Статистический анализ и прогнозирование случайных процессов с помощью смешанных гауссовских моделей

1 Введение

Во многих областях прикладной математики рассматриваются случайные процессы $X(t)$, задаваемые стохастическим дифференциальным уравнением вида

$$dX(t) = a(t)dt + b(t)dW, \tag{1}$$

где $W(t)$ – стандартный винеровский процесс. Коэффициенты $a(t)$ и $b(t)$ – случайны и, вообще говоря, неизвестны. В частности, уравнения вида (1) широко используются в задаче ассимиляции данных при анализе разномасштабной изменчивости геофизических переменных [4]. В финансовой математике популярны специальные версии уравнения (1). В частности, модель геометрического броуновского движения

$$dX(t) = aX(t)dt + bX(t)dW, \tag{2}$$

где $a \in \mathbb{R}$, $b > 0$. Известно много обобщений модели (2) с конкретными видами зависимости a и b от $X(t)$ и других случайных процессов, например, модели Леланда [16], Барлса–Сонера [3], Хестона [13], Кокса–Ингерсолла–Росса [7], Халла–Уайта [14] и другие так называемые модели стохастической волатильности (см. также [8, 9, 2]).

При отсутствии априорной информации о структуре процесса $X(t)$ для успешного изучения и прогнозирования его эволюции первостепенную важность приобретает задача определения коэффициентов $a(t)$ и $b(t)$, то есть задача статистической реконструкции этих коэффициентов. Различным аспектам решения этой задачи посвящены многие десятки работ, см., например, [18, 10, 11, 12, 15, 17] и дальнейшие ссылки в этих работах. Однако в большинстве работ, посвященных этой задаче, коэффициенты стохастического дифференциального уравнения (1) трактуются как *известные* функции от времени и самогó процесса, зависящие от неизвестных числовых параметров.

Здесь же мы трактуем эти коэффициенты как *неизвестные случайные процессы*. В силу случайности этих функциональных коэффициентов задача их реконструкции допускает как минимум две разные формулировки: 1) можно пытаться найти (случайные же) приближения к значениям самих функций $a(t)$ и $b(t)$, т. е. найти их *точечные аппроксимации* и 2) можно пытаться найти (статистически оценить) *распределения* случайных величин $a(t)$ и $b(t)$. Во втором случае, зная какие-либо свойства этих коэффициентов, например, структуру их функциональной зависимости от исходного процесса $X(t)$ (скажем, как в моделях Леланда, Барлса–Сонера, Хестона, Кокса–Ингерсолла–Росса или Беляева и др.), можно найти оценки числовых параметров, входящих в эти модели.

2 Оценивание конечномерных распределений коэффициентов $a(t)$ и $b(t)$

Сначала рассмотрим вторую задачу. Пусть $n \geqslant 1$ и $t_0 = 0 < t_1 < \ldots < t_n$ – моменты времени, в которые наблюдается процесс $X(t)$. Для простоты предположим, что $t_i - t_{i-1} = 1$ для любого $i \geqslant 1$. Обозначим $X_i = X(t_i)$, $i = 1, \ldots, n$. Таким образом, анализируется временнóй ряд

$X_1, \ldots, X_n$. Поскольку приращения винеровского процесса имеют нормальные распределения, из вида уравнения (1) вытекает, что распределение приращения $X_i - X_{i-1}$ процесса $X(t)$ можно аппроксимировать распределением вида

$$\mathsf{P}\big(X_i - X_{i-1} < x\big) \approx \mathsf{E}\Phi\Big(\frac{x - A_i}{B_i}\Big), \tag{3}$$

где $\Phi(x)$ – стандартная нормальная функция распределения,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy, \quad x \in \mathbb{R},$$

$A_i \in \mathbb{R}$ и $B_i > 0$ – случайные величины. В свою очередь, для распределений случайных величин A_i и B_i, по отношению к которым берется математическое ожидание в (3), можно использовать дискретную аппроксимацию. Тогда вместо (3) для распределения приращения $X_i - X_{i-1}$ можно применить приближение вида конечной смеси нормальных распределений

$$\mathsf{P}\big(X_i - X_{i-1} < x\big) \approx \sum\nolimits_{k=1}^{K} p_k \Phi\Big(\frac{x - a_k}{b_k}\Big), \tag{4}$$

где $K \in \mathbb{N}$, $p_k \geqslant 0$, $k = 1, \ldots, K$, $p_1 + \ldots + p_K = 1$. Очевидно, параметры p_k, a_k и b_k зависят также от i и изменяются при переходе от t_i к t_{i+1}.

