

# **ИНФОРМАТИКА И ЕЁ ПРИМЕНЕНИЯ**

**Научный журнал Отделения информационных технологий  
и вычислительных систем Российской академии наук**

Издается с 2007 года  
Журнал выходит ежеквартально

**Учредители:**  
**Российская академия наук**  
**Институт проблем информатики Российской академии наук**

## **РЕДАКЦИОННАЯ КОЛЛЕГИЯ**

академик С. В. Емельянов (главный редактор, член Редакционного совета)  
академик Ю. И. Журавлев (председатель Редакционного совета)  
академик С. К. Коровин  
академик Г. И. Савин  
академик И. А. Соколов (зам. главного редактора, член Редакционного совета)  
академик А. Л. Сتمпковский  
академик Ю. И. Шокин (член Редакционного совета)  
член-корреспондент РАН В. Л. Арлазаров  
член-корреспондент РАН А. Б. Жижченко  
член-корреспондент РАН И. А. Каляев  
член-корреспондент РАН Ю. С. Попков  
член-корреспондент РАН К. В. Рудаков  
член-корреспондент РАН Ю. А. Флеров  
член-корреспондент РАН Б. Н. Четверушкин  
член-корреспондент РАН Р. М. Юсупов  
профессор, д.т.н. В. И. Будзко  
профессор, д.т.н. А. А. Зацаринный  
профессор, д.ф.-м.н. А. В. Печинкин  
профессор, д.т.н. И. Н. Синицын  
профессор, д.ф.-м.н. С. Я. Шоргин (ответственный секретарь)

## **Редакция**

профессор, д.г.-м.н. Р. Б. Сейфуль-Мулюков; к.ф.-м.н. Е. Н. Арутюнов;  
О. В. Ломакина

© Институт проблем информатики Российской академии наук, 2008

## **Адрес редакции:**

Москва 119333, ул. Вавилова 44, корп. 2, ИПИ РАН,  
редакция журнала «Информатика и её применения»  
Тел. 8(499)135-86-92, e-mail [rust@ipiran.ru](mailto:rust@ipiran.ru)

**Журнал «Информатика и её применения» включен в «Перечень ведущих  
рецензируемых научных журналов и изданий, в которых должны быть  
опубликованы основные научные результаты диссертации на соискание  
ученой степени доктора и кандидата наук», утвержденный ВАК**

Подписной индекс журнала в каталоге «Пресса России» 88018 (годовая подписка)

# Информатика и её применения

Том 2 Выпуск 2 Год 2008

*Тематический выпуск*

*Вероятностно-статистические методы  
и задачи информатики и информационных технологий*

## СОДЕРЖАНИЕ

Предисловие	
<b>И. А. Соколов, В. Ю. Королёв</b>	<b>2</b>
Сеточные методы разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых индексов	
<b>В. Ю. Королёв, Е. В. Непомнящий, А. Г. Рыбальченко, А. В. Виноградова</b>	<b>3</b>
Некоторые статистические задачи, связанные с распределением Лапласа	
<b>В. Е. Бенинг, В. Ю. Королёв</b>	<b>19</b>
Некоторые оценки для близких к поглощающим марковских моделей	
<b>А. И. Зейфман, А. В. Чегодаев, В. С. Шоргин</b>	<b>35</b>
Оптимизация расположения станций обслуживания в пространстве	
<b>Т. В. Захарова</b>	<b>41</b>
Отсев эктопических импульсов из ритмограммы с использованием робастных оценок	
<b>А. В. Маркин, О. В. Шестаков</b>	<b>47</b>
Об асимптотическом распределении максимальной порядковой статистики в выборке случайного объема	
<b>В. И. Пагурова</b>	<b>55</b>
Оценивание распределения задержки в биологических динамических системах на примере модели, описывающей ВИЧ-инфекцию	
<b>А. Н. Ушакова</b>	<b>60</b>
Существование состоятельных последовательностей статистических критериев в дискретных статических задачах при сложной нулевой гипотезе	
<b>А. А. Грушо, Е. Е. Тимонина, В. М. Ченцов</b>	<b>64</b>
Стохастические разложения несмещенных оценок в случае однопараметрического экспоненциального семейства	
<b>В. В. Чичагов</b>	<b>67</b>
Abstracts	<b>76</b>
Об авторах	<b>79</b>
About Authors	<b>80</b>

---

Выпускающий редактор *Т. Торжкова*; Технический редактор *Л. Кокушкина*  
Художественный редактор *М. Седакова*

Сдано в набор 31.04.08. Подписано в печать 17.06.08. Формат 60 x 84 / 8  
Бумага офсетная. Печать офсетная. Усл.-печ. л. 10,0. Уч.-изд. л. 8,5. Тираж 200 экз.  
Заказ №

Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43  
[torus@torus-press.ru](mailto:torus@torus-press.ru); <http://www.torus-press.ru>

Отпечатано в ППП «Типография «Наука» с готовых диапозитивов, Москва 121099, Шубинский пер., д. 6.

## Предисловие

Статьи, собранные в данном тематическом выпуске журнала «Информатика и её применения», объединены тем, что все они посвящены разработке новых вероятностно-статистических методов и их применению к решению конкретных задач информатики и информационных технологий. Проблематика, охватываемая публикуемыми работами, развивается в рамках научного сотрудничества между Институтом проблем информатики Российской академии наук и факультетом вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, в частности в ходе работ над совместными научными проектами. Многие из авторов статей, включенных в данный номер журнала, являются активными участниками традиционного международного семинара по проблемам устойчивости стохастических моделей, руководимого В. М. Золотарёвым и В. Ю. Королёвым; ежегодные сессии этого семинара проводятся под эгидой МГУ и ИПИ РАН.

В статье В. Ю. Королёва, Е. В. Непомнящего и А. В. Виноградовой предлагается принципиально новый метод исследования стохастической структуры хаотических процессов и приводятся примеры его применения к анализу волатильности (степени изменчивости) финансовых индексов. Этот метод также может быть использован для исследования статистической структуры других информационных потоков.

Статья В. Е. Бенинга и В. Ю. Королёва посвящена исследованию задач проверки статистических гипотез о параметрах распределения Лапласа. Это распределение очень часто успешно используется в качестве альтернативы нормальному закону в тех прикладных задачах, в которых статистические закономерности поведения исследуемых характеристик демонстрируют склонность к большим вероятностям существенных отклонений от среднего значения, нежели при нормальном распределении.

Статья Т. В. Захаровой посвящена исследованию оптимальных пространственных размещений обслуживающих приборов в сложных системах обслуживания, в частности в сложных технических (транспортных) и информационных системах.

В статье А. И. Зейфмана, А. В. Чегодаева и В. С. Шоргина методы теории дифференциальных уравнений применяются к исследованию сложных марковских цепей, которые описывают поведение сложных систем массового обслуживания.

Статья О. В. Шестакова и А. В. Маркина посвящена исследованию методов статистического анализа ритмограмм, что имеет важное значение при создании систем анализа кардиограмм в реальном времени (в “on-line” режиме).

В статье В. И. Пагуровой исследуется точность аппроксимации распределения экстремальных порядковых статистик в выборках случайного объема. Подобные задачи имеют первостепенное значение при анализе рисков в неоднородных потоках экстремальных событий, в частности в хаотических информационных потоках в сложных инфотелекоммуникационных системах.

Статья А. Н. Ушаковой посвящена математическому моделированию процессов распространения инфекционных вирусных заболеваний, в частности ВИЧ-инфекции.

В статье А. А. Грушо, Е. Е. Тимониной и В. М. Ченцова решаются статистические задачи, имеющие принципиальное значение в проблемах информационной безопасности, в частности при исследовании возможности существования скрытых каналов передачи информации.

Наконец, статья В. В. Чичагова посвящена исследованию свойств несмещенных оценок параметров распределений из экспоненциального семейства. В этой статье строятся асимптотические разложения, позволяющие проводить сравнительное исследование точности несмещенных оценок параметров разнообразных вероятностных моделей реальных процессов.

Редакционная коллегия журнала выражает надежду, что данный специальный выпуск будет интересен специалистам в области информатики, теории вероятностей и математической статистики.

Заместитель главного редактора журнала «Информатика и её применения»,  
директор ИПИ РАН, академик

*И. А. Соколов*

Редактор-составитель тематического выпуска,  
профессор кафедры математической статистики  
факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова,  
ведущий научный сотрудник ИПИ РАН,  
доктор физико-математических наук

*В. Ю. Королёв*

# СЕТОЧНЫЕ МЕТОДЫ РАЗДЕЛЕНИЯ СМЕСЕЙ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ И ИХ ПРИМЕНЕНИЕ К ДЕКОМПОЗИЦИИ ВОЛАТИЛЬНОСТИ ФИНАНСОВЫХ ИНДЕКСОВ\*

В. Ю. Королёв<sup>1</sup>, Е. В. Непомнящий<sup>2</sup>, А. Г. Рыбальченко<sup>3</sup>, А. В. Виноградова<sup>4</sup>

**Аннотация:** Предложены методы статистического разделения смесей вероятностных распределений, основанные на минимизации невязки между теоретической и эмпирической функциями распределения. Основное внимание уделено минимизации  $\sup$ - и  $L_1$ -норм невязки. Показано, что такие задачи могут быть сведены к задачам линейного программирования. Для их численной реализации используется симплекс-метод. Предложенные методы применены к решению задачи декомпозиции волатильности финансовых индексов. Приведены примеры декомпозиции волатильности индексов AMEX, CAC 40, NIKKEI, NASDAQ.

**Ключевые слова:** разделение смесей вероятностных распределений; задача линейного программирования; симплекс-метод; волатильность

## 1 Введение

Для численного решения задачи разделения конечных смесей вероятностных распределений (т. е. задачи отыскания статистических оценок весов компонент смеси и параметров компонент смеси) при относительно большом числе компонент традиционно применяется EM-алгоритм. Если функция правдоподобия регулярна, то этот метод, как правило, находит наиболее правдоподобные оценки параметров смеси. Однако если функция правдоподобия нерегулярна, имеет много локальных максимумов (возможно, к тому же бесконечных), то EM-алгоритм становится крайне неустойчивым. К сожалению, последнее обстоятельство является серьезным препятствием при интерпретации результатов применения EM-алгоритма к разделению конечных смесей нормальных законов. Именно такие смеси повсеместно применяются при математическом моделировании многих явлений в самых разных областях — от биологии до экономики и от физики до финансового анализа.

В частности, как было экспериментально установлено, EM-алгоритм обладает сильной неустойчивостью по начальным данным. Например, в случае четырехкомпонентной смеси нормальных

законов при объеме выборки 200–300 наблюдений замена лишь одного наблюдения другим может кардинально изменить итоговые оценки, полученные с помощью EM-алгоритма [1].

Поэтому крайне необходимо иметь альтернативные методы разделения смесей, ориентированные не на максимизацию функции правдоподобия, а на оптимизацию других разумных критериев качества получаемых оценок. Наличие таких альтернативных достаточно эффективных методов является принципиально важным с точки зрения возможности адекватной практической интерпретации получаемых оценок. Именно такие альтернативные методы и предлагаются в данной статье.

## 2 Основная идея сеточных методов разделения смесей вероятностных распределений

Идея, которая лежит в основе метода разделения смесей, описываемого в данной статье, очень близка к идее гармонического анализа, когда периодическая функция раскладывается в ряд Фурье, т. е. представляется взвешенной комбинацией (рядом),

\* Работа выполнена при поддержке РФФИ, гранты 08-01-00345, 08-01-00363, 08-07-00152.

<sup>1</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, Институт проблем информатики Российской академии наук, vkorolev@comtv.ru

<sup>2</sup>45 ЦНИИ МО РФ

<sup>3</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, alex-rybalchenko@yandex.ru

<sup>4</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, a.nuta@mail.ru

возможно, бесконечно большого числа синусов и косинусов с различными (но кратными) периодами (т.е. с различными параметрами масштаба — частотами). Возможность приближения исходной функции с помощью такого разложения обоснована тем обстоятельством, что семейство синусов и косинусов с указанными свойствами образует базис, т.е. полную систему (линейно независимых) функций в пространстве (регулярных) периодических функций.

Аналогия будет видна более отчетливо, если рассмотреть семейство чисто масштабных смесей нормальных законов с нулевым средним. Поскольку при некоторых условиях регулярности каждое распределение вероятностей, сосредоточенное на неотрицательной полуоси, может быть приближено решетчатым распределением с произвольно высокой точностью (скажем, в метрике Леви, метризирующей слабую сходимость), можно заключить, что семейство нормальных распределений  $\mathbf{N}_{\text{rat}} = \{\mathcal{N}(0, s) : s — рациональное\}$  образует счетную полную систему функций в пространстве масштабных смесей нормальных законов с нулевым средним. Это означает, что для любой масштабной смеси нормальных законов и произвольно малого  $\epsilon > 0$  существует конечная линейная комбинация распределений из семейства  $\mathbf{N}_{\text{rat}}$  такая, что расстояние (скажем, расстояние Леви) между смесью и линейной комбинацией не превышает  $\epsilon > 0$ . Поскольку число слагаемых в такой линейной комбинации конечно и все параметры масштаба рациональны, найдется такое (минимальное) значение параметра масштаба, что параметры масштаба всех членов рассматриваемой линейной комбинации будут кратными этому минимальному значению параметра масштаба.

Конечные смеси нормальных законов находят самое широкое применение как модели распределений многих величин, наблюдаемых на практике (см., например, [2]). Именно поэтому данная статья посвящена применению сеточных методов разделения смесей к решению задачи оценивания параметров конечных смесей нормальных законов. Всюду далее стандартная нормальная функция распределения будет обозначаться  $\Phi(x)$ .

Рассмотрим смесь функций распределения вида

$$F(x) = \sum_{i=1}^k p_i \Phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in \mathbb{R}, \quad (1)$$

где  $k \geq 1$  — целое. В классической задаче разделения смесей параметрами, подлежащими статистическому оцениванию, являются тройки  $(p_i, a_i, \sigma_i)$ ,  $i = 1, \dots, k$ , где  $a_i \in \mathbb{R}$ ,  $\sigma_i > 0$ ,  $p_i \geq 0$ ,  $p_1 + \dots + p_k = 1$ .

Предположим, что заранее известны числа  $\underline{a}$ ,  $\bar{a}$  и  $\bar{\sigma}$  такие, что  $\underline{a} \leq a_i \leq \bar{a}$  и  $\sigma_i \leq \bar{\sigma}$  при всех  $i = 1, \dots, k$ . Другими словами, известны диапазоны изменения неизвестных параметров  $a_i$  и  $\sigma_i$ .

Идея, лежащая в основе рассматриваемого подхода, заключается в замене *интервалов*  $[\underline{a}, \bar{a}]$  и  $(0, \bar{\sigma}]$  возможных значений неизвестных параметров масштаба  $\sigma_i$  и сдвига  $a_i$  *дискретными* множествами известных точек. Эти точки могут быть выбраны, например, исходя из следующих соображений.

Пусть  $\epsilon_a$  и  $\epsilon_\sigma$  — положительные числа, определяющие априорные требования к точности оценивания параметров  $a_i$  и  $\sigma_i$ :

$$\max_i |a_i - \hat{a}_i| \leq \epsilon_a, \quad \max_i |\sigma_i - \hat{\sigma}_i| \leq \epsilon_\sigma, \quad (2)$$

где  $\hat{a}_i$  и  $\hat{\sigma}_i$  — искомые оценки параметров. Числа  $\epsilon_a$  и  $\epsilon_\sigma$  также можно интерпретировать как пороги различимости возможных значений параметров: значения  $a'$ ,  $a''$  и  $\sigma'$ ,  $\sigma''$  соответственно считаются неразличимыми, если

$$|a' - a''| \leq \epsilon_a, \quad |\sigma' - \sigma''| \leq \epsilon_\sigma. \quad (3)$$

Положим  $k_a = [(\bar{a} - \underline{a})/\epsilon_a] + 1$ ,  $k_\sigma = [\bar{\sigma}/\epsilon_\sigma] + 1$ , где символ  $[z]$  обозначает целую часть числа  $z$ . Для  $r = 1, 2, \dots, k_a + 1$  положим  $\tilde{a}_r = \underline{a} + (r - 1)\epsilon_a$ . Аналогично для  $l = 1, 2, \dots, k_\sigma$  положим  $\tilde{\sigma}_l = l\epsilon_\sigma$ . Тогда точки с координатами  $(\tilde{a}_r, \tilde{\sigma}_l)$  образуют узлы конечной сети, накрывающей прямоугольник  $\{(a, \sigma) : \underline{a} \leq a \leq \bar{a}, 0 \leq \sigma \leq \bar{\sigma}\}$ , представляющий собой множество возможных значений параметров сдвига и масштаба компонент смеси (1) (чтобы избежать возможной некорректности, исключена возможность равенства параметра масштаба нулю). Число узлов полученной сети равно  $K = (k_a + 1)k_\sigma$ . Для удобства записи и упрощения обозначений перенумеруем каким-либо образом узлы указанной сети, вводя *единый* индекс  $i$  для координат  $(\tilde{a}_i, \tilde{\sigma}_i)$  узла с номером  $i$  после перенумерации,  $i = 1, \dots, K$ .

Подход, рассматриваемый в данной статье, предложен в [1]. Базовая посылка этого подхода заключается в аппроксимации смеси (1) смесью с заведомо большим числом *известных* компонент:

$$\begin{aligned} F(x) &= \sum_{i=1}^k p_i \Phi\left(\frac{x - a_i}{\sigma_i}\right) \approx \\ &\approx \sum_{i=1}^K \tilde{p}_i \Phi\left(\frac{x - \tilde{a}_i}{\tilde{\sigma}_i}\right) \equiv \tilde{F}(x), \quad x \in \mathbb{R}. \quad (4) \end{aligned}$$

Такое приближение практически допустимо, поскольку в силу соотношений (2) и (3) для любой пары  $(a_r, \sigma_r)$  параметров компоненты смеси (1) обязательно найдется практически не отличимая

от нее пара  $(\tilde{a}_i, \tilde{\sigma}_i)$  параметров компоненты смеси  $\tilde{F}(x)$ . Веса же остальных компонент смеси  $\tilde{F}(x)$ , для параметров которых не найдется «близкой» пары параметров  $(a_r, \sigma_r)$  компоненты смеси (1), можно считать равными нулю. Действительно, если бы в соотношении (4) вместо *приближенного* было бы *точное* равенство, то в силу идентифицируемости семейства конечных смесей нормальных законов и в полном соответствии с определением идентифицируемости конечных смесей с точностью до переиндексации были бы справедливы равенства:

$$k = K, \quad p_i = \tilde{p}_i, \quad a_i = \tilde{a}_i, \quad \sigma_i = \tilde{\sigma}_i, \quad i = 1, \dots, k.$$

Заметим, что неизвестными параметрами смеси  $\tilde{F}(x)$  являются *только* веса  $\tilde{p}_1, \dots, \tilde{p}_K$ .

Пусть  $\mathbf{x} = (x_1, \dots, x_n)$  — (независимая) выборка наблюдений, каждое из которых представляет собой реализацию случайной величины с функцией распределения  $F(x)$ , задаваемой соотношением (1). Пусть  $(x_{(1)}, \dots, x_{(n)})$  и  $F_n(x)$  — соответственно вариационный ряд и эмпирическая функция распределения, построенные по выборке  $\mathbf{x}$ ,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(x_j < x), \quad x \in \mathbb{R}.$$

При этом очевидно, что

$$F_n(x_{(j)}) = \frac{j}{n}, \quad j = 1, \dots, n. \quad (5)$$

В силу теоремы Гливенко при больших  $n$  равномерно по  $x \in \mathbb{R}$  выполняется соотношение

$$F_n(x) \approx F(x). \quad (6)$$

Обозначим

$$\Phi_{ij} = \Phi\left(\frac{x_{(j)} - \tilde{a}_i}{\tilde{\sigma}_i}\right), \quad j = 1, \dots, n; \quad i = 1, \dots, K.$$

Отметим, что величины  $\Phi_{ij}$  *известны*.

Из (4) и (5) вытекает, что при больших  $n$

$$\tilde{F}(x) \approx F_n(x), \quad x \in \mathbb{R},$$

откуда с учетом (6) получается приближенное соотношение

$$\tilde{F}(x_{(j)}) = \sum_{i=1}^K \tilde{p}_i \Phi_{ij} \approx \frac{j}{n}, \quad j = 1, \dots, n. \quad (7)$$

Соотношение (7) можно использовать для отыскания оценок параметров  $\tilde{p}_1, \dots, \tilde{p}_K$  с помощью метода наименьших квадратов и метода наименьших модулей, рассматриваемых в последующих разделах. Реализация этих методов в данном случае довольно проста, поскольку  $\tilde{F}(x)$  зависит от параметров  $\tilde{p}_1, \dots, \tilde{p}_K$  *линейно*, так что мы не выходим за рамки линейной модели регрессионного анализа.

### 3 Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи метода наименьших квадратов

Рассматриваемую задачу удобно записать в векторно-матричном виде. Обозначим

$$\mathbf{p} = \begin{pmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \dots \\ \tilde{p}_K \end{pmatrix}; \quad \mathbf{\Phi} = \begin{pmatrix} \Phi_{11} & \Phi_{21} & \dots & \Phi_{K1} \\ \Phi_{12} & \Phi_{22} & \dots & \Phi_{K2} \\ \dots & \dots & \dots & \dots \\ \Phi_{1n} & \Phi_{2n} & \dots & \Phi_{Kn} \end{pmatrix};$$

$$\mathbf{u} = \begin{pmatrix} \frac{1}{n} \\ \frac{2}{n} \\ \dots \\ \frac{n-1}{n} \\ 1 \end{pmatrix}.$$

Тогда соотношение (7) можно переписать в матричной форме:

$$\mathbf{\Phi p} \approx \mathbf{u}$$

или

$$\mathbf{\Phi p} + \mathbf{d} = \mathbf{u},$$

где  $\mathbf{d}$  — вектор-столбец невязок

$$\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{pmatrix}$$

с компонентами

$$d_j = \frac{j}{n} - \sum_{i=1}^K \tilde{p}_i \Phi_{ij}.$$

Предположим, что  $n \geq K$ . В этом случае ранг случайной (в силу случайности выборки  $\mathbf{x} = (x_1, \dots, x_n)$ ) матрицы  $\mathbf{\Phi}$  с вероятностью единица равен  $K$ . Тогда из общей теории метода наименьших квадратов для линейных моделей вытекает, что решение

$$\hat{\mathbf{p}}^* = \begin{pmatrix} \hat{p}_1^* \\ \hat{p}_2^* \\ \dots \\ \hat{p}_K^* \end{pmatrix}$$

безусловной задачи наименьших квадратов

$$\arg \min_{\mathbf{p}} \mathbf{d}^T \mathbf{d} = \arg \min_{\mathbf{p}} \sum_{j=1}^n \left[ \sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n} \right]^2 \quad (8)$$

с вероятностью единица существует, единственно и имеет вид

$$\hat{\mathbf{p}}^* = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{u}. \quad (9)$$

При этом сумма компонент вектора  $\hat{\mathbf{p}}^*$  отнюдь не обязана быть равной единице. Обозначим

$$\gamma = 1 - \sum_{i=1}^K \hat{p}_i^*. \quad (10)$$

Пусть  $\mathbf{a}$  —  $K$ -мерный вектор-столбец, все компоненты которого равны единице:

$$\mathbf{a}^T = (1, 1, \dots, 1).$$

Тогда соотношение (10) можно записать в виде

$$\gamma = 1 - \mathbf{a}^T \hat{\mathbf{p}}^*.$$

Но в рассматриваемом случае компоненты искомого вектора  $\mathbf{p}$ , будучи вероятностями в дискретном распределении, связаны очевидным условием

$$\tilde{p}_1 + \dots + \tilde{p}_K = 1. \quad (11)$$

В терминах вектора  $\mathbf{a}$  условие (11) запишется как

$$\mathbf{a}^T \mathbf{p} = 1. \quad (12)$$

Хорошо известно, что если ранг матрицы  $\mathbf{\Phi}$  равен  $K$ , то решение  $\hat{\mathbf{p}}$  задачи (8) при условии (12) имеет вид

$$\begin{aligned} \hat{\mathbf{p}} &= \hat{\mathbf{p}}^* + \\ &+ (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{a} \left[ \mathbf{a}^T (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{a} \right]^{-1} (1 - \mathbf{a}^T \hat{\mathbf{p}}^*) \end{aligned} \quad (13)$$

(см., например, с. 85–86 в [3]), где вектор  $\hat{\mathbf{p}}^*$  определен в (9). Несложно видеть, что выражение  $\mathbf{a}^T (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{a}$  равно сумме всех элементов матрицы  $(\mathbf{\Phi}^T \mathbf{\Phi})^{-1}$ . Обозначим эту сумму  $S$ . Тогда с учетом (10) соотношение (13) примет вид

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}^* + \frac{\gamma}{S} (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{a}. \quad (14)$$

Несложно видеть, что  $i$ -я компонента  $s_i$  вектор-столбца  $(\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{a}$  равна сумме всех элементов  $i$ -й строки матрицы  $(\mathbf{\Phi}^T \mathbf{\Phi})^{-1}$ . Таким образом, из (14) окончательно получается, что

$$\hat{p}_i = \hat{p}_i^* + \frac{\gamma s_i}{S}, \quad i = 1, \dots, K. \quad (15)$$

Условие  $n \geq K$ , упрощающее вычисления, определяет выбор числа  $k_a + 1$  возможных значений параметров сдвига и числа  $k_\sigma$  возможных значений параметров масштаба компонент исходной смеси (1). А именно: чтобы обеспечить существование единственного решения  $\hat{\mathbf{p}}$  условной задачи наименьших квадратов, параметры сетки  $k_a$  и  $k_\sigma$  должны быть связаны с объемом выборки  $n$  соотношением

$$(k_a + 1)k_\sigma \leq n.$$

Случай  $n < K$  нуждается в более тщательном анализе, поскольку в таком случае оценки наименьших квадратов определены неоднозначно.

Если оценки  $\hat{p}_i$  удовлетворяют неравенствам

$$0 \leq \hat{p}_i \leq 1 \quad (16)$$

для всех  $i = 1, \dots, k$ , то задача решена. Однако может случиться так, что некоторые из последних неравенств не реализуются на полученном векторе. В таком случае можно идти несколькими путями.

Во-первых, в работе [4] показано, что задачу наименьших квадратов с ограничениями типа неравенств можно свести к задаче последовательного решения нескольких задач наименьших квадратов без ограничений. Вытекающий из этого результата алгоритм решения рассматриваемой задачи таков.

Если для некоторых  $i = 1, \dots, k$  соотношения (16) не выполнены, то минимум функции

$$SS(\mathbf{p}) = \sum_{j=1}^n \left[ \sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n} \right]^2$$

на множестве

$$\mathcal{P} = [0, 1]^K \cap \{\mathbf{p} : \tilde{p}_1 + \dots + \tilde{p}_K = 1\}$$

достигается в одной из граничных точек, в которых хотя бы для одного  $i$  выполнено равенство  $p_i = 0$  (напомним, что  $\hat{p}_1 + \dots + \hat{p}_K = 1$  по построению). Пусть  $I$  — некоторое подмножество множества  $\{1, 2, \dots, K\}$ . Как известно, всего таких подмножеств  $2^K$ . Для каждого такого подмножества  $I$  можно найти  $\hat{\mathbf{p}}_I$  — точку минимума функции

$$SS_I(\mathbf{p}_I) = \sum_{j=1}^n \left[ \sum_{i \in I} \tilde{p}_i \Phi_{ij} - \frac{j}{n} \right]^2$$

при условии  $\mathbf{p}_I^T \mathbf{p}_I = 1$  по формулам, аналогичным (14) и (15) (здесь  $\mathbf{p}_I$  — это вектор  $\mathbf{p}$ , компоненты которого с номерами, не попавшими в множество  $I$ , равны нулю). Если при этом  $\hat{\mathbf{p}}_I \in \mathcal{P}$ ,

то вычисляется значение  $SS_I(\hat{\mathbf{p}}_I)$ . Перебрав все возможные подмножества  $I$ , можно найти решение исходной задачи — точку

$$\hat{\mathbf{p}} = \arg \min_I SS_I(\hat{\mathbf{p}}_I).$$

Однако при таком подходе требуется решить  $2^K$  задач наименьших квадратов с линейным ограничением, что при больших  $K$  занимает чрезвычайно много времени.

