УДК 51.7:532.546

Моделирование ансамбля нестационарных траекторий с помощью уравнения Фоккера-Планка

© **Л. В. Клочкова[1], Ю. Н. Орлов[2], С. А. Федоров[3]**

Аннотация. Рассматривается кинетическое уравнение Фоккера-Планка для выборочной плотности функции распределения случайных величин, наблюдаемых на практике в виде временных рядов. Вместо изучения одной из возможных реализаций временного ряда предлагается рассмотреть ансамбль случайных траекторий, порождаемых эмпирической функцией распределения. Предлагается модель для описания изменения во времени функционалов, заданных на случайных траекториях и имеющих практическое значение. Это, например, индикатор уровня загрязнения мегаполиса в виде средней концентрации вредных веществ за определенный период времени, индикатор изменения эпидемиологической обстановки в регионе, функционал эффективности управления загрязнением в виде снижения уровня загрязнения в результате определенных действий и т.п. Формулируется метод тестирования управляющего функционала на нестационарном ансамбле траекторий.

Ключевые слова: выборочная функция распределения, уравнение Фоккера-Планка, ансамбль нестационарных траекторий, тестирование управляющего функционала.

1. Введение

Настоящая работа продолжает исследования по моделированию и прогнозированию временных рядов с помощью кинетических уравнений типа Лиувилля и Фоккера-Планка, примененных к выборочным функциям распределения [1, 2]. В указанных работах был построен метод прогнозирования функции распределения случайной величины с нестационарным поведением на заданный горизонт вперед. Естественным развитием кинетического подхода применительно к временному ряду является генерация ансамбля возможных траекторий, которые отвечают выявленным тенденциям в изменении выборочных функций распределения в силу того или иного кинетического уравнения.

В работах [3, 4] была построена методика построения кинетических уравнений применительно к практической задаче прогнозирования загрязненности атмосферы мегаполисов, а также распространения инфекций или вредных примесей в случайно-неоднородной и нестационарной среде. В них были построены статистики, которые требуется вычислить по данным временного ряда, чтобы определить параметры кинетического уравнения для выборочной функции распределения этого ряда. Здесь мы описываем, как можно было бы не только прогнозировать коридор возможного изменения случайной величины, но и тестировать эффективность того или иного управляющего воздействия, заданного в виде функционала на фрагменте случайной траектории. В результате модели для описания эволюции распределения случайных параметров, характеризующих интенсивность источника вредных примесей, приобретают дополнительную актуальность.

Практическая трудность использования детерминистических, а не стохастических уравнений для моделирования концентрации загрязнения мегаполисов в зависимости от

[1] Старший научный сотрудник Института прикладной математики им. М.В.Келдыша РАН, г. Москва; klud@imamod.ru.

[2] Ведущий научный сотрудник Института прикладной математики им. М.В.Келдыша РАН, г. Москва; ov3159fd@yandex.ru.

[3] Аспирант Института прикладной математики им. М.В.Келдыша РАН, г. Москва; klud@imamod.ru.

случайных условий водно-воздушной среды состоит в том, что источник загрязнения по интенсивности и составу является случайным и притом нестационарным. В результате, кинетические уравнения, применяемые в условиях неопределенности пространственного распределения примесей, приобретают дополнительные стохастические свойства из-за неопределенности функции источника. Возможность описать эту неопределенность кинетическим уравнением того же типа, что и среду, в которой осуществляется перенос изучаемого фактора, позволяет построить унифицированную кинетическую модель процесса в целом. Построенное численное решение относительно прогнозной функции распределения можно использовать для создания алгоритма генерации пучка нестационарных траекторий.