Для статистического оценивания параметров p_k, a_k и b_k можно использовать подход, основанный на скользящем разделении смесей, описанный в [1]. Статистические закономерности поведения рассматриваемых процессов $X(t)$, $a(t)$, $b(t)$ изменяются во времени, вообще говоря, нерегулярным образом, результатом чего является отсутствие универсального смешивающего закона. Таким образом, чтобы изучить динамику изменения статистических закономерностей в поведении исследуемого процесса, задача статистического разделения конечных смесей нормальных законов должна быть последовательно решена на интервалах времени, постоянно сдвигающихся в направлении «астрономического» времени. Тем самым параметры смесей (параметры сдвига (дрейфа) a_k, масштаба (диффузии) b_k и веса компонент p_k) оцениваются как функции времени.

Задача 1. Для решения задачи оценивания параметров параметров сдвига (дрейфа) a_k, масштаба (диффузии) b_k и весов компонент p_k на каждом окне использовать ЕМ-алгоритм, реализующий метод максимального правдоподобия [1]. Но поскольку ЕМ-алгоритм обладает несколькими недостатками (невысокое быстродействие, неустойчивость по исходным данным, необязательная глобальность найденного максимума и др.), предложить его модификации, устраняющие или смягчающие эти недостатки. В частности, теоретически исследовать состоятельность оценок, получаемых *сеточными* версиями ЕМ-алгоритма.

Задача 2. Для оценивания параметров p_k, a_k и b_k использовать метод, заключающийся в трактовке наблюдений, попавших в каждое окно (отрезка временно́го ряда) как независимой однородной выборки, по которой строится эмпирическая функция распределения

$$F_{t,n}(x) = \frac{1}{n} \sum\nolimits_{j=t}^{t+n-1} \mathbb{I}(X_j < x), \quad x \in \mathbb{R}.$$

Здесь t – параметр, характеризующий положение (например, номер) окна, n – число наблюдений, попавших в окно («ширина окна»), $\mathbb{I}(A)$ – индикатор события A. Затем решается задача минимизации какого-либо расстояния (например, равномерного, L_1 или L_2) между этой $F_{t,n}(x)$ и смесью, стоящей в правой части (4), по параметрам a_k, b_k и p_k. С этой целью можно использовать любую процедуру поиска экстремума из включенных в стандартные программы. Цель – добиться адекватности результата и быстродействия такого метода оценивания. Вместо указанных выше метрик можно использовать модули разностей между $F_{t,n}(x)$ и и смесью, стоящей в правой части (4), в каких-либо произвольно выбранных точках $x_1, \ldots, x_M$, где M – произвольно выбираемое число, удовлетворяющее соотношению $M \geqslant 3K - 1$ (точек должно быть больше, чем оцениваемых параметров). Особо следует отметить, что выбор ширины окна определяется исследователем. Окно (в данном случае объем выборки, считающейся однородной) не должно быть слишком малым, чтобы содержать число наблюдений, обеспечивающее достаточную точность статистических процедур. С другой стороны, окно не должно быть слишком большим, чтобы итоговые выводы не были «пересглажены». Также очевидно, что ширина окна связана с горизонтом прогнозирования.

Для случая, когда в качестве расстояния рассматривается дискретный аналог метрики L_2, задача решена.

Задача 2.1. Доказать, что точки $x_1, \ldots, x_M$ следует выбирать так, чтобы

$$F_{t,n}(x_{r+1}) - F_{t,n}(x_r) = F_{t,n}(x_r) - F_{t,n}(x_{r-1}), \quad r = 1, \ldots, M-1.$$

Для случая, когда в качестве расстояния рассматривается дискретный аналог метрики L_2, задача решена.