Во-вторых, в работе [5] показано, что задачу наименьших квадратов с ограничениями типа неравенств можно свести к задаче квадратичного программирования, и предложена итерационная процедура — модифицированный симплекс-метод — для ее решения. Эта процедура реализована в виде встроенного средства Optimization Toolbox системы MATLAB и использована, в частности, в работе [6] для решения исходной задачи наименьших квадратов с ограничениями типа неравенств и равенства. Однако при этом объем вычислений трудно оценить, а статистические свойства оценок, получаемых при таком подходе, чрезвычайно трудно исследовать.

Наконец, исходную задачу наименьших квадратов с ограничениями типа неравенств и равенства можно заменить приближенной, накинув конечную сетку и на множество возможных значений вектора  $(p_1, \dots, p_k)$ , сведя таким образом решение задачи к простому перебору.

## 4 Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи метода наименьших модулей

### 4.1 Решение в смысле минимума sup-нормы

Соотношение (7) допускает различные конкретизации, связанные с различными нормами невязки. Решение задачи по методу наименьших квадратов, рассмотренное в предыдущем разделе, связано с минимизацией обычной евклидовой нормы невязки.

В данном разделе будет рассмотрено решение, связанное с минимизацией sup-нормы вектора невязок

$$\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{pmatrix},$$

где, как и ранее,

$$d_j = \frac{j}{n} - \sum_{i=1}^K \tilde{p}_i \Phi_{ij}.$$

Будем искать решение задачи

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \max_{1 \leq j \leq n} |d_j|. \quad (17)$$

Как известно, задача (17) может быть сведена (см., например, с. 91 в [7] и с. 137 в [8]) к задаче линейного программирования вида

$$f(\mathbf{p}, \theta) = \theta \longrightarrow \min_{(\mathbf{p}, \theta) \in \mathcal{Q}'}; \quad (18)$$

$$\mathcal{Q}' = \left\{ (\mathbf{p}, \theta) \in \mathbb{R}_+^{K+1} : \tilde{p}_1 + \dots + \tilde{p}_K \geq 1; \right. \\ \left. \tilde{p}_1 + \dots + \tilde{p}_K \leq 1; \theta \geq \sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n}; \right. \\ \left. \theta \geq - \sum_{i=1}^K \tilde{p}_i \Phi_{ij} + \frac{j}{n}, j = 1, \dots, n \right\}. \quad (19)$$

Последняя задача, как известно, решается с помощью стандартного симплекс-метода, заключающегося в направленном переборе угловых точек множества  $\mathcal{Q}'$  (см., например, гл. 1 в [7]).

### 4.2 Решение в смысле минимума $L_1$ -нормы

В данном разделе будет рассмотрено решение, связанное с минимизацией  $L_1$ -нормы вектора невязок  $\mathbf{d}$ . Другими словами, будем искать решение задачи наименьших модулей

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \sum_{j=1}^n |d_j|. \quad (20)$$

Обозначим  $\vec{\theta} = (\theta_1, \dots, \theta_n)^\top$ . Как известно, задача (20) может быть сведена (см., например, с. 91 в [7] и с. 137 в [8]) к задаче линейного программирования вида

$$f(\mathbf{p}, \vec{\theta}) = \sum_{j=1}^n \theta_j \longrightarrow \min_{(\mathbf{p}, \vec{\theta}) \in \mathcal{Q}''}; \quad (21)$$

$$\mathcal{Q}'' = \left\{ (\mathbf{p}, \vec{\theta}) \in \mathbb{R}_+^{K+n} : \tilde{p}_1 + \dots + \tilde{p}_K \geq 1; \right. \\ \left. \tilde{p}_1 + \dots + \tilde{p}_K \leq 1; \theta_j \geq \sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n}; \right. \\ \left. \theta_j \geq - \sum_{i=1}^K \tilde{p}_i \Phi_{ij} + \frac{j}{n}, j = 1, \dots, n \right\}. \quad (22)$$



Последняя задача также решается с помощью стандартного симплекс-метода, заключающегося в направленном переборе угловых точек множества  $Q''$  (см., например, гл. 1 в [7]).

Вычислительные реализации задач линейного программирования (18), (19) и (21), (22) в популярных системах MATHCAD и MATLAB обладают намного более высоким быстродействием, нежели вычислительные реализации EM-алгоритма.

С содержательной точки зрения рассмотренные в данном разделе задачи построения оценок вектора весов  $\mathbf{p}$  вполне равноправны с задачей, рассмотренной в предыдущем разделе. Более того, как известно, оценки наименьших модулей более устойчивы по отношению к наличию резко выделяющихся наблюдений, нежели оценки наименьших квадратов. Поэтому оценки минимума суп-нормы и минимума  $L_1$ -нормы в определенном смысле предпочтительнее оценок наименьших квадратов.

Конечно, рассмотренные методы дают лишь приближенное решение задачи разделения смесей. Однако при их реализации удается избежать итерационных процедур типа EM-алгоритма для поиска решений экстремальных задач. Решение, полученное такими методами, вполне можно использовать как начальное приближение для EM-алгоритма или его модификаций с целью последующего получения более точного решения.

При этом, имея решение (15) задачи разделения смесей, начальное приближение EM-алгоритма можно выбирать следующим образом. Упорядочим полученные оценки весов  $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_K$  по убыванию и получим набор  $\hat{p}_{i_1}, \hat{p}_{i_2}, \dots, \hat{p}_{i_k}$  такой, что  $\hat{p}_{i_1} \geq \hat{p}_{i_2} \geq \dots \geq \hat{p}_{i_k}$ . Выберем порог  $\delta > 0$  из тех соображений, что на практике, если  $p_{i_j} < \delta$ , то будем считать, что компонента смеси, соответствующая весу  $p_{i_j}$ , отсутствует. Положим

$$k = \max\{j : p_{i_j} \geq \delta\}.$$

Тогда в качестве начального приближения для EM-алгоритма возьмем параметры  $a_{i_j}$  и  $\sigma_{i_j}^2$ , соответствующие весам  $\hat{p}_{i_1}, \hat{p}_{i_2}, \dots, \hat{p}_{i_k}$ . Особо следует отметить, что при таком подходе число компонент смеси определяется автоматически по заданному порогу пренебрежимости  $\delta$ .

Обратим внимание, что в рамках рассмотренного в данном разделе подхода возможно классическое детерминистическое истолкование точности в отношении приближения параметров  $a_i$  и  $\sigma_i$ , при котором точность характеризуется *одним* числом, задаваемым шагом дискретной сетки, накидываемой на множества значений указанных параметров. В отношении же весовых параметров  $p_1, \dots, p_k$ , чтобы охарактеризовать точность приближения,

приходится пользоваться статистическим истолкованием, при котором одного числа недостаточно, а нужно задавать еще и надежность вывода (коэффициент доверия или уровень значимости).

## 5 Декомпозиция волатильности с помощью метода скользящего разделения смесей

Возможности описанных выше сеточных методов разделения конечных смесей нормальных законов будут проиллюстрированы на примере решения задачи декомпозиции волатильности (т. е. задачи разложения волатильности на компоненты) некоторых финансовых индексов.

Общая схема теоретического решения этой проблемы выглядит следующим образом (см. [2]).

- (i) Асимптотический подход, основанный на предельных теоремах для обобщенных дважды стохастических пуассоновских процессов как моделей неоднородных хаотических случайных блужданий, естественно приводит к заключению о том, что аппроксимации для распределений (логарифмов) приращений процессов эволюции финансовых индексов на интервалах времени умеренной длины следует искать в виде общих сдвиг/масштабных смесей нормальных законов. Более того, при этом смешивающий закон определяется накопленной (интегральной) интенсивностью потоков соответствующих информативных событий (элементарных скачков, «тиков»).
- (ii) Проблема статистической реконструкции распределений приращений упомянутых процессов (или их логарифмов) сводится к задаче статистического оценивания смешивающего распределения, которое является параметром этой задачи.
- (iii) В самой общей постановке задача статистического оценивания смешивающего распределения является некорректной, так как общие сдвиг/масштабные смеси нормальных законов не являются идентифицируемыми. Таким образом, в рамках общего принципа регуляризации некорректных задач исходная проблема заменяется задачей отыскания решения, наиболее близкого к истинному в классе конечных дискретных сдвиг/масштабных смесей нормальных законов. Эта «редуцированная» задача уже является корректной и имеет единственное решение, так как семейство конеч-

ных дискретных сдвиг/масштабных смесей нормальных законов идентифицируемо. Поскольку сдвиг/масштабные смеси нормальных законов обладают свойством устойчивости относительно смешивающего закона, эта замена оправдана и регулярна. При этом, зная оценки устойчивости, можно вычислить погрешности, образующиеся при замене исходной задачи редуцированной. При упомянутой регуляризации происходит автоматическое выделение типичных или более-менее устойчивых структур в эволюции рассматриваемых сложных систем.

(iv) Представление распределений (логарифмов) приращений процессов эволюции финансовых индексов в виде конечных сдвиг/масштабных смесей нормальных законов естественно приводит к многомерной интерпретации волатильности рассматриваемого процесса и к возможности разложения волатильности на динамическую и диффузионные компоненты. Действительно, если функция распределения логарифмического приращения  $Z$  некоторого финансового индекса имеет вид (1), то для нее справедливо представление

$$F(x) = P(Z < x) = \sum_{j=1}^k p_j \Phi\left(\frac{x - a_j}{\sigma_j}\right) = E\Phi\left(\frac{x - V}{U}\right),$$

где пара случайных величин  $U, V$  имеет дискретное распределение

$$P((U, V) = (\sigma_j, a_j)) = p_j, \quad j = 1, \dots, k.$$

Так что, как продемонстрировано в книге [2],

$$DZ = DV + EU^2, \quad (23)$$

причем

$$DV = \sum_{j=1}^k (a_j - \bar{a})^2 p_j, \quad EU^2 = \sum_{j=1}^k p_j \sigma_j^2, \quad (24)$$

где

$$\bar{a} = \sum_{j=1}^k a_j p_j.$$

Волатильность индекса естественно отождествить с величиной  $DZ$  (или  $\sqrt{DZ}$ ). При этом первое выражение в (24) зависит только от весов  $p_j$  и параметров положения (сдвига)  $a_j$

компонент, и потому характеризует ту часть волатильности, которая обусловлена наличием локальных трендов, т.е. «динамическую» компоненту волатильности, тогда как второе выражение в (24) зависит только от весов  $p_j$  и параметров масштаба («коэффициентов диффузии»)  $\sigma_j$  компонент и потому характеризует «чисто диффузионную» ее компоненту.

Если вспомнить традиционное одномерное представление о волатильности как о стандартном отклонении приращения процесса, то можно заметить, что разложение (23) уточняет это представление: волатильность процесса представляет собой корень квадратный из суммы двух компонент, первая из которых является характеристикой разбросанности локальных трендов, а вторая характеризует диффузию процесса. Если локальные тренды отсутствуют, то классическая волатильность равна корню квадратному из взвешенной суммы квадратов волатильностей компонент, причем веса компонент показывают важность соответствующей диффузионной компоненты.

(v) Статистические закономерности поведения рассматриваемых процессов, формализованные в пункте (i), изменяются во времени, результатом чего является отсутствие *универсального* смешивающего закона. Таким образом, для изучения динамики изменения статистических закономерностей в поведении исследуемого хаотического 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), масштаба (диффузии), а также соответствующие веса) оцениваются как функции времени. При этом естественно возникают задачи, связанные как с выбором подходящих методов оценивания параметров сдвига и масштаба, так и с выбором оптимальных параметров вычислительных процедур, реализующих эти методы: начального приближения, ширины скользящего интервала (окна), правила остановки и др.

(vi) Наконец, для адекватной интерпретации результатов и для идентификации феноменологически выделенных (статистически оцененных) компонент, т.е. для адекватного сопоставления статистически оцененных компонент с реальными процессами или явлениями, необходимо из многих возможных мо-

делей выбрать наиболее адекватную, например проверить, является выделенная динамическая компонента волатильности статистически значимой или нет.

## 6 Применение сеточных методов разделения смесей для декомпозиции волатильности конкретных финансовых индексов

В качестве исходных данных использовались минутные логарифмические приращения четырех биржевых индексов: AMEX, NASDAQ 100, NIKKEI 225 и S&P 40. В силу того что временные ряды логарифмических приращений лишены явной трендовой составляющей и их среднее близко к нулю, была взята нулевая сетка по параметрам  $a_i$ . При этом количество узлов сетки по параметрам  $\sigma_i$  взято равным  $K = 50$ . Для параметров  $\sigma_i$ ,  $i = \overline{1, 50}$ , было выбрано максимальное граничное значение, составляющее три стандартных отклонения по всему временному ряду. Таким образом, исходная задача оценки тройки параметров  $(a, \sigma, p)$  свелась к задаче подбора коэффициентов  $\tilde{p}_i$ , где  $i = \overline{1, 50}$ . При этом ширина скользящего окна  $n$  взята равной 300 отсчетам (что соответствует пяти часам). Задача оценивания весов последовательно решается для каждого положения скользящего окна по выборке (отрезку исходного ряда), соответствующей данному положению окна.

Оценки весов  $p_i$  искались как решения двух задач линейного программирования, к которым сводятся задачи минимизации  $\sup$ -нормы и  $L_1$ -нормы невязки между эмпирической и теоретической функциями распределения (см. разд. 4).

Для решения использовалась стандартная процедура `linprog` пакета MATLAB 7.0.

Результаты представлены на графиках, приведенных на рис. 1–4. На каждом графике горизонтальная ось — это ось времени: каждая точка на горизонтальной оси соответствует конкретному значению правого конца скользящего интервала времени, по которому (т. е. по наблюдениям, попавшим в который) вычисляются оценки весов. Вертикальная ось — это ось значений параметров  $\sigma_j$ . Веса компонент смеси, соответствующих конкретным значениям параметров  $\sigma_j$ , показаны оттенками серого цвета. Чем прямоугольник темнее, тем вес больше.

Для сравнения на рисунках также представлены результаты решения аналогичной задачи с по-

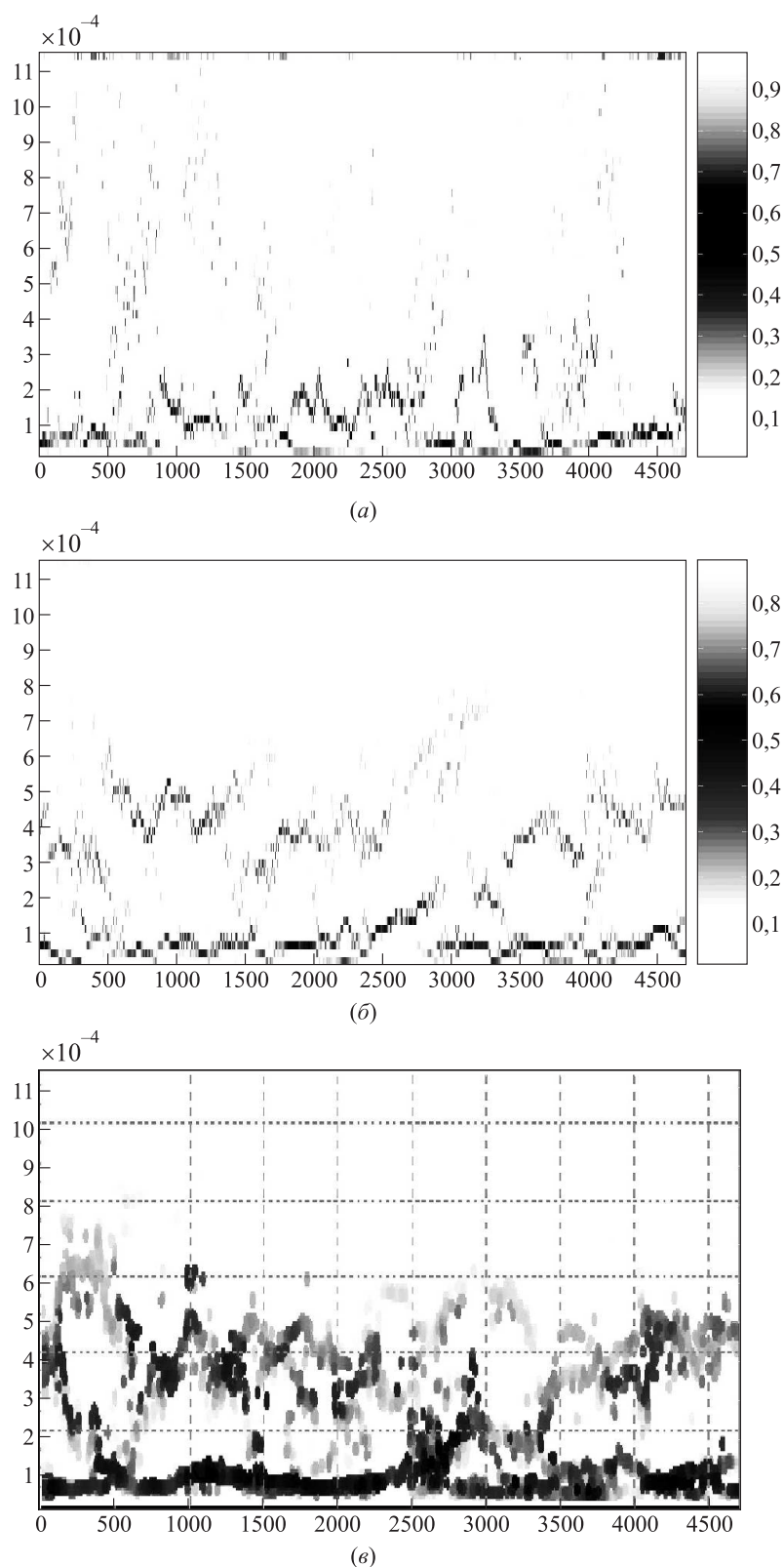
мощью EM-алгоритма. Этот алгоритм реализован программой ZHPlot, разработанной Ю. В. Жуковым.

### 6.1 Декомпозиция волатильности индекса AMEX

Американская фондовая биржа (American Stock Exchange, AMEX) — одна из крупнейших региональных бирж США, расположена в Нью-Йорке. Ведет свою историю с 1911 г., когда нью-йоркские уличные торговцы акциями объединились в ассоциацию New York Curb Market Association. В 1953 г. получила нынешнее название. После обвала рынка в 1987 г. биржа ужесточила правила торгов, подняв уровень биржевой маржи (инструмент гарантийного обеспечения при торгах) и установив максимально допустимый уровень падения цен на акции, после которого торги прекращаются. В начале 1990-х гг. AMEX первой в мире ввела систему электронных торгов с использованием беспроводных терминалов. В 1998 г. биржа была куплена NASDAQ, однако в 2004 г. участники AMEX выкупили площадку и остаются до сегодняшнего дня акционерами. Основной используемый индекс — AMEX Composite — обобщенный индекс, отражающий совокупную рыночную стоимость всех компонентов рынка AMEX (обыкновенных акций, депозитарных расписок и пр.) по отношению к совокупной рыночной стоимости этого рынка по состоянию на 29.12.1995.

В нашем исследовании рассматриваются ежеминутные данные с 04.02.2008 по 03.03.2008, т. е. длина анализируемого временного ряда превышает 7000 значений (отсчетов). Результаты декомпозиции волатильности индекса AMEX по минутным данным представлены на рис. 1.

**Экономическая интерпретация поведения волатильности индекса AMEX.** Данный индекс является низковолатильным. Наиболее весомые компоненты не принимают значения больше 1,5. Первый заметный рост волатильности наблюдается в среду 06.02, когда рынок акций США продемонстрировал самое резкое почти за год снижение во вторник на фоне данных, указавших на значительный спад активности в секторе услуг США и предупреждения рейтингового агентства Standard & Poor's о том, что оно может понизить кредитные рейтинги банков. Опасения по поводу возможности рецессии в США оказали давление на акции различных секторов — от энергетического до телекоммуникационного. Основной тон на фондовом рынке задавали данные Института менеджеров по поставкам об активности в секторе услуг в США за январь. Индекс активности снизился до минимального с октября



**Рис. 1** Декомпозиция волатильности индекса AMEX сеточным методом с минимизацией sup-нормы невязки между эмпирической и теоретической функциями распределения (а), сеточным методом с минимизацией  $L_1$ -нормы невязки между эмпирической и теоретической функциями распределения (б) и EM-алгоритмом (в)

2001 г. уровня, усилив беспокойство по поводу рецессии. В дальнейшем ситуация была достаточно стабильной, но 11.02 и 12.02 волатильность стала опять расти, потому что американские фондовые рынки в понедельник 11.02 выросли благодаря акциям компаний высокотехнологического сектора, которые помогли отыграть потери, понесенные после открытия торгов. Акции банков упали из-за заявлений министров финансов стран Большой семерки, которые посчитали, что волнения на финансовых рынках ударят по темпам роста мировой экономики. В итоговом коммюнике, опубликованном по завершении встречи министров финансов G7, говорится, что с момента предыдущей встречи в октябре 2007 г. перспективы экономического роста в ведущих индустриальных державах ухудшились, хотя основы и остаются надежными и американская экономика все еще может избежать рецессии. Ралли на рынке нефти благоприятно сказалось на акциях нефтяных компаний. Цена на нефть резко выросла после того, как президент Венесуэлы Уго Чавес пригрозил прекратить поставки нефти в Соединенные Штаты. Интересная картина наблюдается 14.02, когда одна из компонент остается на уровне близком к 0, а вторая начинает расти. Ее можно соотнести с отраслевой (промышленной) компонентой волатильности, потому что 14.02 наблюдался рост американских фондовых индексов, которые поднялись благодаря хорошим данным о розничных продажах в Соединенных Штатах. Министерство торговли в своем отчете в среду 13.02 сообщило, что объем розничных продаж в январе вырос на 0,3%, превзойдя прогнозы аналитиков благодаря неожиданному увеличению продаж автомобилей и росту цен на бензин. Экономисты ожидали, что розничные продажи в США снизятся на 0,3%. За прошедший год объем розничных продаж увеличился на 3,9%. Рост розничных продаж снизил опасения инвесторов по поводу возможного сокращения потребительских расходов в США.

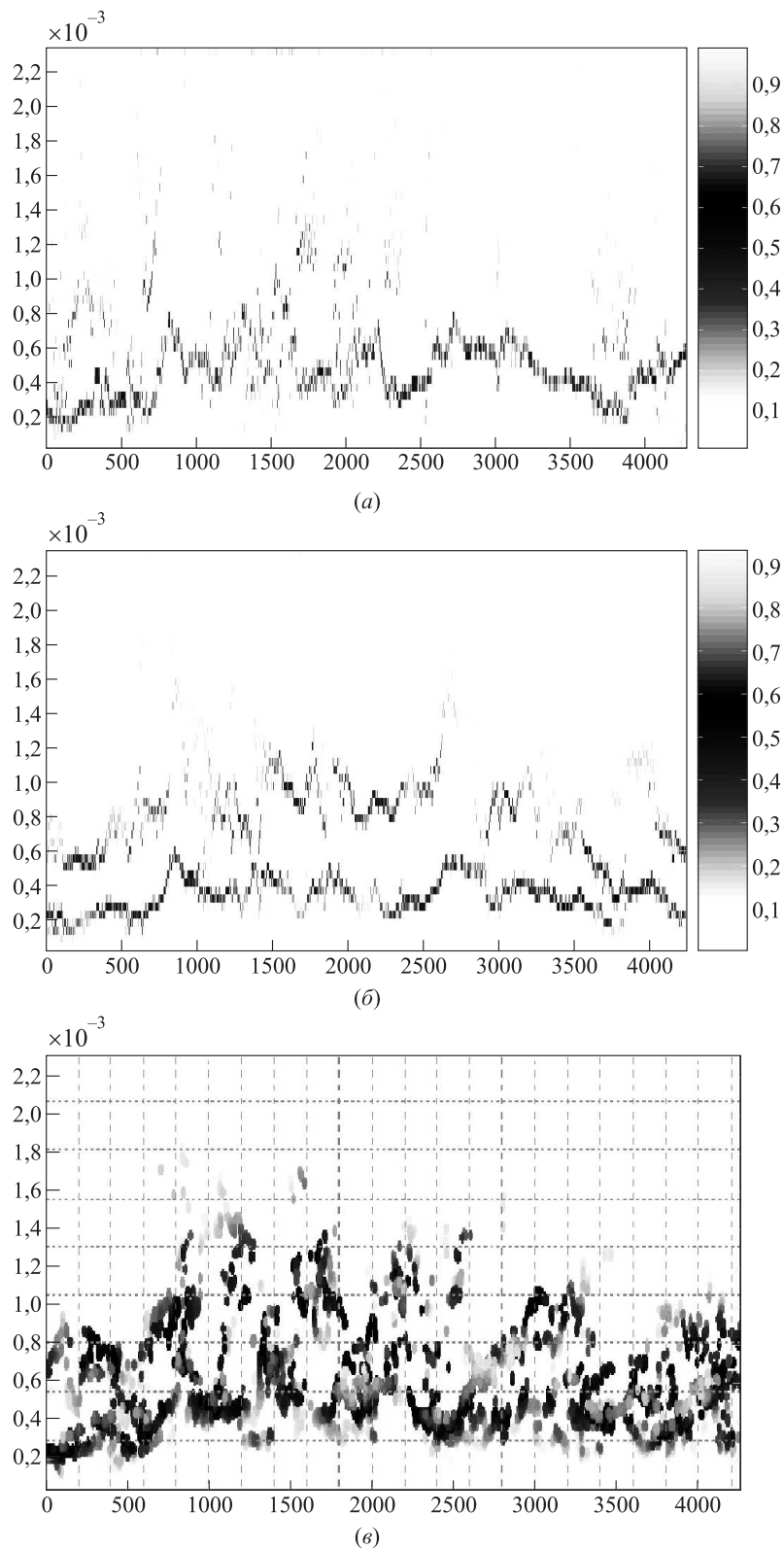
## 6.2 Декомпозиция волатильности индекса CAC 40

Индекс CAC 40 (Compagnie des Agents de Change — Ассоциация французских фондовых брокеров) — один из основных французских индексов, включающий в листинг сорок крупнейших французских корпораций. Индекс CAC 40 вычисляется как среднее взвешенное по капитализации значение цен акций 40 крупнейших компаний Франции, акции которых торгуются на бирже Euronext Paris. Начальное значение индекса — 1000 пунктов — было установлено 31.12.1987. Начиная с 01.12.2003

при подсчете капитализации учитываются лишь акции, находящиеся в свободном обращении. Индекс CAC 40 вычисляется каждые 30 с в рабочие дни биржи с 9:00 до 17:30 по центрально-европейскому времени.

В данной статье анализируются ежеминутные данные индекса CAC 40 в промежутке между 04.02.2008 и 15.02.2008, т.е. анализируется более 4500 значений. Результаты декомпозиции волатильности индекса CAC 40 по минутным данным представлены на рис. 2.