Необходимость тестирования функционала управления, заданного на траектории нестационарного случайного процесса, часто возникает при построении алгоритма распознавания, к которой сводятся многие задачи прикладного статистического анализа [5]. Часто оказывается, что когда изучаемая система, представляемая в виде временного ряда, находится в том или ином состоянии, измеряемые значения случайной величины, через которые и проявляется это состояние, имеют характерные именно для этого состояния функции распределения. Тогда идентификация состояния формулируется как задача распознавания выборочной функции распределения. Задача распознавания выборки как принадлежащей определенной генеральной совокупности решается в статистике либо путем оценивания значений параметров распределения известного функционального вида, либо в рамках непараметрического подхода, когда используется критерий Колмогорова-Смирнова. Однако применение классических критериев корректно только в стационарном случае, когда есть оценки скорости сходимости выборочных распределений к генеральной совокупности. Если функция распределения нестационарна, то обучение алгоритма распознавания на данных за прошлый период часто оказывается несостоятельным. В таком случае для более надежного распознавания надо тестировать решающую функцию на нестационарных временных рядах. Но фактически для тестирования существует только одна траектория, которая в силу нестационарности не позволяет использовать достаточно большой объем выборки. В результате возникает необходимость тестирования тех или иных индикаторов локального поведения временного ряда с целью оценки вероятности их правильного срабатывания.

Индикаторы представляют собой определенные функционалы, заданные на фрагментах траектории случайного процесса. Чтобы оценить эмпирическую условную вероятность того, что определенный интервал значений индикатора отвечает ожидаемому поведению ряда в настоящем или будущем, нужно иметь много реализаций изучаемого процесса. Для этого требуется сгенерировать пучок возможных траекторий временного ряда, выборочная функция распределения которого эволюционирует определенным образом, и проверить на нем устойчивость срабатывания индикатора.

В настоящей работе в качестве модельного уравнения для описания эволюции нестационарных распределений используется уравнение Фоккера-Планка относительно выборочной плотности функции распределения (далее ВПФР) временного ряда. Подход к моделированию нестационарных траекторий на основе решения кинетического уравнения относительно ВПФР был предложен в [6], где в качестве такого уравнения использовалось уравнение Лиувилля. В работе [7] было обосновано уравнение Фоккера-Планка для ВПФР нестационарного временного ряда. Тем самым стало возможным корректное моделирование ансамбля траекторий временного ряда с помощью уравнения типа диффузии со сносом.

2. Метод генерации нестационарных траекторий

Будем для удобства нормировки считать, что изучаемая случайная величина равномерно ограничена по времени, так что все ее значения принадлежат отрезку $[0;1]$.

Пусть $x(t)$ есть значение случайной величины в дискретный момент времени t, где шаг по времени считается единичным, а $f_T(x,t)$ есть ВПФР выборки длины T с окончанием в момент времени t, т.е. выборки фрагмента ряда $\{x(t-T+1),\ldots,x(t)\}$. Выборочная плотность строится по равномерному разбиению гистограммы в соответствии с методикой, описанной в [8], при котором статистическая неопределенность в частотах равна точности, с которой значения ряда различаются одно от другого, т.е. $1/n$, где есть число классовых интервалов.

Число классовых интервалов n в зависимости от длины выборки T и конкретной получающейся формы распределения определяется численно как решение уравнения

$$\frac{\sqrt{T}}{\sum\limits_{i=1}^{n}\sqrt{f_T(i)(1-f_T(i))}}=nt_{1-1/(2n)}, \tag{2.1}$$

где $t_{1-1/(2n)}$ есть квантиль распределения Стьюдента порядка $1-1/(2n)$ с $T-1$ степенью свободы. По кусочно-непрерывной гистограмме ВПФР $f_T(x,t)$ можно построить непрерывную функцию распределения $F_T(x,t)$:

$$F_T(x,t)=\int\limits_0^1 f_T(y,t)dy.$$

Поскольку гистограмма ВПФР имеет вид

$$f(x)=f_j, \quad x\in[(j-1)/n;j/n], \quad j=1\div n, \tag{2.2}$$

то соответствующая ВФР определяется формулой

$$F(x)=(nx-j)\cdot f_{j+1}+\sum_{k=1}^{j}f(k), \quad x\in[(j-1)/n;j/n], \quad j=1\div n. \tag{2.3}$$