Задача 2.2. Также вместо Задачи 1 исследовать метод оценивания параметров смеси в правой части (4), основанный на минимизации расстояния между плотностью смеси и какой-либо эмпирической оценкой плотности.

Задача 3. С помощью оценок смеси (4) осуществить декомпозицию изменчивости (волатильности) процесса $X(t)$ на «динамические» и «диффузионные» компоненты [1]:

$$\mathsf{D}\big(X(t_i) - X(t_{i-1})\big) \approx \sum\nolimits_{j=1}^{K} (a_j - \overline{a})^2 p_j + \sum\nolimits_{j=1}^{K} p_j b_j^2, \tag{5}$$

где

$$\overline{a} = \sum\nolimits_{j=1}^{K} a_j p_j.$$

Первое слагаемое в правой части (5) зависит только от весов p_j и параметров положения (сдвига) a_j компонент смеси (4) и потому характеризует ту часть волатильности, которая обусловлена наличием локальных трендов, то есть «динамическую» компоненту волатильности, тогда как второе выражение в (5) зависит только от весов p_j и параметров масштаба («коэффициентов диффузии») b_j компонент и потому характеризует «чисто диффузионную» компоненту волатильности.

Динамические компоненты связаны с текущим распределением случайного коэффициента $a(t)$ уравнения (1), тогда как диффузионные компоненты порождаются текущим распределением случайного коэффициента $b(t)$. Сопоставить полученную декомпозицию с физической природой исследуемого процесса.

3 Исследование корреляционной структуры временного ряда

Задача 4. Исследовать корреляционную структуру исходного процесса (временного ряда). С этой целью ряд $X_1, X_2, \ldots$ преобразовать в двумерный, попарно объединяя его элементы. В итоге получается двумерный ряд $\{(X_i, X_{i+1})\}_{i \geqslant 1}$. К полученному двумерному ряду применить методы разделения смесей двумерных нормальных распределений. При этом внедиагональные элементы получаемых ковариационных матриц – это ковариации между соседними элементами исходного ряда. Исследовать эффективность разных способов попарного объединения: непересекающиеся пары и пересекающиеся пары.

Задача 5.. Исследовать корреляционную структуру исходного процесса (временного ряда) с помощью преобразования ряда $X_1, X_2, \ldots$ преобразовать в трехмерный, последовательно объединяя его элементы по три. К полученному трехмерному ряду применить методы разделения смесей трехмерных нормальных распределений. При этом внедиагональные элементы получаемых ковариационных матриц – это ковариации между элементами исходного ряда, отстоящими друг от друга на один или два отсчета. Исследовать эффективность разных способов объединения: непересекающиеся тройки и тройки, пересекающиеся на одно или два наблюдения.

4 Реконструкция коэффициентов $a(t)$ и $b(t)$ как функций времени

Задача 6. Реконструировать коэффициенты $a(t)$ и $b(t)$, то есть построить их «точечные» оценки путем использования оценок *распределений* коэффициентов уравнения (1), полученных в результате решения Задач 1 – 2, для построения оценок *самих* коэффициентов. В качестве таких оценок берутся математическое ожидание и среднеквадратическое отклонение оцененного распределения:

$$a(t) \approx \overline{a}(t) = \sum\nolimits_{k=1}^{K} p_k a_k, \qquad b^2(t) \approx \overline{b}(t) = \sum\nolimits_{k=1}^{K} p_k b_k. \tag{6}$$

Здесь t – время (положение окна), параметры a_k, b_k, p_k также зависят от положения окна. Если $\Delta_t \equiv t_i - t_{i-1} \neq 1$, то вместо оценок (6) надо использовать «масштабированные» оценки

$$\Delta_t \overline{a}(t) = \sum\nolimits_{k=1}^{K} p_k a_k, \qquad \sqrt{\Delta_t} \cdot \overline{b}(t) = \sum\nolimits_{k=1}^{K} p_k b_k.$$

Как известно, математическое ожидание случайной величины является наилучшим среднеквадратическим прогнозом этой случайной величины. Так как вектор $\big(\overline{a}(t), \overline{b}(t)\big)$ является математическим ожиданием случайного вектора, принимающего значения (a_i, b_i) с вероятностями p_i, $i = 1, \ldots, K$, то оценки (6) в определенном смысле дают наилучший (минимизирующий среднеквадратический риск) прогноз значения случайных коэффициентов $a(t)$ и $b(t)$.