**Экономическая интерпретация поведения волатильности индекса CAC 40.** Рассмотрим факторы, влиявшие на всплески волатильности. Четвертого февраля 2008 г. наблюдался спад волатильности. Это происходило на фоне трех сессий роста основных европейских индексов благодаря новостям из банковского и добывающего секторов. Но уже 05.02 волатильность пошла резко вверх из-за того, что бумаги добывающих компаний, поднимавшиеся в последнее время, опустились на фоне противоречивой динамики цен на металлы. Акции европейских банков подешевели вслед за бумагами американских финансовых компаний, которые снизились на торгах в США днем ранее. При этом компонента высокой волатильности теряет свой вес. Седьмого февраля наблюдается очередной взрыв волатильности. Это происходит из-за того, что большинство европейских фондовых индексов в этот четверг упало под давлением неблагоприятных прогнозов крупных европейских и американских компаний, которые стали еще одним признаком замедления темпов роста мировой экономики. Французский CAC 40 закрылся со снижением на 92,63 пункта, или на 1,91%, на отметке 4723,80. Перед выходными волатильность стабилизируется, на французском рынке наблюдается снижение индекса, несмотря на то что большинство европейских рынков выросли благодаря акциям сырьевых компаний, подорожавшим на фоне роста цен на нефть и металлы. Но это связано с падением акций Societe Generale SA, второго крупнейшего банка во Франции, которые снизились на 3,5% на фоне новых обстоятельств расследования одной из крупнейших афер в банковском секторе. При этом компонент волатильности с меньшими весами становится больше. Следующий взлет волатильности наблюдается 12.02, когда на европейских рынках произошел самый значительный рост за прошедшие две недели благодаря предложению инвестора Уоррена Баффетта (Warren Buffett) перестраховать муниципальные облигации на общую сумму 800 млрд долл. Как отмечают аналитики, решение такого влиятельного инвестора взять на себя риски по страхованию облигаций вселило опти-



**Рис. 2** Декомпозиция волатильности индекса S&P 40 сеточным методом с минимизацией  $\sup$ -нормы невязки между эмпирической и теоретической функциями распределения (а), сеточным методом с минимизацией  $L_1$ -нормы невязки между эмпирической и теоретической функциями распределения (б) и EM-алгоритмом (в)

мизм в игроков на фондовом рынке. Предложение Баффетта, в случае его принятия указанными компаниями, смогло бы разрядить ситуацию на рынке и снять угрозу серьезных потерь в банковском секторе. Рост цен на медь, цинк и другие металлы в Лондоне благоприятно сказался на котировках акций горнодобывающих компаний. В дальнейшем наблюдается стабилизация ситуации и небольшой рост индекса. Но 15.02 происходит очередной спад в экономике из-за усилившихся опасений того, что банки и дальше будут нести серьезные потери, а в американской экономике начнется рецессия. Таким образом, подобный вид графика (скачкообразный) четко отражает тенденции на французском фондовом рынке — фазы резких взлетов и падений, которые чередуются с фазами невысокой волатильности.

### 6.3 Декомпозиция волатильности индекса NASDAQ

Индекс NASDAQ — индекс внебиржевого рынка, публикуемый Национальной ассоциацией дилеров по ценным бумагам (National Association of Securities Dealers, NASD) и основанный на ее котировках. Сводный индекс NASDAQ (NASD Automated Quotations) строится на основе взвешенной рыночной стоимости акций эмитентов, специализирующихся в области высоких технологий. Это означает, что ценная бумага каждой компании влияет на индекс пропорционально рыночной стоимости компании. Торговля на первом в мире электронном рынке NASDAQ началась 08.02.1971. А поскольку на NASDAQ котируются акции не только Hi-Tech компаний, то возникла целая система индексов, каждый из которых отражает ситуацию в соответствующей отрасли экономики. Существует 13 различных индексов, в основе которых лежат котировки ценных бумаг, торгуемых в электронной системе NASDAQ. Индекс NASDAQ 100 составляют крупнейшие нефинансовые американские и иностранные компании, входящие в листинг NASDAQ, — производители компьютерного «железа» и «софта», телекоммуникационные компании, розничные и оптовые торговые фирмы, а также компании, связанные с биотехнологиями. Первоначально главным критерием попадания в верхние строчки индекса был вес компании, выраженный в долларах США, — рыночная капитализация, но в декабре 1998 г. была проделана хитрая «перебалансировка», по которой рыночная капитализация каждой компании уже умножалась на некий ежеквартально пересматриваемый весовой коэффициент. Результатом этого новшества, по мнению

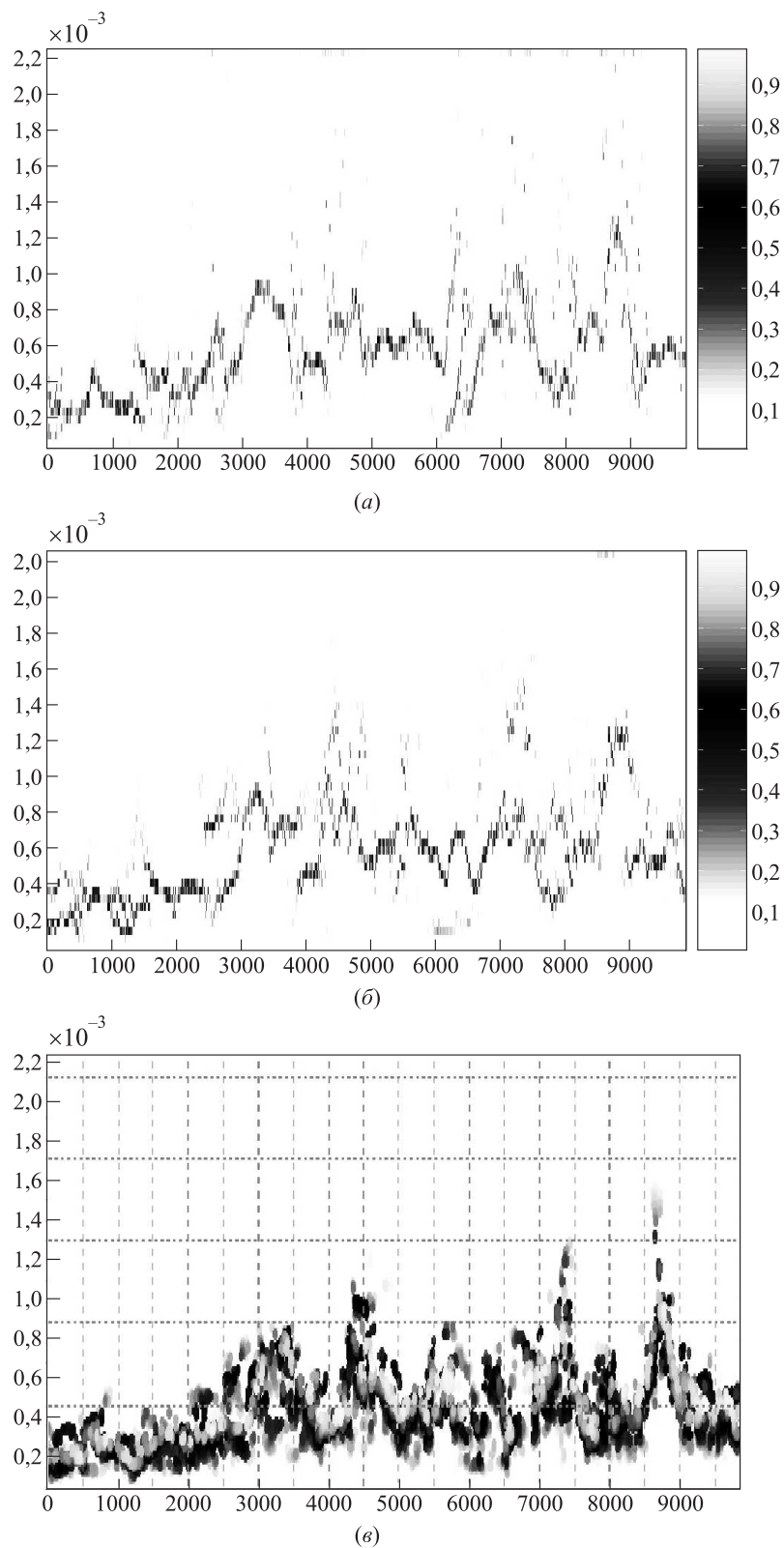
наздаковцев, должна была стать более совершенная диверсификация. В итоге, например, Microsoft уступает Cisco Systems, хотя по капитализации Microsoft в несколько раз превосходит последнюю.

Рассматриваются ежеминутные данные в промежутке с 16.07.07 по 22.08.07, т. е. анализируется более 10 000 данных. Результаты декомпозиции волатильности индекса NASDAQ 100 по минутным данным представлены на рис. 3.

**Экономическая интерпретация поведения волатильности индекса NASDAQ.** В промежуток времени с 16.07 по 24.07 наблюдается некоторая стабильность на рынке. Но во вторник 25.07 американские фондовые рынки рухнули под давлением разочаровавших квартальных отчетов и признаков распространения кризиса ипотечного кредитования на клиентов с высоким рейтингом. Вторник стал самым плохим днем для американских фондовых рынков за прошедшие четыре месяца. Основным фактором, спровоцировавшим такое резкое падение индексов, стал квартальный отчет компании Countrywide Financial Corp. (CFC), из которого стало понятно, что кризис в секторе ипотечного кредитования продолжает разрастаться, и в категорию неплательщиков стали попадать и клиенты с высоким рейтингом. Это спровоцировало рост волатильности. А последующий рост цен на нефть вызвал небольшой рост фондовых рынков, что обеспечило спад волатильности. Следующий значительный подъем волатильности наблюдается с 07.07, что объясняется ростом фондовых рынков США благодаря скупке инвесторами акций финансовых компаний, снизившихся в конце прошлой недели на фоне роста опасений касательно стабильности кредитных рынков. На следующий день акции электроэнергетических, финансовых и нефтедобывающих компаний возглавили рост на американских фондовых рынках после опубликования доклада Федеральной резервной системы. Вслед за подъемом волатильности, как и обычно, наблюдается ее стабилизация, так как продолжается рост рынка, но он не является значительным. Очередной ее всплеск можно наблюдать в среду 15.08. Во вторник американские фондовые индексы испытали значительное снижение на фоне вновь появившихся опасений относительно слабости кредитной сферы, предполагаемого снижения потребительских расходов и новых свидетельств кризиса ипотечного рынка. NASDAQ потерял 1,7%.

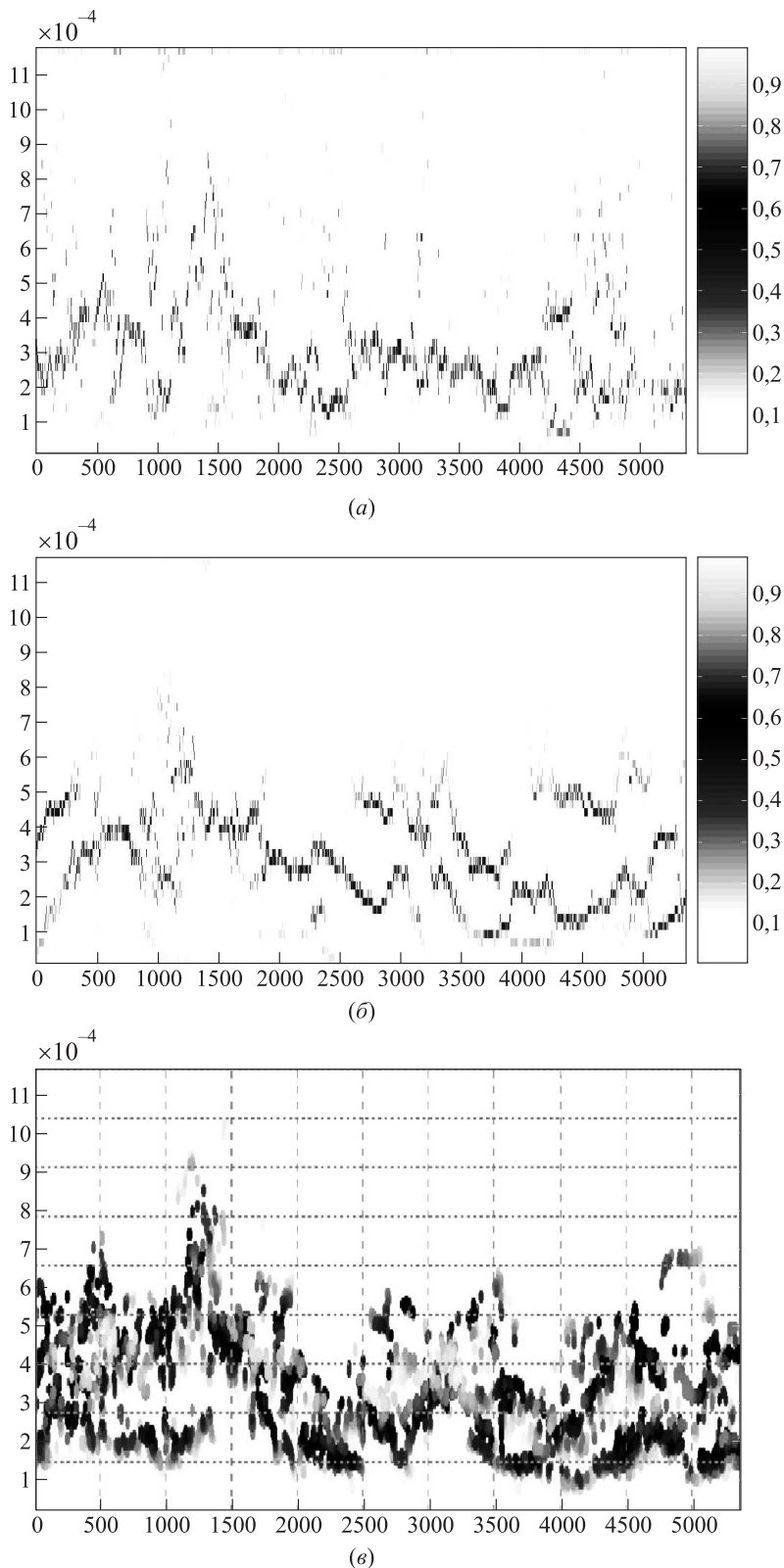
### 6.4 Декомпозиция волатильности индекса NIKKEI

Индекс NIKKEI 225 — наиболее широко используемый и изучаемый показатель на Японском



**Рис. 3** Декомпозиция волатильности индекса NASDAQ сеточным методом с минимизацией  $\sup$ -нормы невязки между эмпирической и теоретической функциями распределения (а), сеточным методом с минимизацией  $L_1$ -нормы невязки между эмпирической и теоретической функциями распределения (б) и EM-алгоритмом (в)





**Рис. 4** Декомпозиция волатильности индекса NIKKEI сеточным методом с минимизацией sup-нормы невязки между эмпирической и теоретической функциями распределения (а), сеточным методом с минимизацией  $L_1$ -нормы невязки между эмпирической и теоретической функциями распределения (б) и EM-алгоритмом (в)

фондовом рынке. Индекс NIKKEI 225 впервые был опубликован в 1950 г. Токийской фондовой биржей под названием TSE Adjusted Stock Price Average. С 1970 г. индекс вычисляется японской газетой Nihon Keizai Shimbun. Новое название индекса произошло от сокращенного названия газеты — Nikkei. Он вычисляется как среднее взвешенное значение цен акций 225 компаний, наиболее активно торгуемых в первой секции Токийской фондовой биржи. Эти акции выбраны из 450 наиболее ликвидных компаний. Они разделены на шесть отраслевых секторов и отражают различные тенденции Японского рынка. Список компаний, охваченных индексом NIKKEI 225, пересматривается как минимум раз в год, в октябре.

Рассматриваются ежеминутные данные с 15.01.2008 по 13.02.2008, т. е. анализируется более 5500 данных. Результаты декомпозиции волатильности индекса NIKKEI по минутным данным представлены на рис. 4.

**Экономическая интерпретация поведения волатильности индекса NIKKEI.** В начале можно наблюдать резкий взлет обеих компонент волатильности, что связано с ожиданием негативной финансовой отчетности американских банков, которая может еще сильнее ухудшить перспективы самой крупной экономики в мире. Индекс токийской биржи NIKKEI к завершению торгов снизился на 0,98% до 13.972,63 пункта, опустившись ниже отметки 14.000 пунктов впервые с ноября 2005 г. Опасения по поводу японской экономики усилились после того, как глава Банка Японии Тосихико Фукуи предупредил, что темп экономического роста может замедлиться на некоторое время, несмотря на продолжающееся развитие экономики. Рынок ждал публикации финансовых результатов американского банка Citigroup Inc позднее во вторник. Новый генеральный директор Citi Викрам Пандит, вероятно, сообщит о значительном сокращении дивидендов, об увеличении капитала как минимум на 10 млрд долл., о списании 20 млрд долл., а также о сокращении более 20 000 рабочих мест, сообщила газета Wall Street Journal в понедельник. Дальнейший спад мировой экономики сопровождался стабильной, но довольно высокой волатильностью, а на фоне сильнейшей регрессии на японской фондовой бирже, связанной с опасением замедления двух крупнейших в мире экономик — США и Японии, выделились две противоположные компоненты, одна из которых пошла вверх, а другая — вниз. К 23.01.08 сильное падение японских фондовых индексов, продолжавшееся два предыдущих дня, в среду прекратилось, и основные индикаторы вышли, наконец, в положительную зону, что привело к тому, что самые значимые компоненты

волатильности сошлись в одну. В последующие дни происходил значительный рост индексов как в США и Европе, так и в Японии, что происходило на фоне объявления мер помощи экономике США, неожиданного улучшения на американском рынке труда, а также благодаря ожиданиям очередного сокращения процентных ставок Федеральной резервной системой с целью избежать обращения крупнейшей в мире экономики в рецессию. Рынки акций Азии росли вслед за Уолл-стрит после того, как президент США Джордж Буш и лидеры Конгресса в четверг приняли решение о выделении 150 млрд долл. для предоставления налоговых льгот с целью поддержать экономику, которой грозит рецессия и спад ВВП два и более кварталов подряд. Акции банковского сектора оправались после новости о мошенничестве трейдера Societe Generale, из-за которого компания потеряла 7 млрд долл. Индекс токийской биржи NIKKEI вырос на 4,10% до 13.629,16 пункта, продолжив восстановление после резкого падения в начале недели.

Если рассматривать результаты, получающиеся сеточным методом с минимизацией  $sup$ -нормы, то в последующие дни волатильность в среднем находится на одном и том же уровне. Ее колебания происходят внутри одного дня. Но сеточным методом с минимизацией  $L_1$ -нормы можно выделить две компоненты, которые движутся в унисон, при этом одна компонента — высоковолатильная, ее значения превосходят уровень 5, а значения второй компоненты волатильности флуктуируют вокруг уровня 2.

## 7 Выводы

Из результатов анализа волатильности индексов AMEX, S&P 500, NASDAQ и NIKKEI, приведенных на рис. 1–4, можно сделать несколько выводов.

1. Наличие нескольких выделенных компонент волатильности, по-видимому, соответствует наличию различных уровней активности разных секторов экономики в разные моменты времени.
2. Обращает на себя внимание, что из 50 возможных значений диффузионной компоненты волатильности значимыми (ненулевыми) оказываются веса лишь двух–четырех компонент. Остальные автоматически обнуляются.
3. Результат применения EM-алгоритма оказывается практически аналогичен результату применения сеточного метода с минимизацией  $L_1$ -нормы.
4. Минимизация  $sup$ -нормы приводит к выделению меньшего числа компонент волатильно-

сти, нежели минимизация  $L_1$ -нормы. Это, по-видимому, означает, что  $L_1$ -норма более чувствительна к расхождениям между данными и моделью. В то же время минимизация  $L_1$ -нормы дает более четкие «портреты волатильности», менее зашумленные по сравнению с теми, что получены минимизацией  $\text{sup}$ -нормы.

Авторы считают своим приятным долгом выразить признательность Ф. П. Васильеву за полезное обсуждение вопросов, затронутых в данной статье.

## Литература

1. *Королёв В. Ю.* EM-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор. — М.: Изд-во ИПИРАН, 2007.
2. *Королёв В. Ю.* Вероятностно-статистический анализ хаотических процессов с помощью смешанных гауссовских моделей. Декомпозиция волатильности финансовых индексов и турбулентной плазмы. — М.: Изд-во ИПИРАН, 2007.
3. *Себер Дж.* Линейный регрессионный анализ. — М.: Мир, 1980.
4. *Waterman M. S.* A restricted least squares problem // *Technometrics*, 1974. Vol. 16. No. 1. P. 135–136.
5. *Judge G. G., Takayama T.* Inequality restrictions in regression analysis // *J. of American Statistical Association*, 1966. Vol. 61. No. 1. P. 166–181.
6. *Королёв В. Ю., Ломской В. А., Пресняков Н. Н., Рэй М.* Анализ компонент волатильности с помощью метода скользящего разделения смесей // *Системы и средства информатики. Спец. вып.* — М.: ИПИРАН, 2005. С. 180–206.
7. *Васильев Ф. П., Иваницкий А. Ю.* Линейное программирование. — М.: Факториал Пресс, 2003.
8. *Ашманов С. А., Тимохов А. В.* Теория оптимизации в задачах и упражнениях. — М.: Наука, 1991.

# НЕКОТОРЫЕ СТАТИСТИЧЕСКИЕ ЗАДАЧИ, СВЯЗАННЫЕ С РАСПРЕДЕЛЕНИЕМ ЛАПЛАСА\*

В. Е. Бенинг<sup>1</sup>, В. Ю. Королёв<sup>2</sup>

**Аннотация:** В работе развивается подход, предложенный в статье [1]. Обосновывается естественность возникновения распределения Лапласа в задачах теории вероятностей и математической статистики. В качестве статистической иллюстрации рассмотрено приложение распределения Лапласа к задачам асимптотической проверки гипотез.

**Ключевые слова:** распределение Лапласа; проверка статистических гипотез; функция мощности; асимптотически наиболее мощные критерии

## 1 Введение

Классическое распределение Лапласа с нулевым средним и дисперсией  $\sigma^2$  было введено П. С. Лапласом в 1774 г. [2]. С тех пор, наряду с нормальным, оно стало одной из наиболее активно используемых симметричных вероятностных моделей. Это распределение задается характеристической функцией

$$f(s) = \frac{2}{2 + \sigma^2 s^2}, \quad s \in \mathbb{R}^1,$$

или плотностью

$$\ell(x) = \frac{1}{\sigma\sqrt{2}} \exp\left\{-\frac{\sqrt{2}|x|}{\sigma}\right\}, \quad \sigma > 0, x \in \mathbb{R}^1.$$

Популярность распределения Лапласа как математической (вероятностной) модели обусловлена тем, что его хвосты тяжелее, чем у нормального распределения (см., например, книги [3–5], где описывается роль распределения Лапласа в методах робастного оценивания; работу [6], где обосновывается целесообразность использования распределения Лапласа как модели распределения погрешностей измерений в энергетике; статью [7], посвященную применению распределения Лапласа для моделирования ошибок в навигации; работу [8], в которой распределение Лапласа применяется в метеорологии; статьи [9, 10], посвященные применению распределения Лапласа в управлении запасами и радиоэлектронике). Во многих работах описано успешное применение распределения Лапласа для моделирования распределения приращений логарифмов финансовых индексов [11–13].

В работах [14–17] распределение Лапласа используется как модель логарифма доходов фирм и индивидуумов. Многие работы посвящены применению распределения Лапласа для моделирования распределения логарифма размера частиц при дроблении [18–20]. Наконец, распределение Лапласа применяется при моделировании статистических закономерностей поведения некоторых характеристик атмосферной [21] и плазменной [22] турбулентности. В [23, 24] можно найти дальнейшие ссылки на работы, в которых описывается применение распределения Лапласа к решению прикладных задач в самых разнообразных областях.

## 2 Распределение Лапласа как симметризация

### 2.1 Распределение Лапласа как сверточная симметризация

Пусть  $X$  и  $X'$  — независимые случайные величины с одной и той же функцией распределения  $F(x)$ . Характеристическую функцию, соответствующую функции распределения  $F(x)$ , обозначим  $f(s)$ . Тогда характеристическая функция случайной величины  $X^{(s)} \equiv X - X'$  имеет вид

$$\begin{aligned} E \exp\{isX^{(s)}\} &= E \exp\{is(X - X')\} = \\ &= f(s)f(-s) = f(s)\overline{f(s)} = |f(s)|^2, \quad s \in \mathbb{R}^1 \end{aligned}$$

(черта сверху означает комплексное сопряжение).

\* Работа выполнена при поддержке РФФИ, гранты 08-01-00567, 08-01-00563 и 08-07-00152.

<sup>1</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, Институт проблем информатики Российской академии наук, bening@yandex.ru

<sup>2</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, Институт проблем информатики Российской академии наук, vkorolev@comtv.ru

Последняя характеристическая функция вещественна, следовательно, распределение, ей соответствующее, является симметричным в том смысле, что для любого  $x > 0$

$$P(X^{(s)} < -x) = P(X^{(s)} > x).$$

Распределение, соответствующее характеристической функции  $|f(s)|^2$ , называется *сверточной симметризацией* распределения  $F(x)$ , а случайная величина  $X^{(s)}$ , соответственно, называется *сверточной симметризацией* случайной величины  $X$ .

Рассмотрим случайную величину  $X$  со стандартным показательным распределением

$$P(X < x) = (1 - e^{-x})\mathbf{1}(x \geq 0) \equiv E(x). \quad (1)$$

Здесь символом  $\mathbf{1}(A)$  обозначается индикаторная функция множества  $A$ . Как известно, функции распределения  $E(x)$  соответствует характеристическая функция

$$f(s) = \frac{1}{1 - is}, \quad s \in \mathbb{R}^1.$$

Тогда в соответствии с приведенным выше определением сверточной симметризации функции распределения  $E(x)$  соответствует характеристическая функция

$$|f(s)|^2 = \frac{1}{1 - is} \frac{1}{1 + is} = \frac{1}{1 + s^2},$$

что совпадает с характеристической функцией распределения Лапласа с  $\sigma^2 = 2$ . Таким образом, распределение Лапласа является сверточной симметризацией экспоненциального (показательного) распределения.

## 2.2 Распределение Лапласа как рандомизационная симметризация

Рассмотрим произвольную случайную величину  $X$  и обозначим ее функцию распределения  $F(x)$ . Назовем *рандомизационной симметризацией* случайной величины  $X$  случайную величину  $\tilde{X}$  такую, что

$$\tilde{X} = \begin{cases} X & \text{с вероятностью } 0,5, \\ -X & \text{с вероятностью } 0,5. \end{cases}$$

Очевидно, что

$$\begin{aligned} P(\tilde{X} < x) &= \frac{1}{2} P(X < x) + \frac{1}{2} P(-X < x) = \\ &= \frac{1}{2} [F(x) + 1 - F(-x + 0)] = \\ &= \frac{1}{2} + \frac{1}{2} [F(x) - F(-x + 0)]. \end{aligned}$$

Если при этом случайная величина  $X$  абсолютно непрерывна, то, обозначив плотности случайных величин  $X$  и  $\tilde{X}$  соответственно  $p(x)$  и  $q(x)$ , из последнего равенства получим соотношение

$$q(x) = \frac{1}{2} [p(x) + p(-x)].$$

Более того, если  $P(X \geq 0) = 1$ , то

$$P(\tilde{X} < x) = \begin{cases} \frac{1}{2} [1 + F(x)], & \text{если } x > 0, \\ \frac{1}{2} [1 - F(-x + 0)], & \text{если } x \leq 0, \end{cases}$$

и

$$q(x) = \frac{1}{2} p(|x|).$$

По аналогии функцию распределения  $\tilde{F}(x) = P(\tilde{X} < x)$  будем называть *рандомизационной симметризацией* функции распределения  $F(x)$ .

