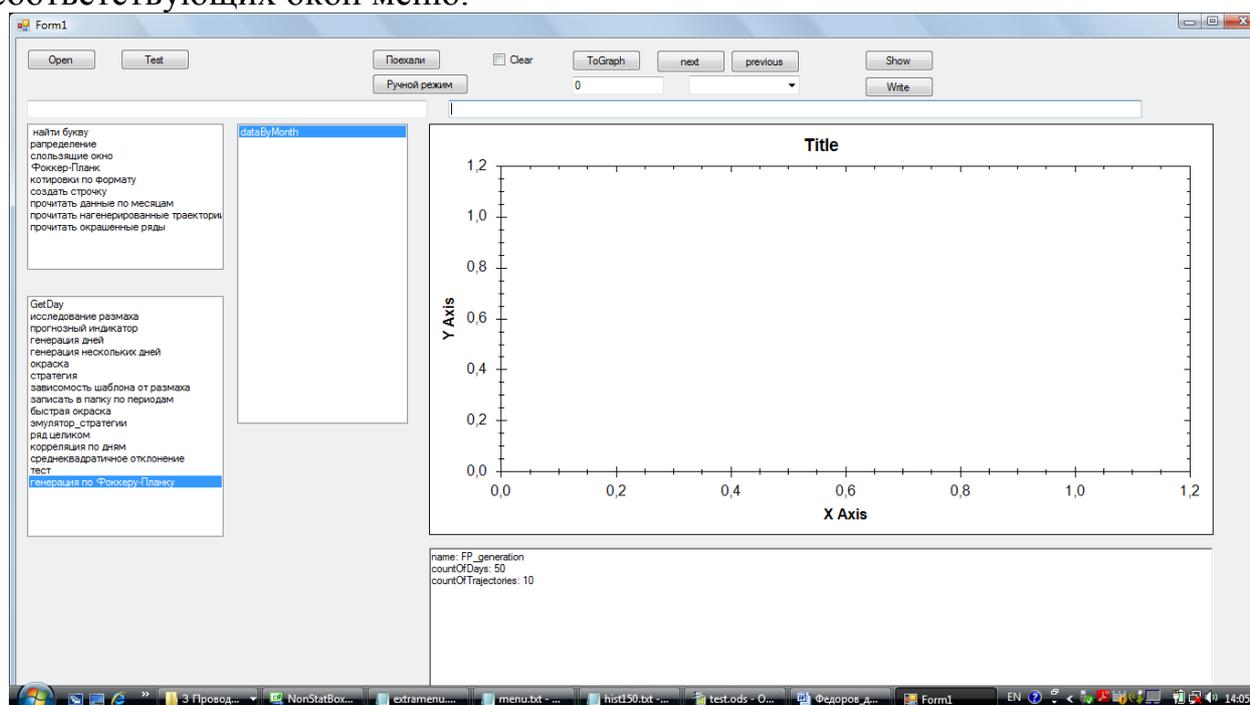


Рис. 21. Скрин-шот рабочего окна программы

Кнопка «Open» открывает файл в текстовом формате из папки, путь к которой прописывается в строке окна, расположенного под указанной кнопкой.

Кнопка «Test» обеспечивает быстрый доступ к файлу – либо последнему, который использовался, либо заданному специальной командой в коде программы.

Кнопка «Ручной режим» осуществляет запуск расчета для конкретного исследования – последнего или специально прописанного в коде.

Верхнее левое поле рабочего окна содержит первичные функции, которые можно выполнять с выбранным файлом данных.

Нижнее левое поле рабочего окна содержит функции, которые можно выполнять с результатами расчетов.

В поле окна справа отображается список проведенных исследований.

Поле с разметкой осей графика предназначено для визуализации результатов расчетов.

Нижнее поле под графиком предназначено для ввода-вывода числовых или текстовых параметров исследования, которое выбирается в поле первичных функций.

Кнопка «Поехали» осуществляет запуск расчета для задания, выбранного в любом из двух левых полей окна (поля не конфликтуют между собой), и после выбора параметра в третьем (нижнем) поле.

Кнопка «Clear» - флаг очистки поля графика при необходимости визуализации других графиков.

Кнопка «ToGraph» дает команду выведения графика или группы графиков на экран в поле визуализации.

Кнопки «next/previous» обеспечивают переход к визуализации следующего или предыдущего графика в выбранной серии графиков.

Кнопка «Show» является командой вывода в нижнее поле под графиком числовых результатов исследования.

Кнопка «Write» осуществляет запись результатов исследования в папку, путь к которой указан в строке над полем графика, а название файла совпадает с названием исследования, выбранного в поле списка проведенных исследований.