Задача 6.1. Предложить строгую формулировку и строгое доказательство оптимальности оценок $\overline{a}(t)$ и $\overline{b}(t)$ в смысле среднеквадратического риска.

Полученные таким образом *полупараметрические* оценки $\overline{a}(t)$ и $\overline{b}(t)$ коэффициентов являются более точными, нежели *непараметрические* оценки, получаемые по методу, примененному при решении задач анализа процессов теплообмена между атмосферой и океаном в [5, 6]. В работах [5, 6] задача оценивания коэффициентов $a(t)$ и $b(t)$ решалась с помощью «частотного» подхода к оцениванию условных переходных вероятностей, сводящегося к усреднению «по пространству», что в ситуациях с нестационарной неопределенностью (в частности, в задачах финансовой математики) малоэффективно. Фактически во второй задаче предлагается подход, при котором сглаживание по пространству заменяется сглаживанием по времени (за счет интерпретации наблюдений, попадающих в окно, как однородной выборки). Как уже было отмечено, ширина окна не должна быть слишком малой, чтобы обеспечить приемлемую точность при решении задачи разделения смесей (определения оценок параметров смеси (4)), и не должна быть слишком большой, чтобы упомянутое выше сглаживание не стало чрезмерным и чреватым потерей информации о довольно быстрых изменениях исходного процесса. В связи с этим традиционные постановки задач статистического оценивания, связанные с исследованием состоятельности получаемых оценок, в которых подразумевается возможность неограниченного объема выборки, что в рассматриваемом случае эквивалентно неограниченному увеличению ширины окна, в определенном смысле противоречат цели достижения приемлемой точности реконструкции (возможно, быстро) изменяющихся коэффициентов стохастического дифференциального уравнения (1).

Задача 7. Исследовать корреляционную структуру процессов $a(t)$ и $b(t)$ с помощью двумерных или трехмерных оценок, полученных в результате решения Задач 4 и 5.

5 Исследование структуры зависимости коэффициентов $a(t)$ и $b(t)$ от процесса $X(t)$

Задача 8. Исследовать структуру зависимости коэффициентов $a(t)$ и $b(t)$ от процесса $X(t)$ с помощью методов регрессионного анализа.

6 Интеллектуальное прогнозирование случайных процессов с направленным обучением

Задача 8 (главная !!!). Построить метод интеллектуального прогнозирования процесса $X(t)$, обучающийся на признаковом пространстве, обогащенном оценками коэффициентов $a(t)$ и $b(t)$ или оценками параметров смеси (4) без привлечения информации, являющейся «внешней» по отношению к исследуемому процессу. Например, в результате описанной выше реконструкции коэффициентов уравнения (1) в дополнение к исходному временному ряду $X_1, \ldots, X_n$ возникает $(3K-1)$-мерный временной ряд параметров смеси (4), оцененных на каждом окне. За счет этого дополнительного векторного временного ряда в большей степени учитывается информация о статистических закономерностях поведения исходного случайного процесса $X(t)$. Эту информацию следует использовать для обогащения признакового пространства на этапе обучения прогнозирующего алгоритма. Например, векторной авторегрессии или нейронной сети. Для точного прогнозирования процесса $X(t)$ по временному ряду $X_1, \ldots, X_n$ большое значение имеет правильный выбор архитектуры нейронной сети. Практика показывает, что лучшие результаты можно получить с помощью LSTM-нейронных сетей. В дополнение к $(3K-1)$-мерному (векторному) временному ряду параметров смеси (7) признаковое пространство можно еще более обогатить коэффициентами авторегрессионной модели, подгоняемой к временному ряду $X_1, \ldots, X_n$ на каждом окне. Тогда обучение прогнозирующего алгоритма будет в бо́льшей степени ориентировано на статистические закономерности поведения рассматриваемого процесса.