Рандомизационную симметризацию  $\tilde{X}$  случайной величины  $X$  удобно интерпретировать следующим образом. Пусть  $Z$  — независимая от  $X$  случайная величина такая, что

$$P(Z = 1) = 1 - P(Z = -1) = \frac{1}{2}.$$

Тогда

$$\tilde{X} = XZ.$$

Если  $f(s)$  — характеристическая функция случайной величины  $X$ , то

$$\begin{aligned} E \exp \{is\tilde{X}\} &= \frac{1}{2} f(s) + \frac{1}{2} f(-s) = \\ &= \frac{1}{2} [E \cos sX + iE \sin sX + E \cos sX - iE \sin sX] = \\ &= E \cos sX = \mathbf{Re} f(s). \end{aligned}$$

Рассмотрим теперь характеристическую функцию стандартного экспоненциального распределения, для которой справедливо представление

$$f(s) = \frac{1}{1 - is} = \frac{1 + is}{(1 - is)(1 + is)} = \frac{1 + is}{1 + s^2}.$$

Стало быть,

$$\mathbf{Re} f(s) = \frac{1}{1 + s^2}.$$

Но последняя функция есть не что иное как характеристическая функция распределения Лапласа с дисперсией  $\sigma^2 = 2$ . Поэтому, так как в данном случае  $\mathbf{Re} f(s) = |f(s)|^2$ , рандомизационная симметризация показательного распределения совпадает с его сверточной симметризацией, которая, как уже было показано, представляет собой распределение Лапласа.

### 3 Распределение Лапласа как смесь

В этом разделе будет показано, что распределение Лапласа допускает несколько представлений в виде «масштабной» смеси некоторых хорошо известных распределений вероятностей. Стандартную нормальную функцию распределения будем обозначать  $\Phi(x)$ ,

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

**Лемма 3.1.** *Распределение Лапласа является масштабной смесью нормальных законов с нулевым средним при экспоненциальном смешивающем распределении:*

$$\Lambda(x) = \int_0^{\infty} \Phi\left(\frac{x}{\sqrt{z}}\right) dE(z), \quad x \in \mathbb{R}^1.$$

**Доказательство.** Имеем

$$\begin{aligned} \int_0^{\infty} \Phi\left(\frac{x}{\sqrt{z}}\right) dE(z) &= \\ &= \int_0^{\infty} \left[ \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{x/\sqrt{z}} e^{-u^2/2} du \right] e^{-z} dz = \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \int_0^{x/\sqrt{z}} \exp\left\{-\frac{u^2}{2} - z\right\} du dz = \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \int_0^{x^2/u^2} e^{-z} dz e^{-u^2/2} du = \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \left(1 - \exp\left\{-\frac{x^2}{u^2}\right\}\right) e^{-u^2/2} du = \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left\{-\frac{u^2}{2} - \frac{x^2}{u^2}\right\} du = \\ &= \begin{cases} \frac{1}{2} e^{\sqrt{2}x}, & x \leq 0, \\ 1 - \frac{1}{2} e^{-\sqrt{2}x}, & x > 0 \end{cases} \end{aligned}$$

(см., например, [25], формула 3.335). Но в правой части стоит не что иное как функция распределения Лапласа с параметром  $\sigma = 1$ , которому соответствует плотность

$$\ell(x) = \frac{1}{\sqrt{2}} \exp\{-\sqrt{2}|x|\}, \quad x \in \mathbb{R}^1.$$

Лемма доказана.

Из леммы 3.1, безграничной делимости показательного закона и хорошо известного утверждения о том, что масштабные смеси нормальных законов безгранично делимы, если безгранично делимы смешивающие распределения (см., например, [26]), вытекает безграничная делимость распределения Лапласа.

Пусть  $W$  — случайная величина, имеющая распределение Лапласа,  $X$  и  $U$  — случайные величины, имеющие соответственно стандартное нормальное и стандартное показательное распределения. Из леммы 3.1 вытекает следующая факторизация случайной величины  $W$ :

$$W \stackrel{d}{=} X\sqrt{U}, \quad (2)$$

где случайные величины  $X$  и  $U$  независимы.

Пусть  $G(x)$  — функция распределения максимума стандартного винеровского процесса на единичном отрезке:

$$G(x) = 2\Phi(\max\{0, x\}) - 1, \quad x \in \mathbb{R}^1.$$

Несложно видеть, что  $G(x) = P(|X| < x)$ , где  $X$  — случайная величина, имеющая стандартное нормальное распределение.

Легко убедиться, что рандомизационная симметризация функции распределения  $G(x)$  совпадает со стандартным нормальным распределением:

$$\tilde{G}(x) = \Phi(x). \quad (3)$$

Пусть  $Z$  — случайная величина, принимающая каждое из значений 1 и  $-1$  с вероятностью  $1/2$ . Тогда в терминах случайных величин  $X$  и  $Z$  соотношение (3) принимает вид

$$X \stackrel{d}{=} |X|Z, \quad (4)$$

где случайные величины в правой части независимы.

Далее, пусть  $X$  и  $U$  — независимые случайные величины, имеющие соответственно стандартное нормальное и стандартное показательное распределения. Тогда для  $x > 0$

$$\begin{aligned} P(|X|\sqrt{U} < x) &= EG\left(\frac{x}{\sqrt{U}}\right) = 2E\Phi\left(\frac{x}{\sqrt{U}}\right) - 1 = \\ &= 2\left(1 - \frac{1}{2}e^{-\sqrt{2}x}\right) - 1 = 1 - e^{-\sqrt{2}x} = P(U < \sqrt{2}x). \end{aligned}$$

Отсюда вытекает следующая «рекуррентная» факторизация случайной величины  $U$  со стандартным показательным распределением:

$$U \stackrel{d}{=} \sqrt{2U}|X|. \quad (5)$$

Из соотношений (2)–(5) получается представление

$$W \stackrel{d}{=} |X|Z\sqrt{U}. \quad (6)$$

Применяя в представлении (6) рекурсию (5), окончательно получаем следующее утверждение.

**Теорема 3.1.** *Справедливы представления*

$$W \stackrel{d}{=} Z\sqrt{U}|X|$$

и для любого натурального  $k \geq 2$

$$W \stackrel{d}{=} 2^{(2^{k-1}-1)/2^k} ZU^{1/2^k} \prod_{m=1}^k |X_m|^{1/2^{m-1}},$$

где случайные величины  $X, X_1, X_2, \dots$  имеют стандартное нормальное распределение, причем все случайные величины в правых частях независимы.

Устремляя  $k$  к бесконечности, можно получить следующее утверждение.

**Следствие 3.1.** *Справедливо представление*

$$W \stackrel{d}{=} \sqrt{2} \cdot Z \prod_{m=1}^{\infty} |X_m|^{1/2^{m-1}},$$

где случайные величины  $X_1, X_2, \dots$  имеют стандартное нормальное распределение, причем все случайные величины в правой части независимы.

Второе утверждение теоремы 3.1 и утверждение следствия 3.1 могут быть записаны в эквивалентной форме: для любого натурального  $k \geq 2$

$$W \stackrel{d}{=} Z \exp \left\{ \frac{(2^{k-1} - 1) \ln 2}{2^k} + \frac{\ln U}{2^k} + \sum_{m=1}^k \frac{\ln |X_m|}{2^{m-1}} \right\}$$

и

$$W \stackrel{d}{=} Z \exp \left\{ \frac{\ln 2}{2} + \sum_{m=1}^{\infty} \frac{\ln |X_m|}{2^{m-1}} \right\}.$$

Наряду с представлением распределения Лапласа в виде масштабной смеси нормальных законов, о котором говорится в лемме 3.1, из представления (6) вытекает еще одна возможность представления распределения Лапласа в виде масштабной

смеси. А именно: при  $x \geq 0$  справедливо очевидное соотношение

$$P(\sqrt{U} < x) = P(U < x^2) = E(x^2) = 1 - e^{-x^2},$$

откуда вытекает, что случайная величина  $\sqrt{U}$  имеет распределение Рэлея–Райса с плотностью

$$r(x) = 2xe^{-x^2}, \quad x \geq 0.$$

Но тогда распределение произведения  $Z\sqrt{U}$ , являющееся рандомизационной симметризацией распределения Рэлея–Райса, имеет плотность

$$p(x) = |x|e^{-x^2}, \quad x \in \mathbb{R}^1.$$

Этой плотности соответствует функция распределения

$$P(x) = \begin{cases} 1 - \frac{1}{2} e^{-x^2}, & \text{если } x > 0, \\ \frac{1}{2} e^{-x^2}, & \text{если } x \leq 0. \end{cases}$$

Далее, очевидно,  $|X| = \sqrt{X^2}$ , причем случайная величина  $X^2$  имеет распределение хи-квадрат с одной степенью свободы, поскольку  $X$  — случайная величина со стандартным нормальным распределением. Как известно, плотность распределения хи-квадрат с одной степенью свободы имеет вид

$$g(x) = \frac{e^{-x/2}}{\sqrt{2\pi x}}, \quad x \geq 0.$$

Таким образом, из представления (6) получаем следующее утверждение.

**Следствие 3.2.** *Распределение Лапласа является масштабной смесью симметризованного распределения Рэлея–Райса, если смешивающее распределение является распределением хи-квадрат с одной степенью свободы:*

$$\Lambda(x) = \int_0^{\infty} P\left(\frac{x}{\sqrt{z}}\right) \frac{e^{-z/2}}{\sqrt{2\pi z}} dz, \quad x \in \mathbb{R}^1.$$

Пусть  $\xi, \xi_1, \xi_2, \dots$  — случайные величины, имеющие распределение хи-квадрат с одной степенью свободы каждая, а  $R$  — случайная величина, имеющая распределение Рэлея–Райса. Тогда следующая факторизация является переформулировкой следствия 3.1:

$$W \stackrel{d}{=} ZR\sqrt{\xi}.$$

Более того, следующее утверждение, обобщающее следствие 3.2, является простой переформулировкой второго утверждения теоремы 3.1.

**Следствие 3.3.** Для любого натурального  $k \geq 2$  справедливо представление

$$W \stackrel{d}{=} 2^{(2^{k-1}-1)/2^k} Z R^{1/2^{k-1}} \prod_{m=1}^k \xi_m^{1/2^m},$$

где все случайные величины в правой части независимы.

Устремляя в следствии 3.3  $k$  к бесконечности, замечаем, что имеет место следующее утверждение (по сути являющееся переформулировкой следствия 3.1).

**Следствие 3.4.** Справедливо представление

$$W \stackrel{d}{=} \sqrt{2} \cdot Z \prod_{m=1}^{\infty} \xi_m^{1/2^m},$$

где все случайные величины в правой части независимы.

Утверждения следствий 3.3 и 3.4 могут быть записаны в эквивалентной форме: для любого натурального  $k \geq 2$

$$W \stackrel{d}{=} Z \exp \left\{ \frac{(2^{k-1}-1) \ln 2}{2^k} + \frac{\ln R}{2^{k-1}} + \sum_{m=1}^k \frac{\ln \xi_m}{2^m} \right\}$$

и

$$W \stackrel{d}{=} Z \exp \left\{ \frac{\ln 2}{2} + \sum_{m=1}^{\infty} \frac{\ln \xi_m}{2^m} \right\}.$$

**Замечание 3.1.** Пусть  $\zeta_\gamma$  — случайная величина, имеющая распределение Вейбулла с параметром  $\gamma$ . Как известно, плотность такого распределения имеет вид

$$h(x) = \gamma x^{\gamma-1} e^{-x^\gamma} \mathbb{I}(x \geq 0).$$

Несложно видеть, что  $Y^{1/2^{k-1}} \stackrel{d}{=} \zeta_{2^{k-1}}$  (при этом распределение Рэля является распределением Вейбулла с  $\gamma = 2$ ). Таким образом, следствие 3 может быть переформулировано в терминах распределения Вейбулла.

Теорема 3.1 указывает алгоритм построения других представлений распределения Лапласа в виде масштабной смеси вероятностных законов.

## 4 Распределение Лапласа как асимптотическая аппроксимация

### 4.1 Распределение Лапласа как асимптотическая аппроксимация в схеме суммирования случайных величин

Пусть  $X_1, X_2, \dots$  — одинаково распределенные случайные величины с конечным математиче-

ским ожиданием  $EX_1 = a$  и конечной дисперсией  $DX_1 = \sigma^2 > 0$ . Для произвольного натурального  $n$  обозначим

$$S_n = X_1 + \dots + X_n.$$

Пусть  $\{N_p\}_{0 < p < 1}$  — семейство натуральнозначных случайных величин, причем при каждом  $p \in (0, 1)$  случайные величины  $N_p, X_1, X_2, \dots$  независимы в совокупности. Предположим, что случайная величина  $N_p$  имеет геометрическое распределение с параметром  $p$ :

$$P(N_p = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots \quad (7)$$

Наша цель — описать асимптотическое поведение случайных величин

$$S_{N_p} = X_1 + \dots + X_{N_p} \quad (p \rightarrow 0).$$

Суммы  $S_{N_p}$ , называемые геометрическими случайными суммами, играют важную роль в теории надежности (см., например, [27]). Здесь будут приведены версии «центральной предельной теоремы» для геометрических случайных сумм.

Хорошо известно, что

$$ES_{N_p} = EN_p EX_1, \quad DS_{N_p} = EN_p DX_1 + DN_p (EX_1)^2.$$

Поэтому с учетом того, что

$$EN_p = \frac{1}{p}, \quad DN_p = \frac{1}{p^2},$$

получаем

$$ES_{N_p} = \frac{a}{p}, \quad DS_{N_p} = \frac{\sigma^2}{p} + \frac{a^2}{p^2}.$$

Центральными предельными теоремами принято называть утверждения, описывающие асимптотическое поведение распределений стандартизованных сумм случайных величин с конечными дисперсиями, т.е. поведение распределений сумм случайных величин, центрированных своими математическими ожиданиями и нормированных своими среднеквадратичными отклонениями.

Оказывается, что асимптотическое поведение распределений стандартизованных геометрических случайных сумм  $\hat{S}_{N_p}$ , во-первых, существенно отличается от асимптотического поведения распределений стандартизованных сумм, описываемого классической предельной теоремой, и, во-вторых, существенно зависит от того, равно нулю или нет математическое ожидание  $a$  отдельного слагаемого.



В рассматриваемом случае

$$\tilde{S}_{N_p} \equiv \frac{S_{N_p} - ES_{N_p}}{\sqrt{DS_{N_p}}} = \frac{S_{N_p} - a/p}{(1/p)\sqrt{p\sigma^2 + a^2}},$$

и, если  $a = 0$ , то

$$\tilde{S}_{N_p} = \frac{\sqrt{p}}{\sigma} S_{N_p},$$

а если  $a \neq 0$ , то

$$\tilde{S}_{N_p} = \frac{a}{|a|} \frac{1}{\sqrt{1 + p\sigma^2/a^2}} \left( \frac{p}{a} S_{N_p} - 1 \right). \quad (8)$$

Символом  $\Lambda(x)$  будет обозначаться функция распределения Лапласа с единичной дисперсией, соответствующая плотности

$$\ell(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, \quad x \in \mathbb{R}^1.$$

Стандартную показательную функцию распределения, как и ранее, будем обозначать  $E(x)$  (см. (1)).

**Теорема 4.1.** (i) Если  $a = 0$ , то

$$\sup_x \left| P(\tilde{S}_{N_p} < x) - \Lambda(x) \right| = o(1) \quad (p \rightarrow 0).$$

(ii) Если  $a \neq 0$ , то

$$\sup_x \left| P(\tilde{S}_{N_p} < x) - E(x \operatorname{sign} a + 1) \right| = o(1) \quad (p \rightarrow 0).$$

**Доказательство.** Пункт (i) представляет собой центральную предельную теорему для геометрических случайных сумм с нулевым средним, доказательство которой можно найти, к примеру, в [28], гл. 6. В силу представления (8) пункт (ii) является следствием из теоремы Реньи — хорошо известной версии закона больших чисел для геометрических случайных сумм, согласно которой

$$\sup_x \left| P\left(\frac{p}{a} S_{N_p} < x\right) - E(x) \right| = o(1)$$

(см., например, [29] или [30]).

Рассмотрим теперь случайную величину

$$R_p = \sum_{j=1}^{N_p^{(1)}} X_j^{(1)} - \sum_{j=1}^{N_p^{(2)}} X_j^{(2)},$$

где  $N_p^{(1)}$  и  $N_p^{(2)}$  — случайные величины с одинаковым геометрическим распределением (7), а случайные величины  $\{X_j^{(i)}\}$ ,  $j \geq 1$ ,  $i = 1, 2$ , имеют

одинаковое распределение с конечной дисперсией. Предположим, что случайные величины  $N_p^{(i)}$ ,  $\{X_j^{(i)}\}$ ,  $j \geq 1$ ,  $i = 1, 2$ , независимы в совокупности.

Величины вида  $R_p$  являются математическими моделями значений процесса риска со случайными премиями (см., например, [30], гл. 10) и играют важную роль при исследовании процессов спекулятивной деятельности [31]. Очевидно, что  $ER_p = 0$ .

Из теоремы 4.1 при этом вытекает следующий результат.

**Следствие 4.1.** Если  $a = EX_j^{(i)} \neq 0$ , то

$$\sup_x \left| P\left(\frac{R_p}{\sqrt{DR_p}} < x\right) - \Lambda(x) \right| = o(1) \quad (p \rightarrow 0).$$

Этот результат является прямым следствием пункта (ii) теоремы 4.1, поскольку распределение Лапласа является сверточной симметризацией показательного распределения (см. параграф 2.1).

## 4.2 Геометрическая устойчивость распределения Лапласа

В предельных теоремах для геометрических случайных сумм распределение Лапласа играет ту же роль, что нормальное распределение. При этом аналогия отнюдь не ограничивается тем свойством, которое описано в предыдущем параграфе и согласно которому распределение Лапласа оказывается предельным для геометрических сумм независимых одинаково распределенных случайных величин с конечными дисперсиями.

В классической теории суммирования независимых случайных величин *строго устойчивыми* называются те распределения, которые обладают свойством сохранения типа при свертках. А именно: пусть  $X, X_1, X_2, \dots$  — случайные величины с одной и той же функцией распределения  $F(x)$ , причем случайные величины  $X_1, X_2, \dots$  независимы. Распределение  $F(x)$  называется *строго устойчивым*, если для любого натурального  $n$  существуют постоянные  $b_n$  такие, что

$$X_1 + \dots + X_n \stackrel{d}{=} b_n X.$$

Хорошо известно, что в этом определении равенство возможно, только если константы  $b_n$  имеют специальный вид, а именно:

$$b_n = n^{1/\alpha},$$

где  $0 < \alpha \leq 2$ . Нормальное распределение является строго устойчивым с  $\alpha = 2$  и единственным строго устойчивым законом с конечной дисперсией (см., например, [26]).

В теории геометрического суммирования существует аналогичное определение. А именно: пусть  $X, X_1, X_2, \dots$  — случайные величины с одной и той же функцией распределения  $F(x)$  и пусть  $\{N_p\}_{0 < p < 1}$  — семейство случайных величин, имеющих геометрическое распределение (7), причем при каждом  $p \in (0, 1)$  случайные величины  $N_p, X_1, X_2, \dots$  независимы. Распределение  $F(x)$  называется *геометрически строго устойчивым*, если существует  $\alpha > 0$  такое, что для любого  $p \in (0, 1)$  выполнено соотношение

$$X_1 + \dots + X_{N_p} \stackrel{d}{=} p^{-1/\alpha} X.$$

Можно показать, что при этом  $\alpha \leq 2$  (см., например, [32] или [27]).

Несложно убедиться, что распределение Лапласа является геометрически строго устойчивым с  $\alpha = 2$ . Действительно, пусть  $F(x)$  — функция распределения Лапласа с дисперсией  $\sigma^2 = 2$ . Ей соответствует характеристическая функция

$$f(s) = \frac{1}{1 + s^2}, \quad s \in \mathbb{R}^1.$$

Тогда характеристическая функция нормированной геометрической случайной суммы  $\sqrt{p} \cdot S_{N_p}$  имеет вид

$$\begin{aligned} E \exp\{is\sqrt{p} \cdot S_{N_p}\} &= \sum_{k=1}^{\infty} p(1-p)^{k-1} f^k(s\sqrt{p}) = \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} \left(\frac{1-p}{1+ps^2}\right)^k = \\ &= \frac{p}{1-p} \frac{1-p}{1+ps^2} \frac{1+ps^2}{1+ps^2-1+p} = \frac{1}{1+s^2} = f(s). \end{aligned}$$

### 4.3 Распределение Лапласа

как асимптотическая аппроксимация для распределений регулярных статистик, построенных по выборкам случайного объема

Рассмотрим традиционную для математической статистики постановку задачи. Пусть случайные величины  $N_1, N_2, \dots, X_1, X_2, \dots$  определены на одном и том же измеримом пространстве  $(\Omega, \mathcal{A})$ . Пусть на  $\mathcal{A}$  задано семейство вероятностных мер  $\{P_\theta, \theta \in \Theta\}$ . Предположим, что при каждом  $n \geq 1$  случайная величина  $N_n$  принимает только натуральные значения и независима от последовательности  $X_1, X_2, \dots$  относительно каждой из семейства мер  $\{P_\theta, \theta \in \Theta\}$ . Пусть  $T_n = T_n(X_1, \dots, X_{N_n})$  —

некоторая статистика, т.е. измеримая функция от случайных величин  $X_1, \dots, X_{N_n}$ . Для каждого  $n \geq 1$  определим случайную величину  $T_{N_n}$ , положив  $T_{N_n}(\omega) = T_{N_n}(\omega)(X_1(\omega), \dots, X_{N_n}(\omega))$  для каждого элементарного исхода  $\omega \in \Omega$ . Будем говорить, что статистика  $T_n$  асимптотически нормальна, если существуют функции  $\delta(\theta)$  и  $t(\theta)$  такие, что при каждом  $\theta \in \Theta$

$$\begin{aligned} P_\theta(\delta(\theta)\sqrt{n}(T_n - t(\theta)) < x) &\implies \\ &\implies \Phi(x) \quad (n \rightarrow \infty). \end{aligned} \quad (9)$$

Примеры асимптотически нормальных статистик хорошо известны. Свойством асимптотической нормальности обладают, например, выборочное среднее (при условии существования дисперсий), центральные порядковые статистики или оценки максимального правдоподобия (при достаточно общих условиях регулярности) и многие другие статистики.

Дальнейшие рассуждения будут основаны на следующей лемме.

**Лемма 4.1.** Пусть  $\{d_n\}_{n \geq 1}$  — некоторая неограниченно возрастающая последовательность положительных чисел. Предположим, что  $N_n \rightarrow \infty$  по вероятности при  $n \rightarrow \infty$  относительно каждой вероятности из семейства  $\{P_\theta, \theta \in \Theta\}$ . Пусть статистика  $T_n$  асимптотически нормальна в смысле (9). Для того чтобы при каждом  $\theta \in \Theta$  существовала такая функция распределения  $F(x, \theta)$ , что

$$\begin{aligned} P_\theta(\delta(\theta)\sqrt{d_n}(T_{N_n} - t(\theta)) < x) &\implies \\ &\implies F(x, \theta) \quad (n \rightarrow \infty), \end{aligned}$$

необходимо и достаточно, чтобы существовало семейство функций распределения  $\mathcal{H} = \{H(x, \theta) : \theta \in \Theta\}$ , удовлетворяющее условиям

$$H(x, \theta) = 0, \quad x < 0, \quad \theta \in \Theta;$$

$$F(x, \theta) = \int_0^\infty \Phi(x\sqrt{y}) d_y H(y, \theta),$$

$$x \in \mathbb{R}^1, \quad \theta \in \Theta; \quad (10)$$

$$P_\theta(N_n < d_n x) \implies H(x, \theta) \quad (n \rightarrow \infty), \quad \theta \in \Theta.$$

При этом если функции распределения случайных величин  $N_n$  не зависят от  $\theta$ , то не зависят от  $\theta$  и функции распределения  $H(x, \theta)$ , т.е. семейство  $\mathcal{H}$  состоит из единственного элемента.

Доказательство. Данная лемма, по сути, лишь переобозначениями отличается от Теоремы 3 из [33], доказательство которой, в свою очередь,

основано на общих теоремах о сходимости суперпозиций независимых случайных последовательностей [34, 35].

Смесь (10) отличается от той, которая фигурирует в лемме 3.1, тем, что смешивающий параметр стоит не в знаменателе аргумента подынтегральной функции распределения, а в числителе. Поэтому из сказанного в разд. 3 вытекает, что если  $H(x, \theta)$  совпадает с функцией обратного показательного распределения  $Q(x)$ , то предельная функция  $F(x, \theta)$  является функцией распределения Лапласа.

Обратное показательное распределение — это распределение случайной величины

$$V = \frac{1}{U},$$

где случайная величина  $U$  имеет стандартное показательное распределение  $E(x)$ . При этом

$$\begin{aligned} Q(x) &= P(V < x) = P\left(\frac{1}{U} < x\right) = \\ &= P\left(U > \frac{1}{x}\right) = e^{-1/x}, \quad x \geq 0. \end{aligned}$$

Обратное показательное распределение  $Q(x)$  является частным случаем распределения Фреше, хорошо известного в асимптотической теории экстремальных порядковых статистик как предельное распределение II типа (см., например, [36]).

Приведем пример ситуации, в которой случайный объем выборки имеет предельное распределение вида  $Q(x)$ . Пусть  $Y_1, Y_2, \dots$  — независимые одинаково распределенные случайные величины с одной и той же непрерывной функцией распределения. Пусть  $m$  — произвольное натуральное число. Обозначим

$$N(m) = \min \left\{ n \geq 1 : \max_{1 \leq j \leq m} Y_j < \max_{m+1 \leq k \leq m+n} Y_k \right\}.$$

Случайная величина  $N(m)$  имеет смысл количества дополнительных наблюдений, которые надо произвести, чтобы текущий (по  $m$  наблюдениям) максимум был перекрыт. Распределение случайной величины  $N(m)$  было найдено С. Уилксом, который в работе [37] показал, что распределение величины  $N(m)$  является дискретным распределением Парето:

$$P(N(m) \geq k) = \frac{m}{m+k}, \quad k \geq 1 \quad (11)$$

(см. также [38], с. 85).

Пусть теперь  $N^{(1)}(m), N^{(2)}(m), \dots$  — независимые случайные величины с одним и тем же распределением (11). Целая часть числа  $a$  будет обозначаться  $[a]$ . Так как при любом фиксированном  $x > 0$

$$\begin{aligned} 1 - \frac{m}{nx(1 + (m-1)/(nx))} &= 1 - \frac{m}{m-1+nx} \leq \\ &\leq 1 - \frac{m}{m+[nx]} \leq 1 - m/(m+nx) = \\ &= 1 - \frac{m}{nx(1+m/(nx))}, \end{aligned}$$

то для любого  $x > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{1}{n} \max_{1 \leq j \leq n} N^{(j)}(m) < x\right) &= \\ &= \lim_{n \rightarrow \infty} P\left(\max_{1 \leq j \leq n} N^{(j)}(m) < nx\right) = \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{m}{m+[nx]}\right)^n = \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{m}{nx}\right)^n = e^{-m/x}. \end{aligned}$$

Поэтому, если положить

$$N_n = \max_{1 \leq j \leq n} N^{(j)}(m)$$

при  $m = 1$ , лемма 4.1 с  $d_n = n$  дает иллюстрацию того, как вместо ожидаемого в соответствии с утверждениями классической асимптотической статистики нормального распределения при замене объема выборки случайной величиной в качестве предельного распределения регулярных статистик возникает распределение Лапласа. При этом изменение значения параметра  $m$  влечет изменение параметра масштаба (дисперсии) итогового распределения Лапласа, точнее, если

$$\lim_{n \rightarrow \infty} P\left(\frac{N_n}{n} < x\right) = e^{-m/x}, \quad (12)$$

то соответствующая плотность распределения Лапласа имеет вид

$$\ell(x) = \sqrt{\frac{m}{2}} e^{-\sqrt{2m}|x|}, \quad \sigma^2 = \frac{1}{m}, \quad x \in \mathbb{R}^1. \quad (13)$$

Примеры прикладных статистических задач, в которых объем выборки существенно случаен, можно найти, например, в книгах [30] и [39].