Два окна над полем графика отвечают номеру (справа) и цвету (слева) выбираемого графика в серии.

Результаты третьей главы опубликованы в [1].

В четвертой главе приведен ряд практических результатов, полученных с применением разработанного метода. Эти результаты относятся к анализу биржевых рядов и связанных с ними торговых систем, к определению однородности текста и языка написания зашифрованного сообщения в задачах математической лингвистики, а также к определению уровня значимости оценки региональной сейсмической активности.

Для проверки корректности работы вычислительных алгоритмов, описанных в главе III, в данном разделе приводятся примеры построения прогнозной ВПФР по уравнению Фоккера-Планка, коэффициенты сноса и диффузии в котором определяются с предыдущих шагов по времени по выборке, длина которой равна горизонту прогнозирования в соответствии с методикой, изложенной выше.

Прогнозная ВПФР, обозначаемая $\tilde{f}_T(x, t_0 + T)$, должна обладать тем свойством, что отклонение $\tilde{f}_T(x, t_0 + T)$ от ВПФР начального состояния $f_T(x, t_0)$ в выбранной норме должно быть порядка СУС между встык-

выборками. При этом желательно, чтобы ее отклонение от фактически состоявшейся ВПФР $f_T(x, t_0 + T)$ не превосходило бы СУС.

На рис. 22 показаны две типовые выборочные плотности распределения для встык-выборок ряда приростов цен закрытия 1-минутных интервалов для индекса РТС за неделю, а на рис. 23 – прогнозная и фактическая ВПФР.

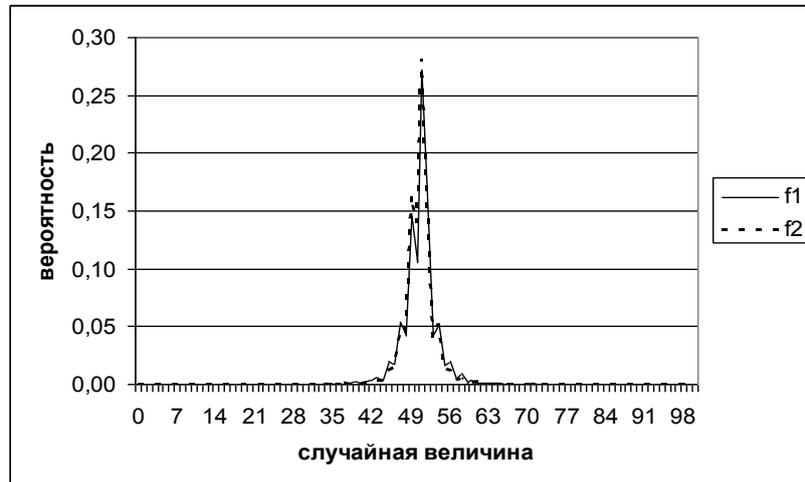


Рис. 22. Пример выборочных плотностей в соседних окнах длиной $T = 3000$

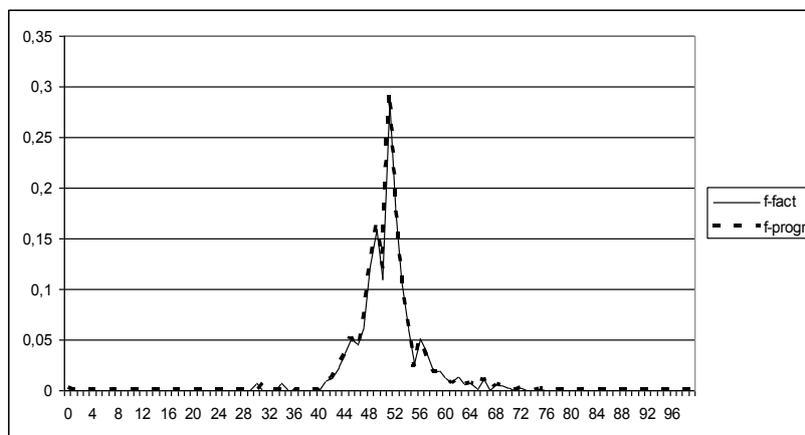


Рис. 23. Пример сравнения фактической и прогнозной ВПФР
в окне длиной $T = 3000$

Рассматриваемому промежутку отвечает длина выборки 3000, ВПФР имеет индекс нестационарности, равный в рассматриваемом примере 3,2. Анализировались данные за 2015 г., всего 400 тыс. значений.