Задача 8.1. Исследовать разные способы пополнения признакового пространства (сравнить результаты прогнозирования при обогащении признакового пространства разными характеристиками). При пополнении признакового пространства $(3K-1)$-мерным дополнительным временны́м рядом параметров смешанной вероятностной модели (4), описывающих "текущие" статистические закономерности поведения наблюдаемых величин, для гарантии надежного обучения алгоритмов важную роль играет алгоритм формирования этого дополнительного ряда, суть которого в фиксации правила сопоставления параметров смеси нормальных законов (весов, средних значений и дисперсий компонент), оцененных на некотором окне, с аналогичными параметрами на предыдущем и следующем окнах. Эта проблема возникает из-за формальной невозможности отождествления индексов (номеров) компонент смеси, получаемых на разных окнах, так как смесь вероятностных (в данном случае нормальных) распределений – это (взвешенная) сумма, а сумма не зависит от порядка слагаемых. Ранее предлагались некоторые методы сопоставления, см., например, [19]. Однако все они не имели достаточного обоснования и основывались главным образом на соображениях гладкости получаемых дополнительных временных рядов.

Вместо $(3K-1)$-мерного дополнительного временно́го ряда параметров смешанной вероятностной модели (4) признаковое пространство можно пополнять двумя рядами реконструкций $\overline{a}(t)$ и $\overline{b}(t)$ коэффициентов $a(t)$ и $b(t)$ уравнения (1), которые в свою очередь, можно пополнить рядами коэффициентами автогрессий, подгоняемых к процессам $\overline{a}(t)$ и $\overline{b}(t)$.

Можно предложить новый подход к формированию дополнительных признаков. Этот подход заключается в том, что дополнительные признаки формируются в соответствии с естественным упорядочением значений ф.р., определенной на текущем окне.

Конкретизируем сказанное. Как уже отмечалось, конечные смеси нормальных законов вида (4) являются идентифицируемыми [20]. Это означает, что между наборами параметров p_k, a_k, b_k, $k = 1, \ldots, K$, и соответствующими функциями распределения $F(x)$ существует *взаимно однозначное соответствие* (с точностью до перестановки индексов i – это обстоятельство и вызывает проблемы с идентификацией параметров на разных окнах). Поэтому вместо "передачи" параметров p_k, a_k, b_k, $k = 1, \ldots, K$ от одного окна к другому можно "передавать" значения соответствующих функций распределения (4). Это можно сделать двумя способами.

Первый способ. Выберем число $M \in \mathbb{N}$ и зафиксируем сетку $\{x_1, x_2, \ldots, x_M\}$ на области изменения приращений анализируемого процесса на всем доступном временном интервале (например, в качестве x_1 можно взять минимальное значение приращения, а в качестве x_M – максимальное). При этом M должно быть довольно большим (например, должно быть выполнено условие $M \geqslant 3K-1$). Точки $x_2, \ldots, x_{M-1}$ можно выбирать в соответствии с **задачей 2.1**. Теперь вместо параметров p_k, a_k, b_k, $k = 1, \ldots, K$, формируемых на каждом (i-м)окне, построим новый вектор параметров $t_1^{(i)},\ldots,t_M^{(i)}$, положив

$$t_j^{(i)} = F(x_j), \quad j = 1,\ldots,M,$$

где в качестве параметров смеси $F(x)$ используются их оценки, построенные на i-м окне. При такой перепараметризации возможна некоторая потеря точности, вызванная, например, тем, что задача

$$\sum\nolimits_{j=1}^{M}\Bigl(t_j^{(i)} - \sum\nolimits_{k=1}^{K} p_k\Phi\Bigl(\frac{x_j - a_k}{b_k}\Bigr)\Bigr)^2 \longrightarrow \min_{p_k,a_k,b_k,\,k=\overline{1,K}}$$

может допускать не одно решение. Такая потеря точности компенсируется увеличением числа M. Но при этом не нужны никакие дополнительные процедуры, связанные с идентификацией параметров (установлением соответствия между параметрами) на разных окнах.