## 5 Экстремальные энтропийные свойства распределения Лапласа

Хорошо известно, что дифференциальная энтропия является практически оптимальной характеристикой неопределенности (непредсказуемости)

вероятностных распределений. Если  $Y$  — абсолютно непрерывная случайная величина с плотностью  $p(x)$ , то ее дифференциальная энтропия определяется как функционал

$$H(Y) = -E \ln p(Y) = - \int_{-\infty}^{\infty} p(x) \ln p(x) dx.$$

Привлекательность распределения Лапласа в качестве вероятностной модели при решении конкретных прикладных задач во многом обусловливается его экстремальными энтропийными свойствами. Согласно энтропийному (информационному) подходу построения вероятностных математических моделей в условиях неопределенности следует выбирать то модельное распределение, которое обладает максимальной энтропией при заданном условиями задачи комплексе ограничений. Выбор максимально неопределенной модели в определенном смысле соответствует реализации минимаксного подхода.

Хорошо известны следующие экстремальные энтропийные свойства распределения Лапласа.

1. Пусть  $\mathcal{F}_1$  — класс всех абсолютно непрерывных распределений, носителем которых является вся вещественная прямая, с нулевым математическим ожиданием и абсолютным моментом первого порядка, равным  $1/\sqrt{2}$ . Тогда, если, как и ранее,  $W$  — случайная величина с распределением Лапласа  $\Lambda(x)$ ,

$$H(W) = \max \{ H(Y) : \mathcal{L}(Y) \in \mathcal{F}_1 \}$$

(см., например, [40]).

2. В разд. 3 было показано (см. лемму 3.1), что случайная величина  $W$  с распределением Лапласа допускает представление (2) в виде произведения независимых случайных величин  $X$  и  $\sqrt{U}$ , где  $X$  — случайная величина со стандартным нормальным распределением, а  $U$  — случайная величина со стандартным показательным распределением. Хорошо известны экстремальные энтропийные свойства нормального и показательного распределений. Пусть  $\mathcal{F}_2$  — класс всех абсолютно непрерывных распределений, носителем которых является вся вещественная прямая, с нулевым математическим ожиданием и единичной дисперсией;  $\mathcal{F}_3$  — класс всех абсолютно непрерывных распределений, носителем которых является неотрицательная полуось, с единичным математическим ожиданием. Тогда

$$H(W) = \max \left\{ H(Y) : Y \stackrel{d}{=} Y' \sqrt{Y''}; \right. \\ \left. \mathcal{L}(Y') \in \mathcal{F}_2, \mathcal{L}(Y'') \in \mathcal{F}_3 \right\}$$

(см., например, [40]). Этим свойством часто мотивируется выбор распределения Лапласа в качестве распределения погрешностей измерений, в которых точность (параметр масштаба) изменяется от измерения к измерению случайным образом (см., в частности, [5, 41]). Это свойство также позволяет построить методику определения характерных временных масштабов в экспериментах с плазменной турбулентностью [22].

## 6 Задача проверки гипотез: асимптотическая постановка

Рассмотрим задачу проверки простой гипотезы в случае однопараметрического семейства. Пусть имеются независимые наблюдения  $\mathbf{X}_n = (X_1, \dots, X_n)$ , каждое из которых принимает значения в произвольном измеримом пространстве  $(\mathcal{X}, \mathcal{A})$  и имеет неизвестную с точностью до параметра  $\theta$  плотность  $p(x, \theta)$  относительно некоторой  $\sigma$ -конечной меры  $\nu(\cdot)$  на  $\mathcal{A}$ . Предположим, что неизвестный параметр  $\theta$  принадлежит открытому множеству  $\Theta \subset \mathbb{R}^1$ , содержащему ноль. Обозначим через  $P_{n,\theta}, E_{n,\theta}$  соответственно распределение и математическое ожидание  $\mathbf{X}_n$ , а через  $P_\theta, E_\theta$  соответственно распределение и математическое ожидание  $X_1$ .

Пусть мы хотим проверить простую гипотезу

$$H_0 : \theta = 0 \tag{14}$$

против сложной альтернативы  $\theta \neq 0$ . В общем случае наилучшего (равномерно наиболее мощного) критерия не существует, и поэтому рассмотрим асимптотический подход, при котором  $n \rightarrow \infty$ . Рассмотрим сначала простую альтернативу ( $\theta_1$  известно)

$$H_1 : \theta = \theta_1 \neq 0. \tag{15}$$

Заметим, что точка  $\theta = 0$  в гипотезе  $H_0$  может быть заменена на любую фиксированную точку  $\theta_0 \in \Theta$ . Этот случай сводится к предыдущему с помощью рассмотрения семейства плотностей вида  $p(x, \theta_0 + \xi)$ , где  $\xi$  принадлежит некоторой окрестности нуля. Согласно фундаментальной лемме Неймана–Пирсона (см., например, [42], теореме 3.2.1) наилучший (наиболее мощный) критерий основан на логарифме отношения правдоподобия

$$\Lambda_n(\theta) = \sum_{i=1}^n (l(X_i, \theta) - l(X_i, 0)), \tag{16}$$

где  $l(x, \theta) = \ln p(x, \theta)$ , и отвергает гипотезу  $H_0$  в случае, если

$$\Lambda_n(\theta_1) > c_n,$$

причем критическое значение  $c_n$  выбирается из условия

$$P_{n,0}(\Lambda_n(\theta_1) > c_n) = \alpha, \quad (17)$$

где  $\alpha \in (0, 1)$  — фиксированный уровень значимости, и для простоты предполагается непрерывность распределения  $\Lambda_n(\theta_1)$  при гипотезе  $H_0$ , т. е. считается, что

$$P_{n,0}(\Lambda_n(\theta_1) = c_n) = 0.$$

Далее предположим также, что существуют все необходимые моменты случайных величин  $l(X_1, \theta)$  и все необходимые производные по  $\theta$  функций  $l(x, \theta)$ .

Обозначим через

$$\begin{aligned} \mu(\theta) &= E_\theta(l(X_1, \theta) - l(X_1, 0)), \\ \sigma^2(\theta) &= D_\theta(l(X_1, \theta) - l(X_1, 0)) \end{aligned} \quad (18)$$

математическое ожидание и дисперсию случайных величин  $l(X_1, \theta) - l(X_1, 0)$  при распределении  $P_\theta$ .

Поскольку  $\Lambda_n(\theta)$  есть сумма независимых одинаково распределенных случайных величин, то согласно центральной предельной теореме имеем при  $n \rightarrow \infty$

$$\mathcal{L}\left(\frac{\Lambda_n(\theta_1) - n\mu(0)}{\sigma(0)\sqrt{n}} \Big| H_0\right) \rightarrow \mathcal{N}(0, 1), \quad (19)$$

где  $\mathcal{L}(Z|H_i)$  означает распределение  $Z$  при гипотезе  $H_i$ ,  $i = 0, 1$ , а  $\mathcal{N}(\mu, \sigma^2)$  — нормальный закон с параметрами  $\mu$  и  $\sigma^2$ . Из соотношения (17) теперь следует (поскольку сходимость функций распределения к нормальной функции распределения равномерна), что

$$c_n = \sqrt{n} \cdot \sigma(0)u_\alpha + n\mu(0) + o(1), \quad (20)$$

где  $u_\alpha = \Phi^{-1}(1 - \alpha)$  и  $\Phi(x)$  — функция распределения стандартного нормального закона.

Обозначим через  $\beta_n^*(\theta_1)$  мощность наилучшего критерия для проверки гипотезы  $H_0$  против альтернативы  $H_1$  (см. (14), (15)), основанного на  $\Lambda_n(\theta_1)$ , т. е.

$$\beta_n^*(\theta_1) = P_{n,\theta_1}(\Lambda_n(\theta_1) > c_n).$$

Покажем, что этот критерий состоятелен, т. е. справедлив следующий хорошо известный результат (см., например, [43], теорему 3.3.1). Приведем здесь короткое доказательство этого факта, основанное на неравенстве Иенсена.

**Лемма 6.1.** Если  $\sigma^2(0) > 0$ ,  $\sigma^2(\theta_1) > 0$ , то

$$\beta_n^*(\theta_1) \rightarrow 1 \quad (n \rightarrow \infty).$$

**Доказательство.** Используя опять центральную предельную теорему, получаем

$$\mathcal{L}\left(\frac{\Lambda_n(\theta_1) - n\mu(\theta_1)}{\sigma(\theta_1)\sqrt{n}} \Big| H_1\right) \rightarrow \mathcal{N}(0, 1),$$

и поэтому с учетом (20) для мощности  $\beta_n^*(\theta_1)$  имеем представление

$$\begin{aligned} \beta_n^*(\theta_1) &= P_{n,\theta_1}(\Lambda_n(\theta_1) > c_n) = \\ &= 1 - \Phi\left(\frac{c_n - n\mu(\theta_1)}{\sigma(\theta_1)\sqrt{n}}\right) + o(1) = \\ &= \Phi\left(\frac{\sqrt{n}(\mu(\theta_1) - \mu(0)) - u_\alpha\sigma(0)}{\sigma(\theta_1)}\right) + o(1). \end{aligned} \quad (21)$$

Применив теперь неравенство Иенсена (см., например, [44], теорему 4.7.5) к математическим ожиданиям  $\mu(\theta_1)$ ,  $\mu(0)$  (см. (18)), имеем

$$\mu(0) = E_0 \ln \frac{p(X_1, \theta_1)}{p(X_1, 0)} \leq \ln E_0 \frac{p(X_1, \theta_1)}{p(X_1, 0)} = 0,$$

$$\begin{aligned} \mu(\theta_1) &= E_{\theta_1} \ln \frac{p(X_1, \theta_1)}{p(X_1, 0)} = -E_{\theta_1} \ln \frac{p(X_1, 0)}{p(X_1, \theta_1)} \geq \\ &\geq -\ln E_{\theta_1} \frac{p(X_1, 0)}{p(X_1, \theta_1)} = 0. \end{aligned}$$

Из этих неравенств следует, что

$$\sqrt{n}(\mu(\theta_1) - \mu(0)) \rightarrow +\infty \quad (n \rightarrow \infty),$$

и поэтому в силу формулы (21) действительно

$$\beta_n^*(\theta_1) \rightarrow 1 \quad (n \rightarrow \infty).$$

Факт стремления мощности  $\beta_n^*(\theta_1)$  к единице (точнее, скорость сходимости к единице) может быть использован для сравнения различных состоятельных критериев (см., например, [43] и обзор, приведенный там). Однако здесь будет рассмотрен несколько иной, асимптотический подход к сравнению различных критериев. Это так называемый подход Питмэна (см. [45]), согласно которому для получения нетривиального предела мощности  $\beta_n^*(\theta_1)$ , заключенного между  $\alpha$  и 1, рассматривают последовательность альтернатив  $\theta_1 = \theta_n$ , стремящуюся к нулю. Из центральной предельной теоремы для схемы серий (см., например, [44], теорему 8.4.5) следует, что в регулярном случае для выполнения этого должно быть

$$\mu(\theta_n) - \mu(0) = \mathcal{O}\left(n^{-1/2}\right), \quad \theta_n = \mathcal{O}\left(n^{-1/2}\right).$$

Поэтому будем рассматривать задачу проверки простой гипотезы  $H_0$  (см. (14)) против последовательности сложных близких альтернатив вида

$$H_{n,1} : \theta = \frac{t}{\sqrt{n}}, \quad 0 < t \leq C, \quad C > 0,$$

где параметр  $t$  неизвестен. Для любого фиксированного  $t \in (0, C]$  наилучший критерий для проверки гипотезы  $H_0$  против простой альтернативы

$$H_{n,t} : \theta = \frac{t}{\sqrt{n}}$$

основан на логарифме отношения правдоподобия

$$\Lambda_n(t) \equiv \sum_{i=1}^n \left( l(X_i, tn^{-1/2}) - l(X_i, 0) \right). \quad (22)$$

Обозначим через  $\beta_n^*(t)$  мощность такого критерия уровня  $\alpha \in (0, 1)$ . Заметим, что поскольку  $t$  неизвестно, нельзя использовать статистику  $\Lambda_n(t)$  для построения критерия проверки гипотезы  $H_0$  против альтернативы  $H_{n,1}$ . Однако  $\beta_n^*(t)$  — так называемая огибающая функция мощности — дает верхнюю границу для мощности любого критерия при проверке гипотезы  $H_0$  против фиксированной альтернативы  $H_{n,t}$ ,  $t > 0$ , и может служить стандартом при сравнении различных критериев.

Найдем предельное выражение для  $\beta_n^*(t)$ . При естественных условиях регулярности формула Тейлора дает

$$\begin{aligned} l(X_i, tn^{-1/2}) - l(X_i, 0) &= \\ &= \frac{t}{\sqrt{n}} l^{(1)}(X_i) + \frac{1}{2} \frac{t^2}{n} l^{(2)}(X_i) + \dots, \end{aligned} \quad (23)$$

где

$$l^{(j)}(x) = \frac{\partial^j}{\partial \theta^j} l(x, \theta) \Big|_{\theta=0}, \quad j = 1, 2, \dots$$

Поэтому из (22) и (23) получаем стохастическое разложение для  $\Lambda_n(t)$  в виде

$$\Lambda_n(t) = tL_n^{(1)} - \frac{1}{2} t^2 I + \frac{1}{2} \frac{t^2}{\sqrt{n}} L_n^{(2)} + \dots, \quad (24)$$

где

$$L_n^{(j)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (l^{(j)}(X_i) - E_0 l^{(j)}(X_i)), \quad j = 1, 2, \dots,$$

и  $I = E_0(l^{(1)}(X_1))^2$  — фишеровская информация. При этом в выражении (24) опущены неслучайный член  $(1/6)(t^3/\sqrt{n})E_0 l^{(3)}(X_1)$ , члены более высокого порядка малости, чем  $n^{-1/2}$ , и использован хорошо известный факт, состоящий в том, что

$$E_0 l^{(1)}(X_1) = 0, \quad E_0 l^{(2)}(X_1) = -I.$$

Критерий, основанный на статистике  $\Lambda_n(t)$ , отвергает гипотезу  $H_0$  в пользу альтернативы  $H_{n,t}$ , если

$$\Lambda_n(t) > c_{n,t},$$

где критическое значение  $c_{n,t}$  выбирается из условия

$$P_{n,0}(\Lambda_n(t) > c_{n,t}) = \alpha.$$

Аналогично (19) и (20) из (24) следует, что

$$\mathcal{L}(\Lambda_n(t)|H_0) \rightarrow \mathcal{N}\left(-\frac{1}{2}t^2 I, t^2 I\right) \quad (25)$$

и

$$c_{n,t} \rightarrow c_t = t\sqrt{I} \cdot u_\alpha - \frac{1}{2}t^2 I. \quad (26)$$

Найдем теперь предельное распределение  $\Lambda_n(t)$  при альтернативе  $H_{n,t}$ . Имеем (см. (24))

$$\begin{aligned} E_{n,t/\sqrt{n}}\Lambda_n(t) &= t\sqrt{n} \cdot E_{t/\sqrt{n}}l^{(1)}(X_1) - \frac{1}{2}t^2 I + \\ &+ \frac{1}{2}t^2 \left( E_{t/\sqrt{n}}l^{(2)}(X_1) - E_0 l^{(2)}(X_1) \right) + \mathcal{O}(n^{-1/2}) = \\ &= \frac{t^2}{2} I + \mathcal{O}(n^{-1/2}), \end{aligned}$$

$$\begin{aligned} D_{n,t/\sqrt{n}}\Lambda_n(t) &= t^2 D_{n,t/\sqrt{n}}L_n^{(1)} + \mathcal{O}(n^{-1/2}) = \\ &= t^2 I + \mathcal{O}(n^{-1/2}), \end{aligned}$$

поэтому

$$\mathcal{L}(\Lambda_n(t) | H_{n,t}) \rightarrow \mathcal{N}\left(\frac{1}{2}t^2 I, t^2 I\right). \quad (27)$$

Теперь с учетом (26), (27) и равенства

$$\beta_n^*(t) = P_{n,t/\sqrt{n}}(\Lambda_n(t) > c_{n,t})$$

имеем

$$\beta_n^*(t) \rightarrow \beta^*(t) = \Phi\left(t\sqrt{I} - u_\alpha\right). \quad (28)$$

Заметим, что для проверки гипотезы  $H_0$  против альтернативы  $H_{n,1}$  существуют критерии, основанные на статистиках, отличных от  $\Lambda_n(t)$ , и имеющие ту же предельную мощность  $\beta^*(t)$ . Такие критерии называются асимптотически наиболее мощными (АНМ), точнее локально АНМ, поскольку альтернатива  $H_{n,1}$  имеет локальный характер. Таковы, например, критерии, основанные на статистиках  $L_n^{(1)}$  и  $\Lambda_n(t_0)$ , где  $t_0 > 0$  фиксировано, оценках максимального правдоподобия и т. п. Заметим, что все эти статистики не зависят от неизвестного параметра  $t$ , и поэтому могут быть использованы при проверке гипотезы  $H_0$  против альтернативы  $H_{n,1}$ .

Покажем, например, что критерий, основанный на статистике  $L_n^{(1)}$ , является АНМ и имеет предельную мощность  $\beta^*(t)$ . Имеем

$$\begin{aligned} L_n^{(1)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n l^{(1)}(X_i), \\ E_{n,0}L_n^{(1)} &= \sqrt{n} \cdot E_0 l^{(1)}(X_1) = 0, \\ D_{n,0}L_n^{(1)} &= D_0 l^{(1)}(X_1) = I, \end{aligned}$$

поэтому, если

$$P_{n,0}(L_n^{(1)} > c_n^{(1)}) = \alpha \in (0, 1),$$

для критического уровня  $c_n^{(1)}$  (в силу центральной предельной теоремы) имеем

$$c_n^{(1)} = \sqrt{I} \cdot u_\alpha + o(1). \quad (29)$$

Далее

$$\begin{aligned} E_{n,t/\sqrt{n}}L_n^{(1)} &= \sqrt{n} \cdot E_{t/\sqrt{n}}l^{(1)}(X_1) = \\ &= \sqrt{n} \cdot \int l^{(1)}(x)p(x, tn^{-1/2}) d\nu(x) = \\ &= tI + \mathcal{O}(n^{-1/2}), \end{aligned}$$

$$\begin{aligned} D_{n,t/\sqrt{n}}L_n^{(1)} &= D_{t/\sqrt{n}}l^{(1)}(X_1) = \\ &= E_0(l^{(1)}(X_1))^2 + \mathcal{O}(n^{-1/2}) = I + \mathcal{O}(n^{-1/2}), \end{aligned}$$

поэтому для мощности  $\beta_n(t)$  критерия, основанного на статистике  $L_n^{(1)}$ , с учетом (29) имеем

$$\begin{aligned} \beta_n(t) &= P_{n,t/\sqrt{n}}(L_n^{(1)} > c_n^{(1)}) = \\ &= \Phi\left(\frac{tI - c_n^{(1)}}{\sqrt{I}}\right) + o(1) = \\ &= \Phi(t\sqrt{I} - u_\alpha) + o(1) \rightarrow \beta^*(t) = \\ &= \Phi(t\sqrt{I} - u_\alpha). \end{aligned}$$

Соотношение (28) создает естественную основу для асимптотического сравнения различных АНМ критериев, однако для различения критериев такого рода, т. е. удовлетворяющих соотношению

$$\beta_n(t) \rightarrow \beta^*(t),$$

где  $\beta_n(t)$  — мощность конкретного рассматриваемого критерия, нужны следующие члены асимптотического разложения  $\beta_n(t)$ , т. е. представление типа

$$\beta_n(t) = \beta^*(t) + \frac{1}{\sqrt{n}}h_1(t) + \frac{1}{n}h_2(t) + \dots \quad (30)$$

Асимптотическим разложениям в статистике посвящены работы [46] и [47]. При получении формул типа (30) для различных критериев было замечено, что при выполнении естественных условий регулярности для АНМ критериев совпадают и члены  $h_1(t)$ , различия наступают в членах порядка  $n^{-1}$ . Этим вопросам посвящены работы [47–50]. При этом величина

$$r(t) = \lim_{n \rightarrow \infty} n(\beta_n^*(t) - \beta_n(t)) \quad (31)$$

допускает статистическую интерпретацию в терминах необходимого числа наблюдений и позволяет находить асимптотический дефект (см. [49–51] и разд. 8, формулы (40), (41)).

Соотношение (30) может быть понято следующим образом. Предположим, что статистику  $T_n$  АНМ критерия можно монотонным преобразованием (не меняющим мощности критерия) преобразовать в статистику  $S_n(t)$  такую, что величина

$$\Delta_n(t) \equiv S_n(t) - \Lambda_n(t) \rightarrow 0 \quad (n \rightarrow \infty) \quad (32)$$

по вероятности относительно распределений  $P_{n,0}$  и  $P_{n,t/\sqrt{n}}$ . Тогда критерий, основанный на статистике  $S_n(t)$ , имеет те же предельные распределения при гипотезах  $H_0$  и  $H_{n,t}$ , что и критерий, основанный на  $\Lambda_n(t)$ , и, следовательно, ту же предельную мощность  $\beta^*(t)$  (см. (28)). Например, в последнем примере

$$T_n = L_n^{(1)},$$

тогда, полагая

$$S_n(t) = tT_n - \frac{1}{2}t^2I,$$

получим (см. (24))

$$\Delta_n(t) = -\frac{1}{2} \frac{t^2}{\sqrt{n}}L_n^{(2)} + \dots \rightarrow 0. \quad (33)$$

В том типичном случае, когда  $\Delta_n(t)$ , как в (32), имеет порядок  $n^{-1/2}$ , т. е. разность между  $S_n(t)$  и  $\Lambda_n(t)$  имеет тот же порядок, можно ожидать, что мощность  $\beta_n(t)$  критерия, основанного на  $S_n(t)$  (или на  $T_n$ ), отличается от  $\beta_n^*(t)$  на величину порядка  $n^{-1/2}$ . Однако было обнаружено, что для широкого класса АНМ критериев это отличие имеет порядок  $n^{-1}$  (см. [49, 50]). Первоначально выражения для  $r(t)$  (см. (31)) строились с помощью асимптотических разложений для  $\beta_n^*(t)$  и  $\beta_n(t)$  (см. [46, 47]). Этот подход технически очень трудоемкий и громоздкий. Однако в работах [49, 50] была получена общая

формула для величины  $r(t)$  без построения асимптотических разложений. Эта формула имеет наглядный вид в терминах условных дисперсий. Для ее демонстрации обозначим через  $\Lambda(t)$  нормальную случайную величину вида  $\mathcal{N}(-(1/2)t^2I, t^2I)$ . Тогда в силу (25)

$$\mathcal{L}(\Lambda_n(t) | H_0) \rightarrow \mathcal{L}(\Lambda(t)).$$

Предположим, что при гипотезе  $H_0$  случайный вектор  $(\sqrt{n} \cdot \Delta_n(t), \Lambda_n(t))$  имеет предельное распределение (типичным образом двумерное нормальное), совпадающее с распределением вектора  $(\Delta(t), \Lambda(t))$ . Тогда можно показать [49, 50], что

$$r(t) = \frac{1}{2t\sqrt{I}} \varphi(u_\alpha - t\sqrt{I}) D(\Delta(t) | \Lambda(t) = c_t), \quad (34)$$

где

$$c_t = t\sqrt{I} \cdot u_\alpha - \frac{1}{2} t^2 I$$

и

$$\varphi(x) = \Phi'(x).$$

Например, для критерия, основанного на статистике  $T_n = L_n^{(1)}$  в работе [50] (формула 1.4.10) получено выражение

$$r(t) = \frac{t^3}{8\sqrt{I}} \varphi(u_\alpha - t\sqrt{I}) (D_0 l^{(2)}(X_1) - I^{-1} E_0^2 l^{(1)}(X_1) l^{(2)}(X_1)).$$

В работе [50] рассмотрен общий случай в терминах общего статистического эксперимента и приведена общая теорема [50], теорема 3.2.1), дающая достаточные условия для существования предела

$$r = \lim_{n \rightarrow \infty} \tau_n^{-2} (\beta_n^* - \beta_n) = \frac{1}{2} e^b p(b) D(\Delta | \Lambda = b), \quad (35)$$

где  $b = \Phi_1^{-1}(1 - \alpha)$ ,  $\Phi_1(x)$  — функция распределения, предельная для логарифма отношения правдоподобия  $\Lambda_n$  при гипотезе  $H_0$  (выше было  $\Phi_1(x) = \Phi((x + (1/2)t^2I)/(t\sqrt{I}))$ ),  $p(x) = \Phi_1'(x)$  и  $\tau_n \rightarrow 0$  — малый параметр (выше было  $\tau_n = n^{-1/2}$ ),  $(\Delta, \Lambda)$  — случайный вектор, предельный для  $(\tau_n^{-1} \Delta_n, \Lambda_n)$ ,  $\Delta_n = S_n - \Lambda_n$ ,  $S_n$  — монотонное преобразование статистики критерия  $T_n$ .

Далее будет приведён пример, когда в формуле (35)  $\tau_n \neq n^{-1/2}$ . С этой целью будет рассмотрено распределение Лапласа с параметром сдвига и на эвристическом уровне показано, что в этом случае  $\tau_n = n^{-1/4}$ .

## 7 Асимптотически наиболее мощный критерий в случае распределения Лапласа

В разд. 4 было показано, как в математической статистике естественно возникает распределение Лапласа (см. формулы (12) и (13)). В этом разделе рассмотрим приложение распределения Лапласа к асимптотическим задачам проверки гипотез, описанным в предыдущем разделе. Итак, рассмотрим распределение Лапласа с неизвестным параметром сдвига, т.е. рассмотрим распределение с плотностью

$$p(x, \theta) = \frac{1}{2} e^{-|x-\theta|}, \quad x, \theta \in \mathbb{R}^1. \quad (36)$$

Заметим, что это семейство не является регулярным, поскольку у  $p(x, \theta)$  не существует производной по  $\theta$  в точке  $\theta = x$ . Это отсутствие регулярности приводит к нарушению естественного порядка  $n^{-1}$  разности  $\beta_n^*(t) - \beta_n(t)$  и приводит к порядку  $n^{-1/2}$ . Факт нарушения обычных порядков при сравнении оценок в случае распределения Лапласа был отмечен в работе [47] (с. 34). При этом там была ссылка на работу [52], в которой на эвристическом уровне построено асимптотическое разложение для оценок максимального правдоподобия. Строгое доказательство для таких оценок дано в работе [53]. Рассмотрим здесь АНМ критерий, основанный на знаковой статистике, и получим на эвристическом уровне без привлечения асимптотических разложений формулу для  $r(t)$  с помощью общей теоремы 3.2.1 из работы [50]. Формальное доказательство полученной формулы для  $r(t)$  (см. (39)), состоящее в проверке условий этой теоремы, будет опубликовано в одной из наших следующих статей.

В работе [54] была доказана следующая лемма об асимптотическом поведении логарифма отношения правдоподобия  $\Lambda_n(t) \equiv \Lambda_n(tn^{-1/2})$  (см. (16) и (22)) в случае распределения Лапласа.