Чтобы имитировать процесс, близкий к реальным наблюдениям, предлагается следующая схема действий. На первом этапе по имеющимся историческим данным строятся выборочные распределения значений $x(t)$ изучаемого случайного параметра за тот промежуток времени T, который представляет интерес. Таким образом, в каждый момент времени t определена ВПФР $f_T(x,t)$ (2.2) и соответствующая ей $F_T(x,t)$ согласно (2.3). Затем генерируется стационарный равномерно распределенный на $[0;1]$ ряд чисел $\{y_k\}$ длиной T. Пусть t_0 есть начальный момент времени, в который ВПФР $f_T(x,t_0)$ известна. 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:

$$y_k=F_T(x_k,t_0+k). \tag{2.4}$$

Подчеркнем, что, согласно (2.4), 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 $\{y_k\}_j, j = 1, \cdot, N$, можно получить пучок из N траекторий, ассоциированных с двумя ВПФР: $f_T(x,t)$ и $f_T(x,t+T)$, согласно наблюдаемой эволюции этих распределений. Каждая j-ая траектория из набора траекторий построенного пучка порождает на отрезке $[t_0+1; t_0+T]$ ВПФР $\tilde{f}_T(y_j; x, t_0+T)$, отличную от наблюдаемой $f_T(x, t_0+T)$. Однако по построению все эти выборочные траектории являются реализациями одного и того же нестационарного распределения вероятностей.

По совокупности сгенерированных траекторий можно оценить, насколько значимы отклонения модельного и фактического распределений. Используем для этого расстояние между функциями распределения в норме C

$$\rho = \|\tilde{F}_T(\{y\}; x, t_0+T) - F_T(x, t_0+T)\|. \tag{2.5}$$

Рассмотрим также все попарные расстояния между ВПФР для сгенерированных траекторий

$$\rho = \|\tilde{F}_T(\{y\}; x, t_0+T) - \tilde{F}_T(\{y'\}; x, t_0+T)\|. \tag{2.6}$$

Если бы распределения $F_T(x,t)$ были стационарны, то расстояния (2.6) подчинялись бы статистике Колмогорова–Смирнова: $P\{\rho\sqrt{T/2} < z\} \to K(z), T \to \infty$, где $K(z)$ — функция Колмогорова. В нашем случае этот критерий неприменим из-за нестационарности процесса, и для оценки близости между распределениями предлагается ввести специальный индикатор, называемый согласованным уровнем стационарности. Для его построения проанализируем статистику расстояний между так называемыми встык-выборками, т.е. между ВФР $F_T(x,t)$ и $F_T(x,t+T)$, сдвинутыми одна относительно другой на величину окна выборки:

$$\rho(T;t) = \|F_T(x,t) - F_T(x,t+T)\|. \tag{2.7}$$

По имеющимся историческим данным построим функцию распределения $G(\rho;T)$ расстояний (2.7), которая представляет эмпирическую вероятность того, что расстояние между распределениями не больше ρ. Определим далее согласованный уровень стационарности (СУС) $\rho^*(T)$ так, что соответствующее расстояние равно значимости критерия, т.е. является решением уравнения

$$G(\rho;T) = 1 - \rho. \tag{2.8}$$

В стационарном случае уравнение (2.8) переходит в уравнение

$$K\left(\varepsilon\sqrt{T/2}\right) = 1 - \varepsilon, \tag{2.9}$$

которое определяет функцию $\varepsilon(T)$, обладающую тем свойством, что при проведении бесконечно большого числа экспериментов по вычислению расстояний между двумя выборочными распределениями длины T в доле случаев ε будет наблюдаться превышение расстояния, равного ε. Если оказалось, что для некоторой длины выборки T отношение ρ^*/ε больше единицы, то ряд нестационарный. Величина

$$J(T) = \frac{\rho^*(T)}{\varepsilon(T)} \tag{2.10}$$

называется индексом нестационарности ряда [9]. Если же на некоторых длинах выборки величина $J(T) \leq 1$, то ряд стационарный, и тогда можно считать, что его выборочные распределения не эволюционируют. Для реального применения кинетического уравнения к описанию эволюции ВПФР следует определить такие длины, на которых $J(T) > 1$.