Естественно, что за исключением n, K, M и точек $x_1,\ldots,x_M$, остающихся неизменными для любого положения окна, все величины, упоминаемые выше, зависят от положения окна. Таким образом, наряду с исходным временным рядом $X_1,\ldots,X_N$ признаковое пространство обогащается M-мерным дополнительным временным рядом $\bigl(t_1^{(i)}, t_2^{(i)},\ldots,t_M^{(i)}\bigr)$, $i = n,\ldots,N$.

В отличие от первого способа, когда значения аргументов изменяющейся от окна к окну ф.р. остаются неизменными при скольжении окна, второй способ заключается в формировании на каждом окне набора квантилей оцененной на окне ф.р. (4) равноотстоящих порядков, например, децилей.

Этот способ вообще можно сделать непараметрическим, заменив квантили ф.р. (4) их приближениями – выборочными аналогами – порядковыми статистиками, построенными по попадающим в окно наблюдениям. При этом в соответствии с **задачей 2.1** номера (*ранги*) “передаваемых” порядковых статистик должны быть одинаковыми для всех окон (а все окна должны иметь одинаковую ширину). Например, ранги “передаваемых” порядковых статистик должны составлять арифметическую прогрессию. Такой выбор информативных порядковых статистик обеспечивает наименьшее значение максимально возможного отклонения эмпирической ф.р., построенной по попавшим в окно наблюдениям, от теоретической ф.р. (4) в остальных точках. “Непараметрический” вариант второго способа хорош тем, что работает существенно быстрее “параметрического”, так как не нужно оценивать параметры смеси (4). Более того, при таком подходе в прогнозирующий алгоритм не вносятся погрешности, обусловленные возможным неправильным выбором теоретической модели распределения. Однако при “непараметрическом” варианте второго способа следует аккуратно применять алгоритм VAR (векторную авторегрессию) в качестве средства прогнозирования, так как при малом числе M могут наблюдаться участки оси времени, на которых векторы выборочных квантилей будут постоянны.

Идея обоих указанных выше способов заключается в том, что исходный временной ряд дополняется многомерным рядом, в котором каждый элемент по сути представляет собой дискретное распределение вероятностей. Такой подход можно считать развитием методов прогнозирования временных рядов гистограммного типа, см. [21, 22, 23] и дальнейшие ссылки в этих работах.

Описанные выше способы обогащения признакового пространства касаются более полного использования информации об *одномерных распределениях* анализируемого процесса. Чтобы также учесть и информацию о взаимозависимости значений процесса в разные моменты времени, указанные признаки можно пополнить значениями коэффициентов авторегрессии (например, первых двух порядков), вычисляемыми на каждом окне.

Задача 8.2. Описанный выше подход основан на скользящем разделении смесей нормальных законов. При этом все наблюдения, попавшие в одно окно, считаются одинаково распределенными. Возникает естественный вопрос: какому моменту времени соответствуют получаемые значения коэффициентов стохастического процесса Ито? Если окно ширины n включает наблюдения $X_\tau, X_{\tau+1},\ldots,X_{\tau+n-1}$, то полученный $3K-1$-мерный вектор оценок параметров смеси (4) и соответствующих реконструкций коэффициентов $a(t)$ и $b(t)$ с равными основаниями можно считать соответствующим любому моменту времени $t \in [\tau,\, \tau+n-1]$, например, первому из них. Другими словами, имеет место своеобразное “запаздывание” получаемых оценок по отношению к исходному ряду. Но при решении задачи прогнозирования естественно иметь адекватные, не запаздывающие, оценки указанных параметров которые можно (и нужно) считать соответствующими последнему по времени из доступных наблюдений (правой границе окна). С этой целью можно учесть то, что $X_1, \ldots, X_N$ – наблюдения, *естественно упорядоченные по времени*, и модифицировать решение **задачи 2**: вместо классической эмпирической функции распределения $F_{t,n}(x)$ использовать функцию (статистику)

$$F^*_{t,n}(x) = \sum_{j=t}^{t+n-1} w_{j-t+1}\mathbb{I}(X_j < x), \quad x \in \mathbb{R},$$