Область изменения случайной величины (нормированных на $[0;1]$ приростов цен) на рис. 22-23 разбита на 100 классовых интервалов, по оси абсцисс отложен номер соответствующего интервала. Несмотря на то, что визуальное распределения представляются весьма близкими, расстояние между ними в норме C равно $\rho = \|F_T(x, t_0) - F_T(x, t_0 + T)\| = 0,13$ при том, что в стационарном случае для выборок представленных длин это расстояние должно

быть порядка 0,04. На рис. 23 показаны две ВПФР для того же ряда, только вместо фактической $f_T(x, t_0) \equiv f_1$ представлена прогнозная $\tilde{f}_T(x, t_0 + T)$, которая сравнивается с фактической $f_T(x, t_0 + T) \equiv f_2$. Расстояние между ними равно 0,06, что примерно в 1,5 раза больше, чем стационарный уровень шума и в два раза меньше, чем расстояние между встык-выборками. Это показывает, что кинетическое уравнение, рассматриваемое как модель изменения ВПФР, дает качественно адекватные результаты прогнозирования.

Отметим, что если не учитывать диффузионный член в уравнении Фоккера-Планка, то расстояние между прогнозом и фактом возрастает почти в два раза до 0,10. При этом основные отличия обусловлены различиями ВПФР в точках локальных экстремумов, когда прогнозные пики вверх по уравнению Лиувилля оказываются выше, а прогнозные минимумы – напротив, ниже, чем сглаженные положения экстремумов с учетом диффузионного члена.

Таким образом, кинетическая модель прогнозирования ВПФР по уравнению Фоккера-Планка дает более точные результаты, чем чисто механическая модель сноса по уравнению Лиувилля, а также значительно более точные, чем модель стационарного приближения.

Другим примером применения программного комплекса является анализ распределения землетрясений, происходящих в определенном регионе, по магнитуде и частоте событий. Актуальным является вопрос о том, насколько стационарны данные наблюдений. Практическая необходимость такого знания состоит, например, в том, что если распределение магнитуд стационарно, то на заданном уровне значимости можно определить сейсмическую опасность и установить нужные СНИП при строительстве тех или иных сооружений. Если же ряд магнитуд нестационарный, то, во-первых, следует определить, на каком фактически уровне значимости следует принимать соответствующие решения, а также на каком горизонте прогнозирования будут справедливы прошлые статистики. Анализировались данные Мирового каталога землетрясений по Японии, начиная с 1900 г. Это магнитуды событий не менее 4 баллов. В указанном Каталоге можно обнаружить утверждение о том, что вероятности магнитуд определены с точностью не хуже 0,025. Однако, построив зависимость СУС выборочных распределений указанных случайных величин в зависимости от длины выборки, обнаруживаем (рис. 41), что наилучшая точность по магнитуде равна 0,10 на выборках примерно 800-1200 данных (2,5-3,5 года), а по промежуткам времени между событиями точность составляет 0,15 на выборках 5-7 тыс. данных (16-24 года).

Результаты рис. 41 показывают, что анализируемые ряды являются сильно нестационарными, по магнитуде наилучшая статистика собирается 2,5 года и существует примерно год, после чего должна быть пересчитана. Если исходить из более реальных требований частоты пересчета, связанных со строительством зданий (например, один раз за 25 лет), то наилучшим уровнем значимости будет 0,15.

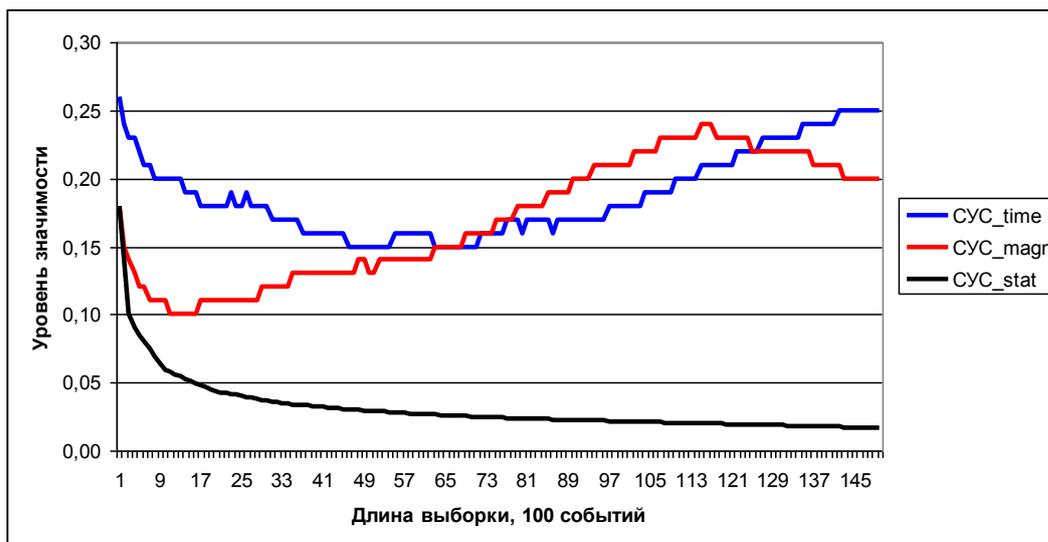


Рис. 41. СУС распределений в зависимости от длины выборки

Результаты четвертой главы приведены в [3, 4, 5, 7, 9, 10].