В результате моделирования по формуле (2.4) получается ансамбль траекторий временного ряда, который обладает следующими свойствами. Во-первых, СУС расстояний (2.5) приближенно равен СУС расстояний (2.6), т.е. характерное расстояние между виртуальными выборками равно характерному расстоянию между виртуальной и фактической траекториями. Во-вторых, если сравнить сгенерированные выборки в окне $[t_0+1; t_0+T]$ с исходной ВФР $F_T(x,t_0)$, то соответствующий СУС будет приблизительно равным $\rho^*(T)$ в соответствии с (2.8).

Описанный подход позволяет построить численный алгоритм моделирования нестационарного временного ряда с определенными непараметрическими свойствами его ВПФР, эволюционирующей в соответствии с уравнением Фоккера-Планка (или иным модельным уравнением, которое описывает эволюцию ВПФР рассматриваемого ряда).

3. Модель эволюции выборочной плотности функции распределения

Считаем, что анализ индекса нестационарности позволил определить длину T выборки, по которой строится ВПФР $f_T(x,t)$, так что далее эта длина фиксирована. Далее для краткости соответствующий индекс функций распределения опускаем. В том же окне длины T строим совместную плотность распределений $\Phi(x,v,t)$ значений временного ряда $x(t)$ и его приращений $v(t)=x(t+1)-x(t)$. При этом справедлива формула

$$f(x,t)=\int_{-1}^{1}\Phi(x,v,t)dv, \tag{3.1}$$

где в пределах интегрирования учтено, что $x\in[0;1]$, так что $v\in[-1;1]$. В качестве модельного уравнения эволюции используем уравнение Фоккера-Планка относительно ВПФР $f(x,t)$:

$$\frac{\partial f}{\partial t}+\frac{\partial}{\partial x}(uf)-\frac{\lambda}{2}\frac{\partial^2 f}{\partial x^2}=0, \tag{3.2}$$

где параметры сноса (средняя скорость $u(x,t)$) и диффузии $\lambda(t)$ в соответствии с [7] определяются формулами:

$$\begin{gathered} u(x,t)f(x,t)=\int v\Phi(x,v,t)dv; \\ \lambda(t)=\frac{d\sigma^2}{dt}-2\mathrm{cov}_{x,u}(t), \quad \sigma^2(t)=\int_0^1(x-\bar{x})^2 f(x,t)dx. \end{gathered} \tag{3.3}$$

В [7] доказано утверждение о том, что оценка величины λ по выборке длины T в соответствии с дискретными аналогами дисперсии и ковариации имеет вид

$$\lambda(t)=\frac{1}{T}\sum_{k=t-T+1}^{t}\left(x(k)-x(k+1)\right)^2-\frac{1}{T^2}\left(x(t+1)-x(t-T+1)\right)^2 \tag{3.4}$$

и при $T>1$ строго положительна. Параметр сноса $u(x,t)$ определяется согласно (3.3). В численной схеме указанные параметры берутся с предыдущего шага по времени. Тогда дискретная форма записи уравнения Фоккера-Планка (3.2) с явной разностной схемой для эволюции по времени с единичным шагом и шаблоном левой разностной производной по пространству с шагом $h = 1/(100n)$, где n определяется в (2.1), имеет вид

$$f_T(x,t+1) = f_T(x,t) + \frac{f_T(x,t)u(x,t-1) - f_T(x+1,t)u(x+1,t-1)}{h} + \\ + \frac{\lambda(t-1)}{2h^2}\left(f_T(x+2,t) - 2f_T(x+1,t) + f_T(x,t)\right). \tag{3.5}$$