**Лемма 7.1.** В случае распределения Лапласа (36) справедливы следующие соотношения: фишеровская информация равна 1, т.е.  $I = 1$ ,

$$\begin{aligned} \mathcal{L}(\Lambda_n(t) | H_0) &\rightarrow \mathcal{N}\left(-\frac{t^2}{2}, t^2\right), \\ \mathcal{L}(\Lambda_n(t) | H_{n,t}) &\rightarrow \mathcal{N}\left(\frac{t^2}{2}, t^2\right), \end{aligned}$$

и поэтому

$$\beta_n^*(t) \rightarrow \beta^*(t) = \Phi(t - u_\alpha) \quad (n \rightarrow \infty).$$



Эта лемма показывает, что отсутствие дифференцируемости по  $\theta$  функции  $p(x, \theta)$  (см. (36)) в точке  $\theta = x$  качественно не влияет на порядок альтернатив  $\theta_n$  (равный  $n^{-1/2}$ ) и вид предельной мощности  $\beta^*(t)$ .

## 8 Формула для предельного отклонения мощностей

В этом разделе на эвристическом уровне будет показано, что справедлива формула (см. (34) и (35))

$$r(t) = \lim_{n \rightarrow \infty} \sqrt{n} (\beta_n^*(t) - \beta_n(t)) = \frac{t^2}{3} \varphi(u_\alpha - t), \quad (37)$$

где  $\beta_n(t)$  — функция мощности АНМ критерия, основанного на знаковой статистике

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(X_i). \quad (38)$$

Формула (37) показывает, что отсутствие регулярности у распределения Лапласа приводит к нарушению естественного порядка разности  $\beta_n^*(t) - \beta_n(t)$  (равного  $n^{-1}$  (см. (31))). Из формулы (37) также следует, что этот порядок равен  $n^{-1/2}$ .

Получим сначала стохастическое разложение для  $\Lambda_n(t)$ . Имеем:

$$\begin{aligned} \Lambda_n(t) &= t\sqrt{n} + 2 \sum_{i=1}^n (X_i - tn^{-1/2}) \mathbf{1}_{[0, tn^{-1/2}]}(X_i) - \\ &- \frac{2t}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{(-\infty, 0)}(X_i) = \frac{t}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{[0, \infty)}(X_i) - \\ &- \frac{t}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{(-\infty, 0)}(X_i) + \\ &+ 2 \sum_{i=1}^n (X_i - tn^{-1/2}) \mathbf{1}_{[0, tn^{-1/2}]}(X_i) = \\ &= \frac{t}{\sqrt{n}} \sum_{i=1}^n \text{sign}(X_i) + \frac{t}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{0\}}(X_i) + \\ &+ 2 \sum_{i=1}^n (X_i - tn^{-1/2}) \mathbf{1}_{[0, tn^{-1/2}]}(X_i). \end{aligned}$$

Поскольку распределение  $X_i$  непрерывно, то

$$P_{n,\theta} \left( \sum_{i=1}^n \mathbf{1}_{\{0\}}(X_i) > 0 \right) = 0, \quad \theta > 0,$$

и поэтому почти всюду справедливо представление

$$\begin{aligned} \Lambda_n(t) &= \frac{t}{\sqrt{n}} \sum_{i=1}^n \text{sign}(X_i) + \\ &+ 2 \sum_{i=1}^n (X_i - tn^{-1/2}) \mathbf{1}_{[0, tn^{-1/2}]}(X_i). \end{aligned}$$

Рассмотрим следующее монотонное преобразование ( $t > 0$ ) статистики  $T_n$  (см. (38)):

$$S_n(t) = tT_n - \frac{1}{2}t^2,$$

тогда (см. (32) и (33))

$$\begin{aligned} \Delta_n(t) &= S_n(t) - \Lambda_n(t) = \\ &= -\frac{1}{2}t^2 - 2 \sum_{i=1}^n (X_i - tn^{-1/2}) \mathbf{1}_{[0, tn^{-1/2}]}(X_i). \end{aligned}$$

Справедливо следующее утверждение (доказательство приведено в работе [54]).

**Лемма 8.1.** Для распределения Лапласа (36) справедливы следующие соотношения

$$\mathcal{L}(\sqrt[4]{n} \cdot \Delta_n(t) \mid H_0) \rightarrow \mathcal{N} \left( 0, \frac{2t^3}{3} \right),$$

$$\begin{aligned} \mathcal{L}((\sqrt[4]{n} \cdot \Delta_n(t), \Lambda_n(t)) \mid H_0) &\rightarrow \\ &\rightarrow \mathcal{N}_2 \left( 0, \frac{2t^3}{3}, 0, -\frac{t^2}{2}, t^2 \right), \end{aligned}$$

где  $\mathcal{N}_2$  — двумерный нормальный закон с соответствующими параметрами.

Из этой леммы следует, что случайные величины  $\sqrt[4]{n} \cdot \Delta_n(t)$  и  $\Lambda_n(t)$  асимптотически независимы, и поэтому формула для  $r(t)$  (см. (35)) с  $\tau_n = n^{-1/4}$  приобретает вид (см. (34) и (37))

$$\begin{aligned} r(t) &= \lim_{n \rightarrow \infty} \sqrt{n} (\beta_n^*(t) - \beta_n(t)) = \\ &= \frac{1}{2t} \varphi(u_\alpha - t) D(\Delta(t) \mid \Lambda(t) = c_t) = \\ &= \frac{1}{2t} \varphi(u_\alpha - t) D\Delta(t) = \frac{t^2}{3} \varphi(u_\alpha - t), \quad (39) \end{aligned}$$

где  $\Lambda(t)$  и  $\Delta(t)$  — независимые нормальные случайные величины с параметрами соответственно  $(-t^2/2, t^2)$  и  $(0, 2t^3/3)$ .

Найдем теперь асимптотическое представление для дефекта (см. [50, 51], с. 40) критерия, основанного на статистике  $T_n$  (см. (38)). Напомним, что дефект  $d_n$  определяется как разность  $(k_n - n)$ , где  $k_n$  — число наблюдений, необходимых критерию, основанному на статистике  $T_n$  для достижения той же мощности, что и критерий, основанный на

$\Lambda_n(t)$ , при одинаковых альтернативах  $tn^{-1/2}$ . Предполагая, что  $d_n$  — непрерывная переменная, получаем равенство для ее определения

$$\beta_n^*(t) = \beta_{k_n}(t\sqrt{k_n n^{-1}}). \quad (40)$$

Из лемм 7.1 и 8.1 и формулы (39) следует, что для мощностей  $\beta_n^*(t)$  и  $\beta_n(t)$  справедливы представления

$$\beta_n^*(t) = \Phi(t - u_\alpha) + \frac{1}{\sqrt{n}} \varphi(t - u_\alpha) h^*(t) + o(n^{-1/2}),$$

$$\beta_n(t) = \Phi(t - u_\alpha) + \frac{1}{\sqrt{n}} \varphi(t - u_\alpha) h(t) + o(n^{-1/2}),$$

где  $h^*(t)$  и  $h(t)$  — некоторые полиномы по  $t$  и  $u_\alpha$ . Из этих соотношений и равенства (40) следует, что  $k_n \rightarrow \infty$ ,  $k_n/n \rightarrow 1$  и

$$\begin{aligned} d_n &= \sqrt{n} \frac{2(h^*(t) - h(t))}{t} + o(\sqrt{n}) = \\ &= \sqrt{n} \frac{2r(t)}{t\varphi(u_\alpha - t)} + o(\sqrt{n}) = \\ &= \sqrt{n} \frac{2t}{3} + o(\sqrt{n}). \quad (41) \end{aligned}$$

Таким образом, в отличие от регулярного случая, в котором  $d_n \rightarrow d < \infty$  [49, 50], т. е. существует конечный асимптотический дефект, в случае распределения Лапласа дефект  $d_n$  стремится к бесконечности со скоростью  $\sqrt{n}$ .

В работе [54] доказано более слабое утверждение (по сравнению с формулой (39)), составляющее содержание следующей леммы.

**Лемма 8.2.** В случае распределения Лапласа (36) для любого  $0 \leq \delta < 1/2$  справедливо соотношение

$$n^\delta (\beta_n^*(t) - \beta_n(t)) \rightarrow 0 \quad (n \rightarrow \infty), \quad 0 < t \leq C, \quad C > 0.$$

## Литература

1. Бенинг В. Е., Королёв В. Ю. Об использовании распределения Стьюдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения, 2004. Т. 49. Вып. 3. С. 419–435.
2. Laplace P. S. Mémoire sur la probabilité des causes par les événements // Mémoires de Mathématique et le Physique, 1774. Vol. 6. P. 621–656. (English translation: Memoir on the probability of the causes of events // Statistical Sciences, 1986. Vol. 1. No. 3. P. 364–378.)
3. Andrews D. F., Bickel P. J., Hampel F. R., Huber P. J., Rogers W. H., Tukey J. W. Robust estimation of location. — Princeton, NJ: Princeton University Press, 1972.
4. Understanding robust and exploratory data analysis / Eds. D. C. Hoaglin, F. Mosteller, J. W. Tukey. — N.Y.: Wiley, 1983.
5. Shevlyakov G. L., Vilchevski N. O. Robustness in data analysis: Criteria and methods. — Utrecht: VSP, 2002.
6. Easterling R. J. Exponential responses with double exponential measurement error. A model for steam generator inspection // In: Proceedings of the DOE Statistical Symposium, U.S. Department of Energy, 1978. P. 90–110.
7. Hsu D. A. Long-tailed distributions for position errors in navigation // Applied Statistics, 1979. Vol. 28. P. 62–72.
8. Okubo T., Narita N. On the distribution of extreme winds expected in Japan // In: National Bureau of Standards Special Publication 560-1, 1980. P. 12.
9. Bagchi U., Hayya J. C., Ord J. K. The Hermite distribution as a model of demand during lead time for slow-moving items // Decision Sciences, 1983. Vol. 14. P. 447–466.
10. Dadi M. I., Marks R. J. Detector relative efficiencies in the presence of Laplace noise // IEEE Transactions in Aerospace Electronic Systems, 1987. Vol. 23. P. 568–582.
11. Damsleth E., El-Shaarawi A. H. ARMA models with double-exponentially distributed noise // J. of The Royal Statistical Society, 1989. Vol. B51. No. 1. P. 61–69.
12. Madan D. B., Seneta E. The variance gamma (V.G.) model for share market return // J. of Business, 1990. Vol. 63. P. 511–524.
13. Kozubowski T. J., Podgorski K. Asymmetric Laplace laws and modeling financial data // Math. Comput. Modelling, 2001. Vol. 34. P. 1003–1021.
14. Frechét M. Sur les formules de répartition de revenus // Revue de l'Institut International de Statistique, 1939. Vol. 7. No. 1. P. 32–38.
15. Frechét M. Letter to the editor // Econometrica, 1958. Vol. 26. P. 590–591.
16. Inoue T. On income distribution: The welfare implication of the general equilibrium model and the stochastic processes of income distribution formation. PhD. Thesis. University of Minnesota, 1978.
17. Ord J. K., Patil G. P., Taillie C. The choice of a distribution to describe personal incomes // Statistical distributions in scientific work / Eds. C. Taillie, G. P. Patil, B. Baldessari. — Dordrecht–Boston: Reidel, 1981. P. 193–202.
18. Bagnold R. A. The size-grading of sand by wind // Proc. Royal Soc. London, 1937. Vol. A163. P. 250–264.
19. Bagnold R. A. The physics of blown sand and desert dunes. — London: Methuen, 1954.
20. Fieller N. R. J., Gilbertson D. D., Olbricht W. Skew log Laplace distributions to interpret particle size distribution data. Manchester-Sheffield School of Probability and Statistics Research Report No. 235, 1984.
21. Barndorff-Nielsen O. E. Models for non-Gaussian variation, with applications to turbulence // Proc. Royal Soc. A, 1979. Vol. 353. P. 401–419.
22. Королёв В. Ю. Вероятностно-статистический анализ хаотических процессов с помощью смешанных Гауссовских моделей. Декомпозиция волатильности фи-

- нансовых индексов и турбулентной плазмы. — М.: Изд-во ИПИРАН, 2007. 363 с.
23. *Johnson N. L., Kotz S., Balakrishnan N.* Continuous univariate distributions. Vol. II. 2nd ed. — N.Y.: Wiley, 1995.
  24. *Kotz S., Kozubowski T. J., Podgorski K.* The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering and finance. — Boston: Birkhauser, 2001.
  25. *Градштейн И. С., Рыжик И. М.* Таблицы интегралов, сумм, рядов и произведений. — М.: Наука, 1971. 1108 с.
  26. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 2. — М.: Мир, 1984.
  27. *Gnedenko V. V., Korolev V. Yu.* Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996.
  28. *Круглов В. М., Королёв В. Ю.* Предельные теоремы для случайных сумм. — М.: Изд-во Московского университета, 1990. 269 с.
  29. *Kalashnikov V. V.* Geometric sums: Bounds for rare events with applications. — Dordrecht: Kluwer Academic Publishers, 1997.
  30. *Королёв В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — М.: Физматлит, 2007. 542 с.
  31. *Артюхов С. В., Базюкина О. А., Королёв В. Ю., Кудрявцев А. А., Шевцова И. Г.* Об оптимизации спекулятивной прибыли на примере пункта обмена валют // Актуарий, 2008. № 1(2). С. 50–56.
  32. *Клебанов Л. Б., Мания Г. М., Меламед И. А.* Одна задача В. М. Золотарёва и аналоги безгранично делимых и устойчивых распределений в схеме суммирования случайного числа случайных величин // Теория вероятностей и ее применения, 1984. Т. 29. Вып. 4. С. 791–794.
  33. *Королёв В. Ю.* Сходимость случайных последовательностей с независимыми случайными индексами. II // Теория вероятностей и ее применения, 1995. Т. 40. Вып. 4. С. 907–910.
  34. *Королёв В. Ю.* Сходимость случайных последовательностей с независимыми случайными индексами. I // Теория вероятностей и ее применения, 1994. Т. 39. Вып. 2. С. 313–333.
  35. *Korolev V. Yu.* A general theorem on the limit behavior of superpositions of independent random processes with applications to Cox processes // J. of Mathematical Sciences, 1996. Vol. 81. No. 5. P. 2951–2956.
  36. *Гумбель Э.* Статистика экстремальных значений. — М.: Мир, 1965.
  37. *Wilks S. S.* Recurrence of extreme observations // J. of American Mathematical Society, 1959. Vol. 1. No. 1. P. 106–112.
  38. *Невзоров В. Б.* Рекорды. Математическая теория. М.: Фазис, 2000.
  39. *Бенинг В. Е., Королёв В. Ю., Соколов И. А., Шоргин С. Я.* Рандомизированные модели и методы теории надежности информационных и технических систем. — М.: Торус Пресс, 2007. 248 с.
  40. *Kapur J. N.* Maximum-entropy models in science and engineering. — N.Y.: Wiley, 1989.
  41. *Новицкий П. В., Зограф И. А.* Оценка погрешностей результатов измерений. — Л.: Энергоатомиздат, 1991.
  42. *Lehmann E. L., Romano J. P.* Testing statistical hypotheses. 3rd ed. — Springer, 2005. 784 p.
  43. *Никитин Я. Ю.* Асимптотическая эффективность непараметрических критериев. — М.: Наука, 1995. 250 с.
  44. *Боровков А. А.* Теория вероятностей. — М.: УРСС, 2003. 470 с.
  45. *Pitman E. J. G.* Lecture notes on nonparametric statistical inference. Lectures given for the University of North Carolina? Institute of Statistics, 1948.
  46. *Bickel P. J.* Edgeworth expansions in nonparametric statistics // Ann. of Statist., 1974. Vol. 2. No. 1. P. 1–20.
  47. *Planzagl J.* Asymptotic expansions in parametric statistical theory // Developments in statistics / Ed. P. R. Krishnaiah. — N.Y.—London: Academic Press, 1989. Vol. 3. P. 1–97.
  48. *Bickel P. J., Chibisov D. M., Van Zwet W. R.* On efficiency of first and second order // Intern. Statist. Review, 1981. Vol. 49. P. 169–175.
  49. *Чибисов Д. М.* Вычисление дефекта асимптотически эффективных критериев // Теория вероятностей и ее применения, 1985. Т. 30. Вып. 2. С. 269–288.
  50. *Bening V. E.* Asymptotic theory of testing statistical hypotheses. — Utrecht: VSP, 2000. 277 p.
  51. *Hodges J. L., Lehmann E. L.* Deficiency // Ann. Math. Statist., 1970. Vol. 41. No. 5. P. 783–801.
  52. *Takeuchi K.* Asymptotic theory of statistical estimation. — 1974, Tokyo (in Japanese).
  53. *Бурнашев М. В.* Асимптотические разложения для медианной оценки параметра // Теория вероятностей и ее применения, 1996. Т. 41. Вып. 4. С. 738–753.
  54. *Королёв В. Ю., Тестова А. В., Бенинг В. Е.* О мощности асимптотически оптимального критерия в случае распределения Лапласа // Вестник Тверского Государственного Университета. Сер. прикладная математика, 2008. № 28. Вып. 1. С. 7–27.

# НЕКОТОРЫЕ ОЦЕНКИ ДЛЯ БЛИЗКИХ К ПОГЛОЩАЮЩИМ МАРКОВСКИХ МОДЕЛЕЙ\*

А. И. Зейфман<sup>1</sup>, А. В. Чегодаев<sup>2</sup>, В. С. Шоргин<sup>3</sup>

**Аннотация:** Изучаются, вообще говоря, нестационарные системы, описываемые счетными марковскими цепями с непрерывным временем, нулевое состояние которых является «почти поглощающим». Такие системы возникают при описании некоторых задач теории массового обслуживания. Исследуются предельные характеристики таких моделей. В качестве примеров рассмотрены модели, описываемые простыми нестационарными блужданиями.

**Ключевые слова:** сети массового обслуживания; цепи Маркова с непрерывным временем; эргодичность; процессы рождения и гибели; простое случайное блуждание

## 1 Введение

Марковские модели в задачах, связанных с системами и сетями массового обслуживания, исследуются и применяются очень давно (см., например, [1, 2]). При этом создание новых методов исследования марковских цепей и расширение круга решаемых задач вызвали появление целого ряда новых работ (см., например, [3] и приведенную там библиографию).

Более реалистические модели массового обслуживания, описываемые нестационарными марковскими цепями, активно изучаются уже более тридцати лет, начиная с заметки [4]. Построение предельного режима и нахождение явных формул для вероятностей состояний таких моделей, как правило, невозможно, поэтому, естественно, основной интерес связан с вопросами аппроксимации характеристик таких систем и получением количественных оценок (см., например, [5–8]).

В данной статье-заметке исследуется класс марковских цепей с непрерывным временем, для которых интенсивность выхода из нулевого состояния в определенном смысле мала. Такие цепи возникают при изучении различных классов задач массового обслуживания. Метод исследования, предложенный в [9] (см. также [10]), основан на применении: (а) логарифмической нормы оператора; (б) некоторых специальных перенормировок.

Пусть  $X(t)$ ,  $t \geq 0$ , — нестационарная марковская цепь со счетным пространством состояний  $E = \{0, 1, \dots\}$ . Обозначим через  $p_{ij}(s, t) =$

$= P(X(t) = j | X(s) = i)$ ,  $i, j \in E$ ,  $0 \leq s \leq t$ , вероятность перехода из состояния  $i$  в состояние  $j$ , а  $p_i(t) = P(X(t) = i)$ ,  $i \in E$ ,  $t \geq 0$ , — вероятность нахождения процесса в состоянии  $i$ . Пусть  $\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^T$  — вектор вероятностей состояний,  $Q(t) = (q_{ij}(t))$ ,  $t \geq 0$ , — соответствующая матрица интенсивностей. Положим  $A(t) = (a_{ij}(t)) = Q^T(t) = (q_{ij}(t))^T$ . Для рассматриваемой в настоящей статье ситуации будет предполагаться, что  $|q_{00}(t)| = |a_{00}(t)|$  достаточно мал. Динамика такого процесса при некоторых дополнительных предположениях описывается прямой системой Колмогорова:

$$\frac{d\mathbf{p}}{dt} = A(t)\mathbf{p}, \quad \mathbf{p} = \mathbf{p}(t), \quad t \geq 0. \quad (1)$$

Далее будем предполагать, что  $A(t)$  локально интегрируема на  $[0, \infty)$  и ограничена почти при всех  $t \geq 0$  как операторная функция на  $l_1$ . Тогда систему (1) можно рассматривать как дифференциальное уравнение с ограниченным оператором в пространстве  $l_1$ . При этом оператор Коши  $U(t, s)$ ,  $0 \leq s \leq t$ , уравнения (1) определяется матрицей  $U^T(t, s) = P(s, t) = (p_{ij}(s, t))$ . Далее в тексте, если не указано противное, будет использоваться  $l_1$ -норма для векторов и матриц  $\|\bullet\|$ , а именно  $\|x\| = \sum_i |x_i|$  и  $\|C\| = \sup_j \sum_i |c_{ij}|$ , где  $C = (c_{ij})$ . Обозначим  $\Omega$  множество всех стохастических векторов:  $\Omega = \{\mathbf{x} = (x_0, x_1, \dots)^T : \mathbf{x} \geq 0, \|\mathbf{x}\| = 1\}$ .

Рассмотрим  $\mathbf{p}(t) \in \Omega$ . Исключим из системы уравнение для нулевой координаты, полагая

\* Работа выполнена при финансовой поддержке РФФИ, грант 06-01-00111.

<sup>1</sup> Вологодский государственный педагогический университет; Институт проблем информатики Российской академии наук; ВНКЦ ЦЭМИ РАН, a\_zeifman@mail.ru

<sup>2</sup> Вологодский государственный педагогический университет, cheg\_al@mail.ru

<sup>3</sup> Институт проблем информатики Российской академии наук, vshorgin@ipiran.ru

$$p_0(t) = 1 - \sum_{i \geq 1} p_i(t),$$

тогда из (1) получается система

$$\frac{d\mathbf{z}}{dt} = B(t)\mathbf{z} + \mathbf{f}(t), \quad (2)$$

где

$$B(t) = \begin{pmatrix} a_{11}(t) - a_{10}(t) & a_{12}(t) - a_{10}(t) & \dots \\ a_{21}(t) - a_{20}(t) & a_{22}(t) - a_{20}(t) & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix};$$

$$\mathbf{z} = (p_1, p_2, \dots)^T; \quad \mathbf{f} = (a_{10}, a_{20}, \dots)^T.$$

## 2 Эргодичность, общий случай

Рассмотрим вспомогательную последовательность положительных чисел  $\{d_i\}$  и будем вначале предполагать, что

$$0 < m = \inf_i d_i; \quad M = \sup_{i,j} \frac{d_i}{d_j} < \infty. \quad (3)$$

Положим

$$\alpha_i^*(t) = \sum_{j \geq 1} \frac{d_j}{d_i} a_{ji}(t)$$

и

$$\beta_*(t) = \inf_{i \geq 1} \alpha_i^*(t).$$

Пусть

$$|a_{00}(t)| \leq \varepsilon \beta_*(t), \quad t \geq 0, \quad (4)$$

причем

$$\int_0^\infty \beta_*(t) dt = +\infty. \quad (5)$$

**Теорема 1.** Пусть  $\{d_i\}$  — последовательность положительных чисел такая, что (3)–(5) выполнены. Пусть  $\varepsilon$  достаточно мало. Тогда  $X(t)$  слабо эргодична, причем при всех  $\mathbf{p}(0)$  справедливы следующие оценки:

$$\begin{aligned} \sum_{i \geq 1} d_i |p_i(t) - \pi_i(t)| &\leq \\ &\leq e^{-\int_0^t (1-M\varepsilon)\beta_*(\tau) d\tau} \sum_{i \geq 1} d_i |p_i(0) - \pi_i(0)| \leq \\ &\leq 2Me^{-\int_0^t (1-M\varepsilon)\beta_*(\tau) d\tau} \end{aligned} \quad (6)$$

и

$$\liminf_{t \rightarrow \infty} p_0(t) \geq 1 - \frac{M\varepsilon}{m(1-M\varepsilon)}, \quad (7)$$

где  $\pi(t) = (\pi_0(t), \pi_1(t), \dots)^T$  — предельное (квазистационарное) распределение вероятностей цепи.

**Доказательство.** Рассмотрим вспомогательную диагональную матрицу

$$D = \text{diag}(d_1, d_2, \dots)$$

и соответствующую обратную матрицу

$$D^{-1} = \text{diag}(d_1^{-1}, d_2^{-1}, \dots).$$

Пусть  $l_{1D}$  — пространство последовательностей таких, что  $\|\mathbf{x}\|_{1D} = \sum_{i=1}^\infty d_i |x_i| < \infty$ .

Оценивая логарифмическую норму  $\gamma(B(t))$  в  $l_{1D}$ , получаем

$$\begin{aligned} \gamma(B(t)) &= \\ &= \sup_{i \geq 1} \left( a_{ii}(t) - a_{i0} + \sum_{j \neq i} \frac{d_j}{d_i} |a_{ji}(t) - a_{j0}(t)| \right) \leq \\ &\leq -\beta_*(t) + M\varepsilon \beta_*(t). \end{aligned}$$

Теперь, если  $V(t, s)$  — оператор Коши уравнения (2), получаем

$$\|V(t, s)\| \leq e^{\int_s^t \gamma(B(\tau)) d\tau} \leq e^{-\int_s^t (1-M\varepsilon)\beta_*(\tau) d\tau}.$$

Отсюда вытекает слабая эргодичность  $X(t)$  и неравенство (6).

Далее, при любом начальном условии в норме  $l_{1D}$  получаем

$$\begin{aligned} \|\mathbf{z}(t)\| &= \sum_{i \geq 1} d_i |p_i(t)| \leq \|V(t, 0)\| \|\mathbf{z}(0)\| + \\ &+ \int_0^t \|V(t, \tau)\| \|\mathbf{f}(\tau)\| d\tau \leq \\ &\leq e^{-\int_0^t (1-M\varepsilon)\beta_*(\tau) d\tau} \|\mathbf{z}(0)\| + \\ &+ \int_0^t e^{-\int_\tau^t (1-M\varepsilon)\beta_*(s) ds} M |a_{00}(\tau)| d\tau \leq \\ &\leq e^{-\int_0^t (1-M\varepsilon)\beta_*(\tau) d\tau} \|\mathbf{z}(0)\| + \frac{M\varepsilon}{1-M\varepsilon}, \end{aligned}$$

а значит,

$$\limsup_{t \rightarrow \infty} \|\pi(t)\|_{1D} \leq \frac{M\varepsilon}{1-M\varepsilon},$$

откуда следует (7).

Перейдем к рассмотрению ситуации, при которой вспомогательная последовательность не ограничена сверху. Аналогично предыдущей теореме устанавливается следующее утверждение.

**Теорема 2.** Пусть  $\{d_i\}$  — неубывающая последовательность положительных чисел такая, что выполнены (4), (5), а кроме того, при некотором  $N$  выполняется условие  $q_{0i}(t) = a_{i0}(t) = 0$  при всех  $i > N, t \geq 0$ . Пусть  $\varepsilon$  достаточно мало. Тогда  $X(t)$  слабо эргодична, причем при всех  $\mathbf{p}(0)$  справедливы следующие оценки:

$$\sum_{i \geq 1} d_i |p_i(t) - \pi_i(t)| \leq e^{-\int_0^t (1-d_N \varepsilon) \beta_*(\tau) d\tau} \sum_{i \geq 1} d_i |p_i(0) - \pi_i(0)|,$$

$$\liminf_{t \rightarrow \infty} p_0(t) \geq 1 - \frac{d_N \varepsilon}{1 - d_N \varepsilon},$$

а кроме того,

$$\limsup_{t \rightarrow \infty} E(t, \mathbf{p}) \leq \frac{d_N \omega \varepsilon}{1 - d_N \varepsilon},$$

где  $\omega = \sup_{k \geq 1} k/d_k$ , а  $E(t, \mathbf{p})$  — среднее (математическое ожидание) для  $X(t)$  при начальном распределении вероятностей состояний  $\mathbf{p}(0) = \mathbf{p}$ .