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В настоящей диссертации разработан новый метод непараметрического моделирования нестационарных временных рядов и построен программный комплекс, реализующий соответствующие численные алгоритмы.

Моделирование ряда основано на решении эмпирического уравнения Фоккера-Планка, что позволяет задавать скользящее среднее, дисперсию, размах и аналогичные величины изменяющимися во времени в соответствии с требованиями статистического эксперимента, причем выборочные распределения остаются в классе распределений, характерных для данного ряда. Разработанный в диссертации метод позволяет тестировать индикаторы-предикторы изменения какого-либо свойства временного ряда и функционалы распознавания состояний ряда в широком диапазоне изменения его выборочных статистик. На практике для такого тестирования обычно используются исторические данные за достаточно большой промежуток времени, но надо помнить, что при этом берется единственная реализация случайного процесса.

К достоинствам метода следует отнести и то, что он позволяет провести стресс-тест на работоспособность индикатора в пределах, контролируемых исследователем. Исторический же ряд данных не предоставляет таких возможностей. Кроме того, для квалифицированного тестирования ряд прошлых данных требует предварительного выявления интересных ситуаций, кластеризации их, определения ошибок при кластеризации, что весьма трудоемко и не дает полного представления об имеющихся локальных паттернах ряда. Таким образом, численный код, генерирующий по фрагменту траектории нестационарного ряда ансамбль его нестационарных же реализаций, представляет практическую важность.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях, рекомендованных ВАК:

1. Орлов Ю.Н., Федоров С.Л. Генерация нестационарных траекторий временного ряда на основе уравнения Фоккера-Планка // Труды МФТИ, 2016. Т. 8. № 2. С. 126-133.
2. Ключкова Л.В., Орлов Ю.Н., Федоров С.Л. Моделирование ансамбля нестационарных траекторий с помощью уравнения Фоккера-Планка // Журнал Средневолжского математического общества, 2016. – Т.18. № 1.
3. Кирина-Лилинская Е.П., Орлов Ю.Н., Федоров С.Л. Метод базисных паттернов в анализе нестационарных временных рядов // Препринты ИПМ им. М.В. Келдыша. 2016. № 7. 20 с.
4. Арутюнов А.А., Борисов Л.А., Зенюк Д.А., Ивченко А.Ю., Кирина-Лилинская Е.П., Орлов Ю.Н., Осминин К.П., Федоров С.Л., Шилин С.А. Статистические закономерности европейских языков и анализ рукописи Войнича // Препринты ИПМ им. М.В. Келдыша. 2016. № 52. 36 с.
5. Орлов Ю.Н., Федоров С.Л. Моделирование распределений функционалов на ансамбле траекторий нестационарного случайного процесса // Препринты ИПМ им. М.В. Келдыша. 2016. № 101. 14 с.

Прочие публикации:

6. Орлов Ю.Н., Федоров С.Л. Моделирование и статистический анализ функционалов, заданных на выборках из нестационарного временного ряда // Препринты ИПМ им. М.В. Келдыша. 2014. № 43. 26 с.
7. Орлов Ю.Н., Федоров С.Л., Давидько В.А. К вопросу классификации нестационарных временных рядов: состав индекса РТС // Препринты ИПМ им. М.В. Келдыша. 2014. № 54. 18с.
8. Босов А.Д., Орлов Ю.Н., Федоров С.Л. О распределении рядов абсолютных приростов цен на финансовых рынках // Препринты ИПМ им. М.В. Келдыша. 2014. № 96. 15 с.

Материалы научных конференций:

9. Федоров С.Л. Анализ функционалов, заданных на выборках из нестационарного временного ряда. // Труды II Международной научно-практической конференции Теоретические и прикладные аспекты современной науки, Белгород, август 2014. С. 9-16.
10. Yu. Orlov, S. Fedorov, A. Samouylov, Yu. Gaidamaka, D. Molchanov. Simulation of Devices Mobility to Estimate Wireless Channel Quality Metrics in 5G Network // Proc. 14-th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 19-25, 2016, Rhodes, Greece.