Здесь, однако, следует учесть, что явные схемы при решении уравнений диффузионного типа неустойчивы, в связи с чем они имеют сравнительно малый горизонт прогнозирования. Для повышения устойчивости далее мы использовали схему, в которой, во-первых, каждый классовый интервал разбит еще на 100 ячеек (с учетом равномерности в них, по построению, выборочной плотности), и, во-вторых, 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 (для краткости нижний индекс T опущен):

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

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

$$f(x,t+1) = f_T(x,t)\frac{f_T(x,t)}{1+\lambda(t-1)/2h^2} + \\ + \frac{f_T(x,t)u(x,t-1) - f_T(x+1,t)u(x+1,t-1)}{h+\lambda(t-1)/2h} + \\ + \frac{\lambda(t-1)}{\lambda(t-1)+2h^2}\left(2f(x-1,t) - f(x,t+1) - f(x-2,t)\right). \tag{3.6}$$

Одновременно с пошаговым решением уравнения (3.2) по схеме (3.6), когда величины λ и $u(x,t)$ пересчитываются на каждом шаге по времени по формулам (3.3), строится и ансамбль соответствующих траекторий согласно методике п. 2.

4. Тестирование управляющего функционала

Рассмотрим некоторый функционал, заданный на случайной траектории. Его статистические свойства такие, как среднее, дисперсия, чувствительность по отношению к тем или иным параметрам, на практике могут быть изучены всего лишь по единственной реализации в виде конкретного временного ряда. С учетом нестационарности поведения ряда этого явно недостаточно, поскольку использование выборок большой длины может привести к ошибочным выводам, ибо анализ будет проводиться по не актуальным данным. Альтернативой является генерация набора траекторий, отвечающих текущим тенденциям анализируемой случайной величины. Статистический анализ управляющего функционала на таком ансамбле состоит в следующем [10].

Пусть на выборке длины T задан некоторый функционал $\psi\{x(t-T+1),\ldots,x(t)\}$. Это может быть, например, статистика в виде скользящей средней, а может быть и некоторая сложная конструкция в виде управления другой случайной траекторией. Последняя задача может быть востребована при анализе эффективности проведения мероприятий по улучшению экологической или эпидемиологической обстановки в регионе.

При тестировании функционала ψ требуется определить, во-первых, его статистические свойства на выборках, отвечающих данной модели эволюции ВПФР, и, во-вторых, изучить устойчивость функционала при изменении параметров уравнения эволюции.

Первая задача решается следующим образом. Пусть выбран интересующий нас фрагмент временного ряда и на нем построен пучок виртуальных траекторий числом N. Обозначим ψ_j значение функционала на j-ой траектории. Его статистические свойства полностью определяются выборочным распределением, которое строится по имеющимся N значениям на траекториях. В частности, можно определить среднее, дисперсию и относительное отклонение:

$$\bar{\psi}=\frac{1}{N}\sum_{j=1}^{N}\psi_j,\quad \sigma_\psi^2=\frac{1}{N}\sum_{j=1}^{N}\left(\psi_j-\bar{\psi}\right)^2,\quad S_\psi=\frac{\bar{\psi}}{\sigma_\psi}. \tag{4.1}$$

Формула (4.1) дает корректный ответ на вопрос, какова, например, средняя скорость восстановления нормальных экологических параметров системы на определенном промежутке времени.

Вторая задача решается посредством вариации параметров уравнения Фоккера-Планка, в результате которой тренд $u(x,t)$ и диффузия $\lambda(t)$ меняются определенным образом. Вычисляя статистику (4.1) функционала управления на новых траекториях, полученных в результате модификации указанных параметров, можно определить допустимые пределы, внутри которых управление устойчиво. Чувствительность функционала определяется как его логарифмическая производная по параметру, например:

$$\Lambda_\psi=\frac{\partial\ln\bar{\psi}}{\partial\ln\lambda}. \tag{4.2}$$

Задавая допустимые границы вариации (4.2), можно в численном эксперименте получить допустимые границы вариации параметров уравнения Фоккера-Планка, т.е. выяснить, предположим, при каком предельном коэффициенте диффузии эффективность управления (4.1) имеет положительное математическое ожидание.