### 3 Эргодичность, процесс рождения и гибели

Пусть теперь  $X(t)$  — «почти поглощающий» процесс рождения и гибели (ПРГ). В этом случае матрица интенсивностей

$$A(t) = \begin{pmatrix} -a_0(t) & b_1(t) & 0 & \dots \\ a_0(t) & -(a_1(t) + b_1(t)) & b_2(t) & \ddots \\ 0 & a_1(t) & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

является якобиевой (тредиагональной).

Положим

$$\alpha_k(t) = a_k(t) + b_{k+1}(t) - d_{k+2} d_{k+1}^{-1} a_{k+1}(t) - d_k d_{k+1}^{-1} b_k(t), \quad k \geq 1,$$

$$\alpha_0(t) = b_1(t) - d_2 d_1^{-1} a_1(t)$$

и

$$\beta(t) = \inf_{i \geq 0} \alpha_i(t).$$

Рассмотрим теперь вместо (4) и (5) условия

$$|a_{00}(t)| = a_0(t) \leq \varepsilon \beta(t), \quad t \geq 0, \quad (8)$$

и

$$\int_0^\infty \beta(t) dt = +\infty \quad (9)$$

соответственно.

**Теорема 3.** Пусть  $\{d_i\}$  — неубывающая последовательность положительных чисел такая, что  $d_1 = 1$ , и выполнены условия (8) и (9). Пусть  $\varepsilon$  достаточно мало. Тогда ПРГ  $X(t)$  слабо эргодичен, причем при всех  $\mathbf{p}(0)$  справедливы следующие оценки:

$$\sum_{i \geq 1} d_i |p_i(t) - \pi_i(t)| \leq 4e^{-\int_0^t \beta(\tau) d\tau} \sum_{i \geq 1} g_i |p_i(0) - \pi_i(0)|, \quad (10)$$

$$\liminf_{t \rightarrow \infty} p_0(t) \geq 1 - 2\varepsilon, \quad (11)$$

а кроме того,

$$\limsup_{t \rightarrow \infty} E(t, \mathbf{p}) \leq 2\omega\varepsilon, \quad (12)$$

где  $g_k = \sum_{i=1}^k d_i$ .

**Доказательство.** Рассмотрим матрицу

$$\mathbf{T} = \begin{pmatrix} d_1 & d_1 & d_1 & \dots \\ 0 & d_2 & d_2 & \dots \\ 0 & 0 & d_3 & \dots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix},$$

введем  $1T$ -норму как  $\|\mathbf{x}\|_{1T} = \|T\mathbf{x}\|$ , соответствующее пространство последовательностей  $l_{1T}$  и вычислим логарифмическую норму  $\gamma(B(t))_{1T}$ . Тогда получаем

$$\gamma(B(t))_{1T} = \gamma(TB(t)T^{-1})_1.$$

Имеем при любом  $a_0(t)$ :

$$\gamma(B(t))_{1T} = \sup_{k \geq 0} (-a_k(t) - b_{k+1}(t) + d_{k+2} d_{k+1}^{-1} a_{k+1}(t) + d_k d_{k+1}^{-1} b_k(t)) \leq -\beta(t);$$

$$\|V(t, s)\|_{1T} \leq e^{-\int_s^t \beta(\tau) d\tau},$$

откуда с учетом известной оценки [6]

$$\|\mathbf{z}(t)\|_{1T} \geq \frac{1}{2} \sum_{i \geq 1} d_i |p_i(t)| \quad (13)$$

получаем слабую эргодичность и неравенство (10).

Далее, при любом начальном условии в норме  $l_{1T}$  получаем

$$\begin{aligned} \|\mathbf{z}(t)\| &\leq \|V(t, 0)\| \|\mathbf{z}(0)\| + \int_0^t \|V(t, \tau)\| \|\mathbf{f}(\tau)\| d\tau \leq \\ &\leq e^{-\int_0^t \beta(\tau) d\tau} \|\mathbf{z}(0)\| + \int_0^t e^{-\int_\tau^t \beta(s) ds} |a_{00}(\tau)| d\tau \leq \\ &\leq e^{-\int_0^t \beta(\tau) d\tau} \|\mathbf{z}(0)\| + \varepsilon, \end{aligned}$$

а значит,

$$\limsup_{t \rightarrow \infty} \|\pi(t)\|_{1T} \leq \varepsilon,$$

откуда следует (11).

Для доказательства (12) достаточно воспользоваться еще раз оценкой (13).

**Следствие.** При любом  $k \geq 1$

$$\sum_{i \geq 1} d_i |p_{0i}(t) - p_{ki}(t)| \leq 2e^{-\int_0^t \beta(\tau) d\tau} \sum_{i=1}^k d_i.$$

**Замечание.** Тот же подход позволяет получать и нижние оценки скорости сходимости (см. [7, 9, 10]), а в случае периодических интенсивностей строить очень важную характеристику процесса — предельное среднее [6, 10].

## 4 Нуль-эргодичность

В двух предыдущих разделах интенсивность выхода из нулевого состояния могла быть и нулевой. Здесь же рассмотрим другую возможную ситуацию.

Положим

$$\alpha_i^*(t) = \sum_{j \geq 0} \frac{d_j}{d_i} a_{ji}(t)$$

и

$$\beta^*(t) = \inf_{i \geq 0} -\alpha_i^*(t).$$

Пусть

$$\int_0^\infty \beta^*(t) dt = +\infty. \quad (14)$$

**Теорема 4.** Пусть существует последовательность  $\{d_i\}$  положительных чисел такая, что  $d_0 = 1, \sup d_i = d < \infty$  и (14) выполнено. Тогда цепь  $X(t)$  нуль-эргодична.

Более того, справедливы следующие оценки:

при всех  $i, \mathbf{p}(0)$

$$p_i(t) \leq \frac{d}{\partial_i} e^{-\int_0^t \beta^*(\tau) d\tau}; \quad (15)$$

при всех  $n, \mathbf{p}(0)$

$$\Pr(X(t) \leq n) \leq \frac{d}{\partial_n} e^{-\int_0^t \beta^*(\tau) d\tau}; \quad (16)$$

при всех  $n, k$

$$\Pr(X(t) \leq n | X(0) = k) \leq \frac{d_k}{\partial_n} e^{-\int_0^t \beta^*(\tau) d\tau}, \quad (17)$$

где  $\partial_n = \min_{0 \leq k \leq n} d_k$ .

**Доказательство.** Доказательство проводится почти так же, как в теореме 1, только в качестве вспомогательного вводится пространство  $l_{1D^*}$  последовательностей таких, что  $\|\mathbf{x}\|_{1D^*} = \sum_{i=0}^\infty d_i |x_i| < \infty$ , а затем оценивается логарифмическая норма  $\gamma(A(t))$  в  $l_{1D^*}$ . Из получаемого в итоге неравенства

$$\|\mathbf{p}(t)\|_{1D^*} \leq e^{-\int_0^t \beta^*(\tau) d\tau} \|\mathbf{p}(0)\|_{1D^*}$$

и сравнения норм и вытекают требуемые оценки.

**Следствие.** При выполнении условий теоремы 4 для любых  $\mathbf{p} = \mathbf{p}(0), t \geq 0$  и любого  $n$  справедлива следующая оценка среднего:

$$E(t, \mathbf{p}) \geq (n+1) \left( 1 - \frac{d}{\partial_n} e^{-\int_0^t \beta^*(\tau) d\tau} \right).$$

Более точная оценка очень просто получается в случае процесса рождения и гибели.

Положим

$$\phi(t) = \inf_{i \geq 1} (a_i(t) - b_i(t)).$$

Пусть

$$\int_0^\infty \phi(t) dt = +\infty. \quad (18)$$

**Теорема 5.** Пусть интенсивности ПРГ  $X(t)$  таковы, что (18) выполнено и вдобавок  $a_0(t) \geq \varepsilon\phi(t)$  при некотором  $\varepsilon \in (0, 1)$ . Тогда для любых  $\mathbf{p} = \mathbf{p}(0)$ ,  $t \geq 0$  справедлива следующая оценка среднего:

$$E(t, \mathbf{p}) \geq \varepsilon \int_0^t \phi(\tau) d\tau + E(0, \mathbf{p}). \quad (19)$$

Доказательство получается непосредственно из оценки

$$\frac{dE}{dt} \geq a_0(t)p_0 + \sum_{i \geq 1} (a_i(t) - b_i(t)) p_i \geq \varepsilon\phi(t).$$

## 5 Примеры

**Пример 1. Простое случайное блуждание.** Пусть  $X(t)$  — процесс рождения и гибели с пространством состояний  $E$  и интенсивностями  $a_0(t) = l(t)$  и  $a_i(t) = a(t)$ ,  $b_i(t) = b(t)$ ,  $i \geq 1$ . Такой процесс служит для описания простых моделей теории массового обслуживания (см. например, [11]). В этом случае возможно применение всех описанных подходов.

**Эргодический случай.** Пусть существует  $s > 1$  такое, что

$$\int_0^\infty (b(t) - sa(t)) dt = +\infty. \quad (20)$$

Положим  $d_k = s^k$  для  $k \geq 1$ . Тогда

$$\alpha_k(t) = (1 - s^{-1})(b(t) - sa(t)), \quad k \geq 1, \\ \alpha_0(t) = b(t) - sa(t)$$

и

$$\beta(t) = (1 - s^{-1})(b(t) - sa(t)),$$

при этом условие (9) выполнено.

Пусть при достаточно малом  $\varepsilon$

$$l(t) \leq \varepsilon\beta(t), \quad t \geq 0. \quad (21)$$

Тогда ПРГ  $X(t)$  слабо эргодичен, и при всех  $\mathbf{p}(0)$  справедливы следующие оценки:

$$\sum_{k \geq 1} s^k |p_i(t) - \pi_i(t)| \leq \\ \leq 4e^{-\int_0^t \beta(\tau) d\tau} \sum_{i \geq 1} g_i |p_i(0) - \pi_i(0)|, \quad (22)$$

$$\liminf_{t \rightarrow \infty} p_0(t) \geq 1 - 2\varepsilon, \quad (23)$$

а кроме того,

$$\limsup_{t \rightarrow \infty} E(t, \mathbf{p}) \leq 2\omega\varepsilon, \quad (24)$$

где  $g_i = \sum_{k=1}^i s^k$ ,  $\omega = \sup_{k \geq 1} \frac{k}{s^k}$ .

Отметим, что в случае стационарного блуждания условие эргодичности (20) означает, что  $a < b$ . Пусть при этом вместо (21) выполнено более простое неравенство  $l \leq \varepsilon$ . Тогда вместо (22)–(24) получаем при всех  $\mathbf{p}(0)$  следующие оценки:

$$\sum_{k \geq 1} s^k |p_i(t) - \pi_i| \leq 4e^{-\beta t} \sum_{i \geq 1} g_i |p_i(0) - \pi_i|;$$

$$\liminf_{t \rightarrow \infty} p_0(t) \geq 1 - \frac{2\varepsilon}{\beta},$$

а кроме того,

$$\limsup_{t \rightarrow \infty} E(t, \mathbf{p}) \leq \frac{2\omega\varepsilon}{\beta},$$

где  $s = \sqrt{b/a}$ ,  $\beta = (\sqrt{b} - \sqrt{a})^2$ .

**Нуль-эргодичность.** Пусть существует  $s < 1$  такое, что

$$\int_0^\infty (sa(t) - b(t)) dt = +\infty, \quad (25)$$

и, кроме того, при некотором достаточно малом  $\varepsilon > 0$

$$l(t) \geq \varepsilon(sa(t) - b(t)). \quad (26)$$

Тогда ПРГ нуль-эргодичен и при  $d = 1$ ,  $d_k = s^k$ ,  $\beta_n = s^n$  и  $\beta^*(t) = \varepsilon(1 - s)(sa(t) - b(t))$  выполняются оценки (15)–(17).

Кроме того, справедлива и оценка типа (19), а именно

$$E(t, \mathbf{p}) \geq \varepsilon \int_0^t (sa(\tau) - b(\tau)) d\tau + E(0, \mathbf{p}).$$

В случае стационарного ПРГ условие (25) эквивалентно тому, что  $a > b$ , а условие (26) — тому, что  $l > 0$ . Выписав для удобства это условие в том же виде (26), получаем при  $s = \sqrt{b/a}$ ,  $\beta^* = \varepsilon(1 - s)(sa - b)$  соответственно оценки:

$$p_n(t) \leq s^{-n} e^{-\beta^* t},$$

$$\Pr(X(t) \leq n) \leq s^{-n} e^{-\beta^* t},$$

$$\Pr(X(t) \leq n | X(0) = k) \leq s^{k-n} e^{-\beta^* t},$$

$$E(t, \mathbf{p}) \geq \varepsilon(sa - b)t + E(0, \mathbf{p}).$$



**Пример 2. Простое случайное блуждание с разными скачками.** Рассмотрим систему массового обслуживания, в которой требования в случае непустой очереди поступают парами (с интенсивностью  $a(t)$ ), а обслуживаются по одному (с интенсивностью  $b(t)$ ). При этом в случае отсутствия требований в системе возможна «иммиграция», т. е. поступление пары требований с интенсивностью  $l(t)$ . В этом случае  $X(t)$  уже не является процессом рождения и гибели.

**Эргодичность.** Пусть существует  $s > 1$  такое, что

$$\int_0^{\infty} (b(t) - s(s+1)a(t)) dt = +\infty. \quad (27)$$

Положим  $d_k = s^k$  для  $k \geq 1$ . Тогда, как легко проверить,

$$\beta_*(t) = (1 - s^{-1})(b(t) - s(s+1)a(t)),$$

далее,  $N = 2$  и при  $l(t) \leq \varepsilon\beta_*(t)$  и достаточно малом  $\varepsilon$  выполнены условия теоремы 2. Следовательно, при всех  $\mathbf{p}(0)$  справедливы следующие оценки:

$$\begin{aligned} \sum_{i \geq 1} s^i |p_i(t) - \pi_i(t)| &\leq \\ &\leq e^{-\int_0^t (1-s^2\varepsilon)\beta_*(\tau) d\tau} \sum_{i \geq 1} s^i |p_i(0) - \pi_i(0)|, \end{aligned}$$

$$\liminf_{t \rightarrow \infty} p_0(t) \geq 1 - \frac{s^2\varepsilon}{1 - s^2\varepsilon},$$

а кроме того,

$$\limsup_{t \rightarrow \infty} E(t, \mathbf{p}) \leq \frac{s^2\omega\varepsilon}{1 - s^2\varepsilon},$$

где  $\omega = \sup_{k \geq 1} k/s^k$ .

Отметим, что в случае стационарной ситуации (интенсивности не зависят от времени) условие (27) эквивалентно тому, что  $2a < b$ . Несложно получить оценки для этого случая, выбрав  $s = \sqrt[3]{b/(2a)} > 1$ .

**Нуль-эргодичность.** Пусть существует  $s < 1$  такое, что

$$\int_0^{\infty} (s(s+1)a(t) - b(t)) dt = +\infty$$

и, кроме того, при некотором достаточно малом  $\varepsilon > 0$

$$l(t) \geq \varepsilon(s(s+1)a(t) - b(t)).$$

Тогда блуждание нуль-эргодично, и при  $d = 1$ ,  $d_k = s^k$ ,  $\partial_n = s^n$  и  $\beta^*(t) = \varepsilon(1 - s^2)(s(s+1)a(t) - b(t))$  выполняются оценки (15)–(17) и (19).

Отметим, что в случае стационарной ситуации блуждание будет нуль-эргодичным при  $2a > b$  и  $l > 0$ .

## Литература

1. Баруча-Рид А. Т. Элементы теории марковских процессов и их приложения. — М.: Наука, 1969. 511 с.
2. Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S. Queueing theory. — Utrecht: VSP, 2004. 446 p.
3. Ching Wai-Ki, Ng M. K. Markov chains: Models, algorithms and applications. — New York: Springer, 2006. 205 p.
4. Гнеденко Б. В., Макаров И. П. Свойства решений задачи с потерями в случае периодических интенсивностей // Дифф. уравнения, 1971. Т. 7. С. 1696–1698.
5. Granovsky B. L., Zeifman A. I. Nonstationary queues: Estimation of the rate of convergence // Queueing Systems, 2004. Vol. 46. P. 363–388.
6. Zeifman A., Leorato S., Orsingher E., Satin Ya., Shilova G. Some universal limits for nonhomogeneous birth and death processes // Queueing Systems, 2006. Vol. 52. P. 139–151.
7. Зейфман А. И., Сатин Я. А. Средние характеристики марковских систем обслуживания // Автоматика и телемеханика, 2007. № 9. С. 122–133.
8. Margolius B. Transient and periodic solution to the time-inhomogeneous quasi-birth death process // Queueing Systems, 2007. Vol. 56. P. 183–194.
9. Zeifman A. I. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes // Stoch. Proc. Appl., 1995. Vol. 59. P. 157–173.
10. Зейфман А. И., Бенинг В. Е., Соколов И. А. Марковские цепи и модели с непрерывным временем. — М.: ЭЛЕКС-КМ, 2008. 168 с.
11. Toyozumi H., Kobayashi Y., Kaiwa K., Shitozawa J. Stochastic features of computer viruses: Towards theoretical analysis and simulation // The 5th St. Petersburg Workshop on Simulation. — StPB.: SPBU, 2005. P. 695–702.

# ОПТИМИЗАЦИЯ РАСПОЛОЖЕНИЯ СТАНЦИЙ ОБСЛУЖИВАНИЯ В ПРОСТРАНСТВЕ

Т. В. Захарова<sup>1</sup>

**Аннотация:** В пространстве  $\mathbb{R}^N$  эффективным образом располагаются станции обслуживания по критерию стационарного среднего времени ожидания начала обслуживания.

**Ключевые слова:** размещение станций обслуживания; критерий оптимальности; время ожидания начала обслуживания

## 1 Введение

В статье рассматривается неклассическая задача теории массового обслуживания. На обслуживание поступают вызовы, распределенные в пространстве  $\mathbb{R}^N$ . Каким образом разместить в этом пространстве станции обслуживания?

Чаще всего рассматривается поток однородных требований, различающихся лишь моментами поступления в систему. Спецификой изучаемого класса систем является довольно сложное фазовое пространство системы, требующее учета информации о положении и числе обслуживающих приборов, положении поступающих вызовов, их вероятностных характеристиках.

Такие модели систем массового обслуживания применяются для изучения реальных систем, где обслуживание производится территориально расположенными объектами. В связи с этим возникают определенные задачи оптимизации.

В данной статье решается задача оптимизации расположения станций обслуживания по критерию стационарного среднего времени ожидания начала обслуживания для систем с дисциплиной обслуживания FIFO (First In, First Out). Станции функционируют как независимые системы массового обслуживания типа  $M|G|1$ . Описываются свойства оптимальных размещений и приводятся алгоритмы построения асимптотически оптимальных размещений.

## 2 Постановка задачи

В пространстве  $\mathbb{R}^N$  возникают требования в случайных точках  $\xi_1, \xi_2, \dots$ , независимых и одинаково распределенных с плотностью распределения  $p$ .

Для обслуживания этих требований имеется  $n$  станций. Моменты поступления требований образуют пуассоновский поток с параметром  $\lambda$ . Интенсивность входящего потока  $\lambda$  изменяется с ростом числа станций  $n$ . В случае, когда нужно подчеркнуть эту зависимость, параметр входящего потока будем обозначать  $\lambda(n)$ .

**Определение 1.** Размещением  $n$  станций обслуживания в пространстве  $\mathbb{R}^N$  назовем множество точек пространства  $\{x_1, \dots, x_n\}$ , в которых они расположены.

Обозначать размещение станций будем символом  $x$ , т. е.  $x = \{x_1, \dots, x_n\}$ . Станцию обслуживания и точку пространства, где она расположена, будем обозначать одним и тем же символом.

**Определение 2.** Зоной влияния станции  $x_i$  назовем множество  $C_i$  тех точек пространства, для которых эта станция является ближайшей, т. е.

$$C_i = \{v \in \mathbb{R}^N : |v - x_i| \leq |v - x_j|, \quad j = 1, 2, \dots, n\}.$$

Расстояние  $|u - v|$  между точками  $u$  и  $v$  пространства  $\mathbb{R}^N$  задается чебышевской нормой:

$$|u - v| = \max_{1 \leq i \leq N} |u_i - v_i|,$$

где  $u = (u_1, \dots, u_N)$ ;  $v = (v_1, \dots, v_N)$ .

Станции обслуживают заявки только из своих зон влияния. Обслуживание осуществляется прибором,двигающимся только по прямой и с постоянной скоростью. При поступлении заявки прибор со станции перемещается в точку вызова, заявка обслуживается прибором некоторое случайное время  $\eta$ , затем прибор возвращается обратно на станцию. Дисциплина обслуживания вызовов следующая: если в момент поступления вызова прибор занят, то поступающий вызов ставится в очередь.

<sup>1</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, lsa@cs.msu.su

После освобождения прибора на обслуживание поступает первая заявка из очереди.

В системе  $M|G|1$  с дисциплиной обслуживания FIFO при загрузке системы меньше единицы стационарное среднее время ожидания начала обслуживания определяется по формуле

$$W = \frac{\lambda\beta_2}{2(1 - \lambda\beta_1)},$$

где  $\lambda$  — интенсивность входящего потока;  $\beta_1, \beta_2$  — первый и второй момент времени обслуживания [1]. Этим определяется целесообразность введения критерия оптимальности

$$W(x) = \frac{1}{2n} \sum_{i=1}^n \frac{\lambda_i\beta_{i2}}{1 - \lambda_i\beta_{i1}}$$

при условии, что  $\max_{1 \leq i \leq n} \lambda_i\beta_{i1} < 1$ , где  $\lambda_i$  — интенсивность потока вызовов, поступающих на станцию  $x_i$ ;  $\beta_{i1}, \beta_{i2}$  — соответственно первый и второй момент времени обслуживания на  $x_i$ .

В статьях [2, 3] рассматривались аналогичные постановки задач, но с вызовами, распределенными на плоскости, и евклидовой нормой для определения расстояния.

**Определение 3.** Размещение  $x^*$  назовем оптимальным, если  $W(x^*) \leq W(x)$  для любого размещения  $x$  такого, что  $|x| = |x^*|$ . Через  $|x|$  здесь обозначено число элементов размещения  $x$ .

Введем еще ряд обозначений:

$$|p|_m = \left( \int p^m(u) du \right)^{1/m};$$

$$E\eta = \beta_1;$$

$$E\eta^2 = \beta_2;$$

$\{x\}$  — последовательность размещений.

### 3 Основные результаты

**Теорема 1.** Пусть плотность  $p$  ограничена,  $p^{N/(N+1)}$  интегрируема по Лебегу,  $E|\xi|^2 < \infty$  и  $\lambda(n) = o(n^{1/2})$ . Тогда для всякой последовательности оптимальных размещений  $\{x^*\}$ :

$$(1) \lim_{n \rightarrow \infty} n^{1/N} \left( \frac{n}{\lambda(n)} W(x^*) - 0,5\beta_2 \right) = \frac{N}{N+1} \beta_1 |p|_{N/(N+1)};$$

$$(2) \lim_{n \rightarrow \infty} \frac{|x_A^*|}{n} = |p|_{N/(N+1)}^{-N/(N+1)} \int_A p^{N/(N+1)}(u) du,$$

где  $A$  — произвольное измеримое по Лебегу множество,  $x_A^* = x^* \cap A$ .

Иные результаты в описанной модели получаются, если  $E\eta = 0$ , т. е. обслуживание заключается лишь в перемещении прибора до вызова и обратно.

**Теорема 2.** Если плотность  $p$  ограничена,  $p^{1/N}$  интегрируема по Лебегу,  $E|\xi|^2 < \infty$  и  $\lambda(n) = o(n^{1,5})$ , тогда для любой последовательности оптимальных размещений  $\{x^*\}$ :

$$(1) \lim_{n \rightarrow \infty} \frac{n^{(N+2)/N}}{\lambda(n)} W(x^*) = \frac{1}{2} \frac{N}{(N+2)} |p|_{1/N};$$

$$(2) \lim_{n \rightarrow \infty} \frac{|x_A^*|}{n} = |p|_{1/N}^{-1/N} \int_A p^{1/N}(u) du$$

для любого измеримого по Лебегу множества  $A$ .

### 4 Вспомогательные утверждения

Все множества, рассматриваемые ниже, являются подмножествами  $N$ -мерного пространства  $\mathbb{R}^N$ , под интегралом понимается  $N$ -кратный интеграл Лебега.

**Лемма 1.** Если  $Q$  — измеримое по Лебегу подмножество метрического пространства,  $S$  — шар с центром в точке  $v$  из того же пространства и меры Лебега множеств  $Q$  и  $S$  равны, тогда

$$\int_Q a(|u-v|) du \geq \int_S a(|u-v|) du$$

для любой неубывающей на  $[0, \infty)$  действительной функции  $a(u)$ .

**Доказательство.** Пусть  $R$  — радиус шара  $S$ . Под знаком интеграла будем подразумевать  $a(|u-v|) du$ . Очевидны следующие соотношения

$$Q = (Q \cap S) \cup (Q \setminus S);$$

$$S = (Q \cap S) \cup (S \setminus Q) \implies \mu(Q \setminus S) = \mu(S \setminus Q);$$

$$\int_{Q \setminus S} \geq a(R) \mu(Q \setminus S) = a(R) \mu(S \setminus Q) \geq \int_{S \setminus Q} \implies$$

$$\implies \int_Q = \int_{Q \cap S} + \int_{Q \setminus S} \geq \int_{Q \cap S} + \int_{S \setminus Q} = \int_S.$$

Следующая лемма является аналогом результатов Л. Ф. Тота [4].

**Лемма 2.** Пусть для  $i = 1, 2, \dots, n$   $S_i$  обозначает шар с центром в нуле,  $\sigma_n$  — шар с центром в нуле и мерой  $\sigma_n = (1/n) \sum_{i=1}^n S_i$ . Справедливо следующее неравенство:

$$\sum_{i=1}^n \int_{S_i} a(|u|) du \geq n \int_{\sigma_n} a(|u|) du \quad (1)$$

для любой неубывающей на  $[0, \infty)$  действительной функции  $a(u)$ .

Доказательство проведем методом математической индукции. Для  $n = 1$  утверждение леммы очевидно, так как неравенство обращается в равенство. Предположим, что (1) верно для  $n = m$ . Докажем его справедливость для  $n = m + 1$ . Под знаком интеграла будем подразумевать  $a(|u|) du$ , тогда

$$\sum_{i=1}^{m+1} \int_{S_i} = \sum_{i=1}^m \int_{S_i} + \int_{S_{m+1}} \geq m \int_{\sigma_m} + \int_{S_{m+1}}. \quad (2)$$

Пусть для определенности  $\sigma_m \subseteq \sigma_{m+1}$ . В случае  $\sigma_{m+1} \subset \sigma_m$  доказательство аналогичное.

$$\begin{aligned} \sigma_{m+1} &= \frac{1}{m+1} \left( \sum_{i=1}^m S_i + S_{m+1} \right) = \\ &= \left( \frac{1}{m} - \frac{1}{m(m+1)} \right) \left( \sum_{i=1}^m S_i + S_{m+1} \right) = \\ &= \sigma_m + \frac{1}{m} S_{m+1} - \frac{1}{m} \sigma_{m+1} = \\ &= \sigma_m + \frac{1}{m} (S_{m+1} - \sigma_{m+1}) \implies S_{m+1} \geq \sigma_{m+1}, \end{aligned}$$

так как по предположению  $\sigma_{m+1} \geq \sigma_m$ .