где $w_1, \ldots, w_n$ – известные числа, $w_i \geqslant 0$ $(i = 1, \ldots, n)$, $w_1 + \ldots + w_n = 1$. Для решения указанной задачи прогнозирования веса $w_1, \ldots, w_n$ следует выбирать так, чтобы они образовывали возрастающую последовательность. Например,

$$w_j = C_p(1 - p^j), \quad j = 1, \ldots, n, \quad p \in [0, 1),$$

где

$$C_p = \left(n - \frac{p(1-p^n)}{1-p}\right)^{-1}.$$

Функция $F_{t,n}(x)$ является частным случаем $F^*_{t,n}(x)$, где $w_1 = \ldots = w_n = \frac{1}{n}$ $(p = 0)$.

Нужно перепробовать разные варианты задания весов (разные значения p) и сравнить получаемые прогнозы. Исследовать зависимость оптимального значения p от ширины окна n и текущей волатильности исходного процесса (см. **задачу 3**).

Задача 8.3. Описать асимптотические свойства обобщенной эмпирической функции распределения $F^*_{t,n}(x)$.

Задача 8.4. Предложить модификацию метода максимального правдоподобия, аналогичную **задаче 8.2**. Например, пусть $X_{\tau+1}, \ldots, X_{\tau+n}$ – последовательные элементы временного ряда, попавшие в очередное окно ширины n. Пусть $f(x; \theta_\tau)$ – предполагаемая плотность каждого наблюдения, θ_τ – вектор неизвестных параметров, соответствующий этому окну. Предположим, что "достоверность" или "важность" не всех элементов $X_{\tau+1}, \ldots, X_{\tau+n}$ одинакова. Тогда вместо классической задачи максимизации (логарифма) функции правдоподобия

$$\sum_{j=1}^{n} \log f(X_{\tau+j}; \theta_\tau) \longrightarrow \max_\theta$$

можно решать (исследовать) задачу

$$\sum_{j=1}^{n} v_j \log f(X_{\tau+j}; \theta_\tau) \longrightarrow \max_\theta,$$

где $v_1, \ldots, v_n$ – неотрицательные числа такие, что $v_1 + \ldots + v_n = n$. Для задачи прогнозирования естественно считать, что при этом $v_j \leqslant v_{j+1}$, $j \in \{1, \ldots, n-1\}$ (см. **задачу 8.2**). Например, $v_j = 1 - p^j$, $j = 1, \ldots, n$, $p \in [0, 1)$. При этом значение $p = 0$ соответствует классической функции правдоподобия.

Задача 8.5. Описать и использовать соответствующую модификацию EM-алгоритма для случая, когда $f_\tau(x; \theta)$ – конечная смесь нормальных плотностей.

Результаты решения **задач 8.1, 8.2, 8.4** и **8.5** применить для решения **задачи 8**!

Список литературы

[1] *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. – Москва: Изд-во Московского университета, 2011.

[2] *Ширяев А. Н.* Основы стохастической финансовой математики. Том 1. Факты. Модели. – Москва: Фазис, 1998.

[3] *Barles G., Soner H. M.* Option pricing with transaction costs and a nonlinear Black–Scholes equation // Finance and Stochastics, 1998. Vol. 2. P. 369–397.

[4] *Belyaev K., Kuleshov A, Tuchkova N. Tanajura C. A. S.*. An optimal data assimilation method and its application to the numerical simulation of the ocean dynamics // Mathematical and Computer Modelling of Dynamical Systems, 2017. P. 1-14. DOI: 10.1080/13873954.2017.1338300

[5] *Беляев К. П., Королев В. Ю., Горшенин А. К., Антипов А. И., Имеев М. А., Кирюшкин Н. И., Лобовский М. А.* Некоторые особенности внутригодовой изменчивости потоков тепла в Северной Атлантике // Известия Российской академии наук. Физика атмосферы и океана, 2021. Т. 57. Вып. 6. С. 707-720.