Описанный метод позволяет тестировать индикаторы-предикторы изменения какого-либо свойства временного ряда и функционалы распознавания состояний ряда в широком диапазоне изменения его выборочных статистик. К его достоинствам следует отнести 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 данных за прошлый период 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.

Работа выполнена при поддержке гранта РФФИ, проект № 14–01–00145.

Список литературы

1. Орлов Ю. Н., Осминин К. П., "Построение выборочной функции распределения для прогнозирования нестационарного временного ряда", *Математическое моделирование*, 2008, № 9, 23–33.

2. Орлов Ю. Н., Осминин К. П., *Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков*, Эдиториал УРСС/Книжный дом "ЛИБРОКОМ", М., 2011, 384 с.

3. Клочкова Л. В., Орлов Ю. Н., Тишкин В. Ф., "Математическое моделирование корреляции эпидемической обстановки в мегаполисах от состояния воздуха", *Журнал Средневолжского математического общества*, **7** (2012), 34–43.

4. Зенюк Д. А.,Клочкова Л. В., Орлов Ю. Н., "Моделирование нестационарных случайных процессов кинетическими уравнениями с дробными производными", *Журнал Средневолжского математического общества*, **17**:1 (2016).

5. Вапник В. Н., Червоненкис А. Я., *Теория распознавания образов*, Наука, М., 1974, 416 с.

6. Босов А. Д., Кальметьев Р. Ш., Орлов Ю. Н., "Моделирование нестационарного временного ряда с заданными свойствами выборочного распределения", *Математическое моделирование*, 2014, № 3, 97–107.

7. Босов А. Д., Орлов Ю. Н., *Эмпирическое уравнение Фоккера-Планка для прогнозирования нестационарных временных рядов.*, Препринт № 3 ИПМ им. М. В. Келдыша РАН, ИПМ им. М. В. Келдыша РАН, Москва, 2013.

8. Орлов Ю. Н., *Оптимальное разбиение гистограммы для оценивания выборочной плотности распределения нестационарного временного ряда.*, Препринт № 14 ИПМ им. М. В. Келдыша РАН, ИПМ им. М. В. Келдыша РАН, Москва, 2013.

9. Орлов Ю. Н., *Кинетические методы исследования нестационарных временных рядов*, МФТИ, Москва, 2014, 276 с.

10. Орлов Ю. Н., Федоров С. Л., *Моделирование и статистический анализ функционалов, заданных на выборках из нестационарного временного ряда.*, Препринт № 43 ИПМ им. М. В. Келдыша РАН, ИПМ им. М. В. Келдыша РАН, Москва, 2014.

Дата поступления 25.04.2016

Modeling of a non-stationary trajectories ensemble using Fokker-Planck equation

© L. V Klochkova[4], J. H. Orlov[5], S. L. Fedorov[6]

Abstract. There is Discusses the kinetic Fokker-Planck equation for sample density distribution functions of random variables seen in practice in the form of the time number. Instead of studying one of the possible realizations of the time series it is proposed to consider an ensemble of random trajectories generated by the empirical distribution function. A model is proposed to describe the time variation functional defined on random trajectories and practical importance. This, for example, the indicator of level of pollution of the metropolis in the form of average pollutant concentration over a certain period of time, a similar indicator of the changing epidemiology in the region, the functional efficiency in the management of pollution in the form of lower pollution levels as a result of certain actions, etc. There is formulated a method of testing the control functions for the non-stationary trajectories ensemble.

Key Words: random distribution function, the Fokker-Planck equation, the ensemble of non-stationary trajectories, the testing control functions

[4] Senior Research Fellow of Keldysh Institute of Applied Mathematics of RAS, Moscow; klud@imamod.ru.

[5] Senior, Researcher Officer of the Institute of applied mathematics by name M.V.Keldysh of RAS, Moscow; ov3159fd@yandex.ru.

[6] Aspirant of the Institute of applied mathematics by name M.V.Keldysh of RAS, Moscow