[6] *Беляев К. П., Горшенин А. К., Королев В. Ю., Плеханов А. Д.* Статистический анализ внутри- и межгодовой изменчивости экстремальных значений явных и скрытых потоков тепла в Северной Атлантике за 1979–2021 гг. // Известия Российской академии наук. Физика атмосферы и океана, 2022. Т. 58. Вып. 6. С. 720-736.

[7] *Cox J. C., Ingersoll J. E., Ross S. A.* A theory of the term structure of interest rates // Econometrica, 1985. Vol. 53. P. 385–407.

[8] *Derman E., Kani J.* Riding on a smile // Risk, 1994. Vol. 7. P. 32–39.

[9] *Dupire B.* Pricing with a smile // Risk, 1994. Vol. 7. P. 18–20.

[10] *Florens-Zmirou D.* On estimating the diffusion coefficient from discrete observations // Journal of Applied Probability, 1993. Vol. 30. No. 4. P. 790–804.

[11] *Genon-Catalot V., Jacod J.* On the estimation of the diffusion coefficient for multi-dimensional diffusion processes // Annales de l'Institute Henri Poincaré, Ser. B, 1993. Vol. 29. No. 1. P. 119–151.

[12] *Genon-Catalot V., Jacod J.* Estimation of the diffusion coefficient for diffusion processes: random sampling // Scandinavian Journal of Statistics, 199. Vol. 21. No. 3. P. 193–221.

[13] *Heston S. L.* A closed-form solution for options with stochastic volatility, with application to bond and currency options // Review of Financial Studies, 1993. Vol. 6. P. 327–343.

[14] *Hull J., White A.* The pricing of options on assets with stochastic volatilities // Journal of Finance, 1987. Vol. 42. P. 281–308.

[15] *Lamouroux D., Lehnertz K.* Kernel-based regression of drift and diffusion coefficients of stochastic processes // Physics Letters A, 2009. Vol. 373. P. 3507–3512.

[16] *Leland H. E.* Option pricing and replication with transactions costs // Journal of Finance, 1985. Vol. 40. P. 1283–1301.

[17] *Wei C., Shu H.* Maximum likelihood estimation for the drift parameter in diffusion processes // Stochastics, 2016. Vol. 88. No. 5. P. 699–710. DOI: 10.1080/17442508.2015.1124879

[18] *Yoshida N.* Estimation for diffusion processes from discrete observation // Journal of Multivariate Analysis, 1992. Vol. 41. P. 220–242.

[19] *Горшенин А. К., Королев В. Ю., Щербинина А. А.* Статистическое оценивание распределений случайных коэффициентов стохастического дифференциального уравнения Ланжевена // Информатика и ее применения, 2020. **14**. Вып.3. С. 3–12.

[20] *Teicher, H.* Identifiability of mixtures // Annals of Mathematical Statistics, 1961. Vol. 32. P. 244–248. DOI: 10.1214/AOMS/1177705155

[21] *Arroyo, J., Maté, C.* Forecasting histogram time series with k-nearest neighbours methods // Int. J. Forecast. 25(1), 192–207 (2009)

[22] *Arroyo, J., González-Rivera, G., Maté, C.* Forecasting with interval and histogram data. Some financial applications / A. Ullah, D. E. A. Giles (Eds). Handbook of empirical economics and finance. – Boca Raton–London–New York: CRC/Chapman and Hall, 2011. P. 247–280. DOI: 10.1201/b10440

[23] *Rakpho, P., Yamaka, W., Zhu, K.* Artificial Neural Network with Histogram Data Time Series Forecasting: A Least Squares Approach Based on Wasserstein Distance / Sriboonchitta, S., Kreinovich, V., Yamaka, W. (Eds). Behavioral Predictive Modeling in Economics. Studies in Computational Intelligence, vol 897. – New York: Springer, Cham. P. 351-362. DOI: 10.1007/978-3-030-49728-6_23