Поэтому

$$\begin{aligned} \int_{\sigma_m} &= \int_{\sigma_{m+1}} - \int_{\sigma_{m+1} \setminus \sigma_m}; \\ \int_{S_{m+1}} &= \int_{\sigma_{m+1}} + \int_{S_{m+1} \setminus \sigma_{m+1}}. \end{aligned}$$

Подставляя эти выражения в (2), получим

$$\begin{aligned} \sum_{i=1}^{m+1} \int_{S_i} &\geq \\ &\geq m \int_{\sigma_{m+1}} - m \int_{\sigma_{m+1} \setminus \sigma_m} + \int_{\sigma_{m+1}} + \int_{S_{m+1} \setminus \sigma_{m+1}} \geq \\ &\geq (m+1) \int_{\sigma_{m+1}} + a(R_{m+1}) (S_{m+1} - \sigma_{m+1}) - \\ &- ma(R_{m+1}) (\sigma_{m+1} - \sigma_m) = (m+1) \int_{\sigma_{m+1}}. \end{aligned}$$

Предположим сначала, что  $p$  — простая, ограниченная функция и  $p^{N/(N+1)}$  интегрируема по Лебегу.

**Лемма 3.** Если  $E\eta > 0$ ,  $\lambda(n) = o(n)$ , тогда для любой последовательности оптимальных размещений  $\{x^*\}$

$$\lim_{n \rightarrow \infty} \frac{n}{\lambda(n)} W(x^*) - 0,5\beta_2 = 0.$$

Доказательство. Для всякого размещения  $x = \{x_1, \dots, x_n\}$  такого, что  $\max_{1 \leq i \leq n} \lambda_i \beta_{i1} < 1$ , выполняется неравенство

$$W(x) \geq \frac{1}{2n} \sum_{i=1}^n \lambda_i \beta_{i2},$$

где

$$\beta_{i2} = E \left( \frac{(2|\xi - x_i| + \eta)^2}{\xi} \in C_i \right) \geq E\eta^2 = \beta_2;$$

$C_i$  — зона влияния станции  $x_i$ , а интенсивность потока поступающих на нее требований есть  $\lambda_i = \lambda(n)P(C_i)$ . С учетом этого получаем оценку снизу для  $W(x)$

$$W(x) \geq 0,5 \frac{\lambda(n)}{n} \beta_2.$$

Пусть  $K_i = \{u \in \mathbb{R}^2 : p(u) = p_i > 0\}$  — измеримые по Лебегу непересекающиеся множества. Множество и его мера Лебега  $\mu$  иногда для краткости будем обозначать одной и той же буквой. Построим размещение  $x^0$ . Для этого каждое множество  $K_i$  заменим элементарным множеством  $L_i$  таким, что  $\mu(K_i \Delta L_i) < \varepsilon$ , затем  $L_i$  замещаем конгруэнтными  $N$ -мерными кубами, пересекающимися лишь по границе и объемом  $\sigma_i = K_i/m_i$ , где

$$m_i = m(1 - \delta) \frac{p_i^{N/(N+1)} K_i}{\sum_j p_j^{N/(N+1)} K_j};$$

$m$  — некоторое натуральное число и  $0 < \delta < 1$ . Здесь множество и его мера Лебега обозначены одной и той же буквой.

Если  $\mu(\sigma_i \cap L_i) > 0$ , то в центр куба  $\sigma_i$  помещается станция. Пусть  $n_i$  — число таких станций. Построим последовательность вложенных расширяющихся кубов  $K$  с центром в нуле, которые удовлетворяют условию  $\mu(K) = o(m)$ .  $([m\delta] + 1)$  станций разместим равномерно на множествах  $(K_i \setminus L_i) \cap K$ .

Итак, построено некоторое размещение  $x = \{x_1, \dots, x_n\}$ , где  $n$  — общее число станций.

Оценим сверху  $W(x^*)$ . По определению  $W(x^*) \leq W(x^0)$ . Для  $\{x^0\}$

$$nW(x^0) \leq \frac{1}{2} \sum_i n_i \lambda p_i \left( \sigma_i^{(N+2)/N} \frac{N}{N+2} + \right)$$

$$+ 2\beta_1 \frac{N}{N+1} \sigma_i^{(N+1)/N} + \beta_2 \sigma_i) / \left( 1 - \lambda p_i \frac{N}{N+1} \sigma_i^{(N+1)/N} + \beta_1 \sigma_i \right).$$

Знаменатель  $1 - \lambda p_i \left( \frac{N}{N+1} \sigma_i^{(N+1)/N} + \beta_1 \sigma_i \right)$  имеет порядок  $1 + o(1)$ , так как  $\lambda(n) = o(n)$  и

$$p_i \sigma_i^{(N+1)/N} = \left( \sum_j p_j^{N/(N+1)} K_j \right)^{(N+1)/N} n^{-(N+1)/N} + o(n^{-(N+1)/N}) = O(n^{-(N+1)/N});$$

$$p_i \sigma_i = p_i^{1/(N+1)} \left( \sum_j p_j^{N/(N+1)} K_j \right) n^{-1} + o(n^{-1}) = O(n^{-1}).$$

Таким образом,

$$nW(x^0) \leq \frac{\lambda \beta_2}{2} (1 + o(1)) \sum_i p_i K_i.$$

Устремляя  $m$ , а тем самым и  $n$ , к бесконечности и учитывая оценку снизу для  $W$ , получаем утверждение леммы

$$\lim_{n \rightarrow \infty} \frac{n}{\lambda(n)} W(x^*) - 0,5\beta_2 = 0.$$

Введем некоторые обозначения.

Пусть  $G$  — некоторый компакт на носителе плотности  $p$ ,  $D_n(x) = \max_{1 \leq i \leq n} \text{diam } A_i$ ,  $A_i = \{u \in \in A'_i : p(u) > 0\}$ ,  $A'_i$  — зона влияния  $x_i$  на компакте  $G$ .

**Лемма 4.** Если для размещения  $\{x\}$

$$\lim_{n \rightarrow \infty} \frac{n}{\lambda(n)} W(x) - 0,5\beta_2 = 0,$$

тогда

$$\lim_{n \rightarrow \infty} D_n(x) = 0.$$

**Доказательство.** Предположим, что  $\limsup_{n \rightarrow \infty} D_n(x) = d > 0$ . Тогда выберем подпоследовательность размещений такую, что для некоторой станции  $y_n$  из  $n$ -го размещения этой подпоследовательности  $\lim_{n \rightarrow \infty} |u_n - v_n| = d$ , где  $u_n, v_n \in C_n$ ,  $C_n = \{u \in C'_n : p(u) > 0\}$ ,  $C'_n$  — зона влияния станции  $y_n$  на  $G$ .

Ввиду компактности  $G$  без ограничения общности можно считать, что  $y_n \rightarrow y_0, u_n \rightarrow u_0, v_n \rightarrow v_0$ .

Пусть  $|u_0 - y_0| \geq d/2$ . Это означает, что лишь конечное число размещений из выбранной подпоследовательности имеет станции в  $d/4$ -окрестности точки  $u_0$ . Через  $R$  обозначим  $d/8$ -окрестность точки  $u_0$ .

Поскольку  $u_0$  является предельной точкой  $C_0$  и  $C_0$  — подмножество носителя плотности, то для некоторого  $\varepsilon > 0$

$$\int_R p(u) du > \varepsilon.$$

Оценим снизу значение критерия  $W$  на размещениях, не имеющих станций в  $d/4$ -окрестности точки  $u_0$ :

$$\beta_{i2} = \beta_2 + \frac{4}{P(C_i)} \int_{C_i} |u - x_i|^2 p(u) du + \frac{4\beta_1}{P(C_i)} \int_{C_i} |u - x_i| p(u) du;$$

$$W(x) = \frac{1}{2n} \sum_{i=1}^n \frac{\lambda_i \beta_{i2}}{1 - \lambda_i \beta_{i1}} \geq \frac{\lambda(n)}{2n} \sum_{i=1}^n P(C_i) \beta_{i2}.$$

Учитывая, что на множестве  $R$   $|u - x_i| > d/8$  для любого  $i$  на выбранном размещении, имеем

$$\begin{aligned} \frac{2n}{\lambda(n)} W(x) &\geq \beta_2 + 4 \int_R \min_{1 \leq i \leq n} |u - x_i|^2 p(u) du + \\ &+ 4\beta_1 \int_R \min_{1 \leq i \leq n} |u - x_i| p(u) du \geq \\ &\geq \beta_2 + 4\varepsilon \left( \frac{d}{8} \right)^2 + 4\beta_1 \varepsilon \frac{d}{8} > 0. \end{aligned}$$

Тем самым

$$\lim_{n \rightarrow \infty} \frac{n}{\lambda(n)} W(x) - 0,5\beta_2 > 0,$$

что противоречит условию теоремы.

Поэтому  $d = 0$ .

**Теорема 3.** Если  $p^{N/(N+1)}$  интегрируема по Лебегу, тогда для всякой последовательности оптимальных размещений  $\{x^*\}$

$$\begin{aligned} \liminf_{n \rightarrow \infty} n^{1/N} \left( \frac{n}{\lambda(n)} W(x^*) - 0,5\beta_2 \right) &\geq \\ &\geq \frac{N}{N+1} \beta_1 |p|_{N/(N+1)}. \end{aligned}$$

Доказательство. Пусть плотность  $p$  определяется так же, как в лемме 3,  $J$  — конечное множество индексов. Тогда

$$\mathbb{R}^N = \bigcup_{i=1}^n C_i \supset \bigcup_{j \in J} K_j.$$

Следовательно,

$$\begin{aligned} n\lambda^{-1}W(x) - 0,5\beta_2 &\geq \\ &\geq \sum_{j \in J} 2p_j \left( \int_{K_j} \min_{1 \leq i \leq n} |u - x_i|^2 du + \right. \\ &\quad \left. + \beta_1 \int_{K_j} \min_{1 \leq i \leq n} |u - x_i| du \right). \end{aligned}$$

Так как  $K_j$  — ограниченное, измеримое по Лебегу множество, то для него найдется такое элементарное множество  $L_j$ , что  $\mu(K_j \setminus L_j) < \varepsilon K_j$ ,  $0 < \varepsilon < 1$ ,  $\mu(L_j \setminus K_j) = 0$ . Пусть

$$L_j = \bigcup_{i \in I_j} P_i,$$

$I_j$  — конечное множество индексов,  $P_i$  — непересекающиеся  $N$ -мерные параллелепипеды.

Число станций, попадающих в  $P_i$  при оптимальном размещении  $x^*$ , обозначим через  $n_i$ . Пусть  $B_i$  — параллелепипед объема  $P_i(1 - \varepsilon)$ , гомотетичный  $P_i$  относительно его центра симметрии. Обозначим через  $z^m = \{z_1, \dots, z_{n_i}\}$  оптимальное размещение  $n_i$  станций для  $p\mathbf{1}_{B_i}$  по критерию  $E \min_{1 \leq l \leq n_i} |u - x_l|^m$  для  $m$  равного 1 или 2.

В силу леммы 4 при достаточно больших  $n$  зоны влияния станций, находящихся в  $P_i^c$ , не пересекаются с  $B_i$ , поэтому

$$\int_{B_i} \min_{1 \leq l \leq n} |u - x_l|^m du \geq \int_{B_i} \min_{1 \leq l \leq n_i} |u - z_l|^m du.$$

Обозначим через  $A_l$  зону влияния станции  $z_l$  на  $B_i$ . Применяя лемму 2, имеем

$$\begin{aligned} \int_{B_i} \min_{1 \leq l \leq n_i} |u - z_l|^m du &= \sum_{l=1}^{n_i} \int_{A_l} |u - z_l|^m du \geq \\ &\geq n_i \int_{\sigma_i} |u|^m du = C n_i \sigma_i^{(N+m)/N}, \end{aligned}$$

где  $\sigma_i$  —  $N$ -мерный куб с центром в нуле и объемом  $B_i/n_i$ ,  $C = (1/2)(N/(N+1))$  для  $m = 1$  и  $C = (1/4)(N/(N+2))$  для  $m = 2$ .

Следовательно,

$$\begin{aligned} \int_{K_j} \min_{1 \leq i \leq n} |u - x_i|^m du &\geq \sum_{i \in I_j} C n_i \sigma_i^{(N+m)/N} = \\ &= C \sum_{i \in I_j} n_i^{-m/N} B_i^{(N+m)/2} \geq \\ &\geq C n_j^{-m/N} K_j^{(N+m)/N} (1 - \varepsilon)^{N+m}, \end{aligned}$$

где  $n_j$  — число станций размещения  $x_{K_j}^*$ .

При выводе последнего соотношения было использовано неравенство Гельдера с параметрами  $(-N/m, N/(N+m))$ . Далее,

$$\begin{aligned} n\lambda^{-1}W(x) - 0,5\beta_2 &\geq \\ &\geq \sum_{j \in J} p_j \left( \frac{1}{2} \frac{N}{N+2} n_j^{-2/N} K_j^{(N+2)/N} (1 - \varepsilon) + \right. \\ &\quad \left. + \frac{N}{N+1} \beta_1 n_j^{-1/N} K_j^{(N+1)/N} \right) (1 - \varepsilon)^{N+1} \geq \\ &\geq \frac{1}{2} \frac{N}{N+2} n^{-2/N} (1 - \varepsilon)^{N+2} \times \\ &\quad \times \left( \sum_{j \in J} p_j^{N/(N+2)} K_j \right)^{(N+2)/N} + \\ &\quad + \frac{N}{N+1} \beta_1 n^{-1/N} (1 - \varepsilon)^{N+1} \times \\ &\quad \times \left( \sum_{j \in J} p_j^{N/(N+1)} K_j \right)^{(N+1)/N}. \end{aligned}$$

Здесь дважды было применено неравенство Гельдера с параметрами  $(N/(N+2), -N/2)$  и  $(N/(N+1), -N)$ .

Из интегрируемости  $p^{N/(N+1)}$  следует неравенство

$$\begin{aligned} \liminf_{n \rightarrow \infty} n^{1/N} \left( \frac{n}{\lambda(n)} W(x^*) - 0,5\beta_2 \right) &\geq \\ &\geq \frac{N}{N+1} \beta_1 |p|_{N/(N+1)} (1 - \varepsilon)^{N+1}. \end{aligned}$$

Так как  $\varepsilon$  может быть выбрано произвольно малым, то

$$\begin{aligned} \liminf_{n \rightarrow \infty} n^{1/N} \left( \frac{n}{\lambda(n)} W(x^*) - 0,5\beta_2 \right) &\geq \\ &\geq \frac{N}{N+1} \beta_1 |p|_{N/(N+1)}. \end{aligned}$$

## 5 Доказательство основных результатов

Приведем только доказательство теоремы 1; доказательство теоремы 2 проводится аналогично. Укажем способ построения таких размещений  $\{x\}$ , на которых неравенство теоремы 3 превращается в равенство. Такие размещения будем называть асимптотически оптимальными второго порядка по критерию  $W$ .

Так как  $E|\xi|^2 < \infty$ , то можно выбрать последовательность вложенных расширяющихся кубов  $K$  с центром в нуле таких, что  $E|\xi|^2 \mathbf{1}_{K^c} = o(m^{-1/N})$ .

Так как  $K_i$  — измеримые по Лебегу множества, следовательно, для любого  $\varepsilon > 0$  найдется такое элементарное множество  $L_i$ , что

$$\mu(K_i \Delta L_i) < \frac{\varepsilon}{K} \text{ для } \forall i.$$

Затем каждое множество  $L_i$  замещаем конгруэнтными  $N$ -мерными кубами, пересекающимися по границе и объемом  $\sigma_i = K_i/m_i$ , где

$$m_i = m(1 - \delta) \frac{p_i^{N/(N+1)} K_i}{\sum_{j=1}^r p_j^{N/(N+1)} K_j}, \quad 0 < \delta < 1.$$

Если  $\mu(\sigma_i \cap L_i) > 0$ , то в центр  $\sigma_i$  помещается станция. Пусть  $n_i$  — число таких  $\sigma_i$ . ( $[m\delta] + 1$ ) станций разместим равномерно на множествах  $(K_i \setminus L_i) \cap K$ . Итак, построено некоторое размещение  $x = \{x_1, \dots, x_n\}$ , где  $n$  — общее число станций.

Тогда

$$\begin{aligned} \frac{n}{\lambda} W(x) &\leq \sum_i n_i \left( p_i \left( 0,5\sigma_i^{(N+2)/N} \frac{N}{N+2} + \right. \right. \\ &\quad \left. \left. + \beta_1 \frac{N}{N+1} \sigma_i^{(N+1)/N} \right) \right) / \left( 1 - \right. \\ &\quad \left. - \lambda p_i \left( \frac{N}{N+1} \sigma_i^{(N+1)/N} + \beta_1 \sigma_i \right) \right) + \\ &\quad + \frac{1}{2} \sum_i \frac{\beta_2 p_i K_i}{1 + o(n^{-1/N})} + o(n^{-1/N}) = \\ &= (1 + o(1)) \sum_i \left( 0,5 \frac{N}{N+2} p_i^{(N-1)/(N+1)} K_i \times \right. \\ &\quad \times \left( \sum_j p_j^{N/(N+1)} K_j \right) (n(1 - \delta))^{-2/N} + \\ &\quad + \frac{N}{N+1} \beta_1 p_i^{N/(N+1)} K_i \left( \sum_j p_j^{N/(N+1)} K_j \right)^{1/N} \times \\ &\quad \times (n(1 - \delta))^{-1/N} + 0,5\beta_2 \left( 1 + o(n^{-1/N}) \right). \end{aligned}$$

Совершая предельный переход, получаем

$$\begin{aligned} \limsup_{n \rightarrow \infty} n^{1/N} \left( \frac{n}{\lambda(n)} W(x) - 0,5\beta_2 \right) &\leq \\ &\leq \frac{N}{N+1} \beta_1 |p|_{N/(N+1)}. \end{aligned}$$

Из этого неравенства и теоремы 3 следует утверждение первого пункта теоремы 1 в случае, когда плотность  $p$  простая.

Пусть  $p$  — произвольная функция, удовлетворяющая условиям доказываемой теоремы. Введем простые функции  $p_k(u)$  по правилу:  $p_k(u) = m/k$ , если  $m/k < p(u) \leq (m+1)/k$ ,  $k \in \mathcal{N}$ ,  $m = 0, 1, \dots$

В определении критерия  $W$  заменим  $p$  на  $p_k$  и полученное выражение обозначим через  $W_k$ . Для простых функций было доказано, что

$$\begin{aligned} n^{1/N} \left( \frac{n}{\lambda(n)} W_k(x^*) - 0,5\beta_2 \right) &= \\ &= \frac{N}{N+1} \beta_1 |p_k|_{N/(N+1)} + o(1). \end{aligned}$$

Применим теорему Лебега о предельном переходе под знаком интеграла к обеим частям равенства, тогда

$$\begin{aligned} n^{1/N} \left( \frac{n}{\lambda(n)} W(x) - 0,5\beta_2 \right) &= \\ &= \frac{N}{N+1} \beta_1 |p|_{N/(N+1)} + o(1). \end{aligned}$$

Устремляя  $n$  к бесконечности, получим

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{1/N} \left( \frac{n}{\lambda(n)} W(x) - 0,5\beta_2 \right) &= \\ &= \frac{N}{N+1} \beta_1 |p|_{N/(N+1)}. \end{aligned}$$

Метод доказательства второго пункта теоремы 1 для случая  $N = 2$  изложен в статье [3], и его не сложно обобщить на случай произвольного  $N$ .

## Литература

1. *Ивченко Г. И., Капитанов В. А., Коваленко И. Н.* Теория массового обслуживания. — М.: Высшая школа, 1982.
2. *Захарова Т. В.* Оптимизация расположения станций обслуживания на плоскости // Изв. АН СССР. Техн. киберн., 1987. № 6. С. 83–91.
3. *Захарова Т. В.* Размещения систем массового обслуживания, минимизирующие среднюю длину очереди // Информатика и её применения, 2008. Т. 2. Вып. 1. С. 35–38.
4. *Том Ф. Л.* Расположение на плоскости, на сфере и в пространстве. — М.: ГИФМЛ, 1958.

# ОТСЕВ ЭКТОПИЧЕСКИХ ИМПУЛЬСОВ ИЗ РИТМОГРАММЫ С ИСПОЛЬЗОВАНИЕМ РОБАСТНЫХ ОЦЕНОК

А. В. Маркин<sup>1</sup>, О. В. Шестаков<sup>2</sup>

**Аннотация:** Предложен метод удаления эктопических импульсов из ритмограммы на основе собственной математической модели. Параметры модели оцениваются с использованием робастного варианта линейной регрессии. Отсев эктопических импульсов производится с помощью доверительных интервалов для разностей длин  $RR$ -интервалов. Приведены результаты работы метода на реальных данных.

**Ключевые слова:** ритмограмма; робастные оценки; линейная регрессия; доверительные интервалы

## 1 Введение

Ритмограмма  $RR$ -интервалов как мера вариабельности сердечного ритма (ВСР) получила в последние годы достаточно широкое распространение в диагностике сердечно-сосудистых заболеваний (ССЗ) [1, 2]. Благодаря развитию компьютерной техники стал возможным анализ десятков тысяч записей и выявление общих закономерностей. В частности, оказалось, что существует зависимость между ССЗ и спектральными характеристиками ритмограммы. Пример электрокардиограммы (ЭКГ) и ритмограммы приведен на рис. 1. График электрокардиограммы построен для промежутка времени 81,25–86 с, ритмограмма же — для пятиминутного промежутка 10–310 с. Часть ритмограммы, соответствующая промежутку 81,25–86 с, обведена пунктирным прямоугольником.

В структуре ЭКГ выделяют несколько волн:  $P$ ,  $Q$ ,  $R$ ,  $S$  и  $T$  (см. рис. 1). Эти волны отражают различные состояния работы сердца. Под  $RR$ -интервалами подразумеваются временные промежутки между пиками соседних  $R$ -волн.

Существует ряд рекомендаций относительно построения ритмограммы. Они устанавливают минимальную продолжительность измерений, при которой, во-первых, возможно вычислить спектр для определенного диапазона частот и, во-вторых, спектральные характеристики будут постоянны [1]. Однако в ритмограмме обычно присутствуют так называемые эктопические импульсы — очень короткие или очень длинные интервалы между биениями. Обычно сначала идет короткий  $RR$ -интервал, а сразу после него — длинный. Эти импульсы имеют иную природу, в отличие от нормальных (или,

как их еще называют, синусовых) интервалов [1, 3]. На рис. 1 видно, что в районе восьмидесяти четвертой секунды структура кардиограммы отклоняется от обычной, пик  $R$ -волны находится близко к предыдущему и далеко от следующего. На ритмограмме это отражается в появлении двух скачков: сначала вниз, затем вверх. Это и есть эктопические импульсы.

Наличие всего нескольких эктопических импульсов оказывает существенное влияние на спектр ритмограммы, поэтому обычно их стараются удалить. Один из методов удаления эктопических импульсов предложен в настоящей статье. Стоит отметить, что ритмограмма относится к сигналам с неравномерным квантованием времени, так как длина отсчета времени равна длине очередного  $RR$ -интервала. Это необходимо учитывать при вычислении спектральных характеристик.

## 2 Классический метод отсева

Весьма популярный метод удаления эктопических импульсов состоит в следующем [2]:

1. Удалить все интервалы короче 300 мс и длиннее 2000 мс.
2. Удалить все интервалы, каждый из которых отличается от предыдущего более чем на 200 мс.
3. Удалить интервалы, отличающиеся по длине более чем на 20% от среднего значения, где среднее вычисляется по пяти предыдущим синусовым интервалам.

<sup>1</sup>Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета ВМиК, artem.v.markin@mail.ru

<sup>2</sup>Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета ВМиК, oshestakov@cs.msu.su



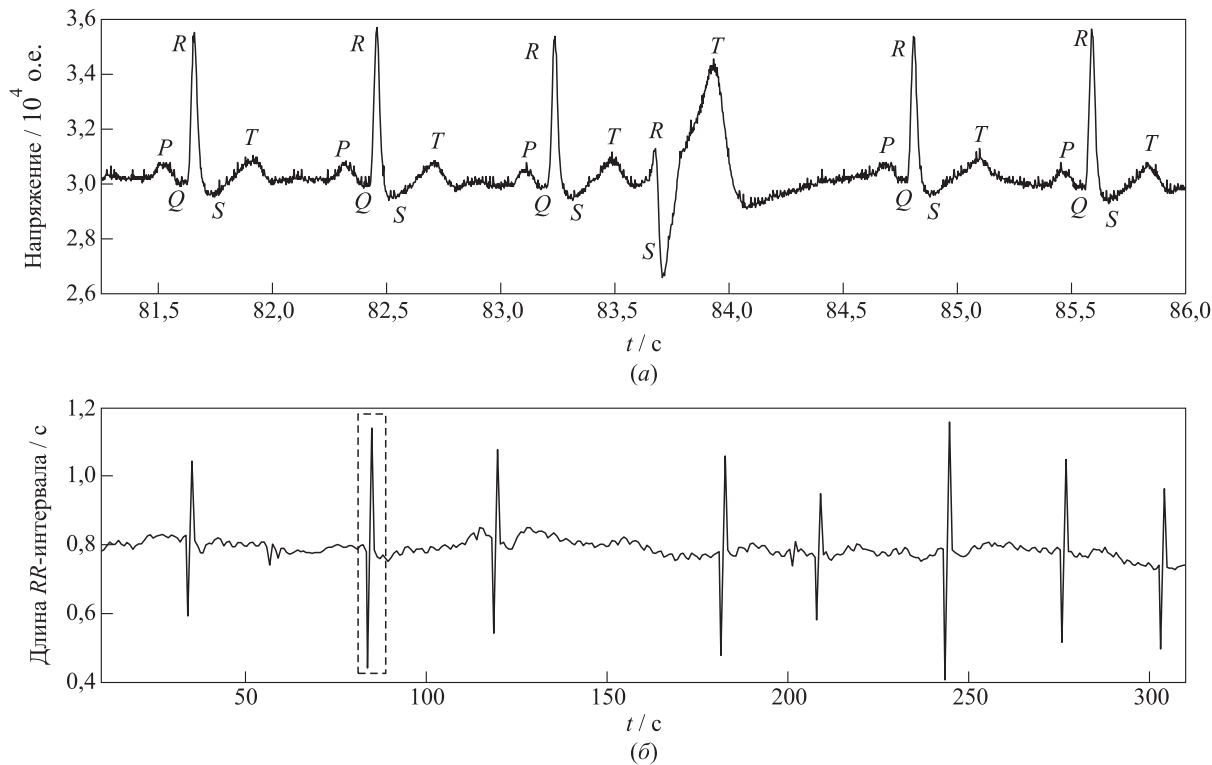


Рис. 1 Электрокардиограмма (а) и соответствующая ей ритмограмма (б)

Критические значения вычислены по базе ЭКГ Массачусетского технологического института (MIT-BIH Normal Sinus Rhythm Database), насчитывающей десятки тысяч записей. Эти значения являются некоторым усреднением, и поэтому вполне может случиться так, что какие-то эктопические значения не будут отсеяны и останутся в сигнале. При таком подходе фактически не учитывается форма (структура) ритмограммы и ее индивидуальные особенности (амплитуда колебаний). В предлагаемом методе присутствует адаптация под конкретный сигнал.

### 3 Математические модели ритмограммы

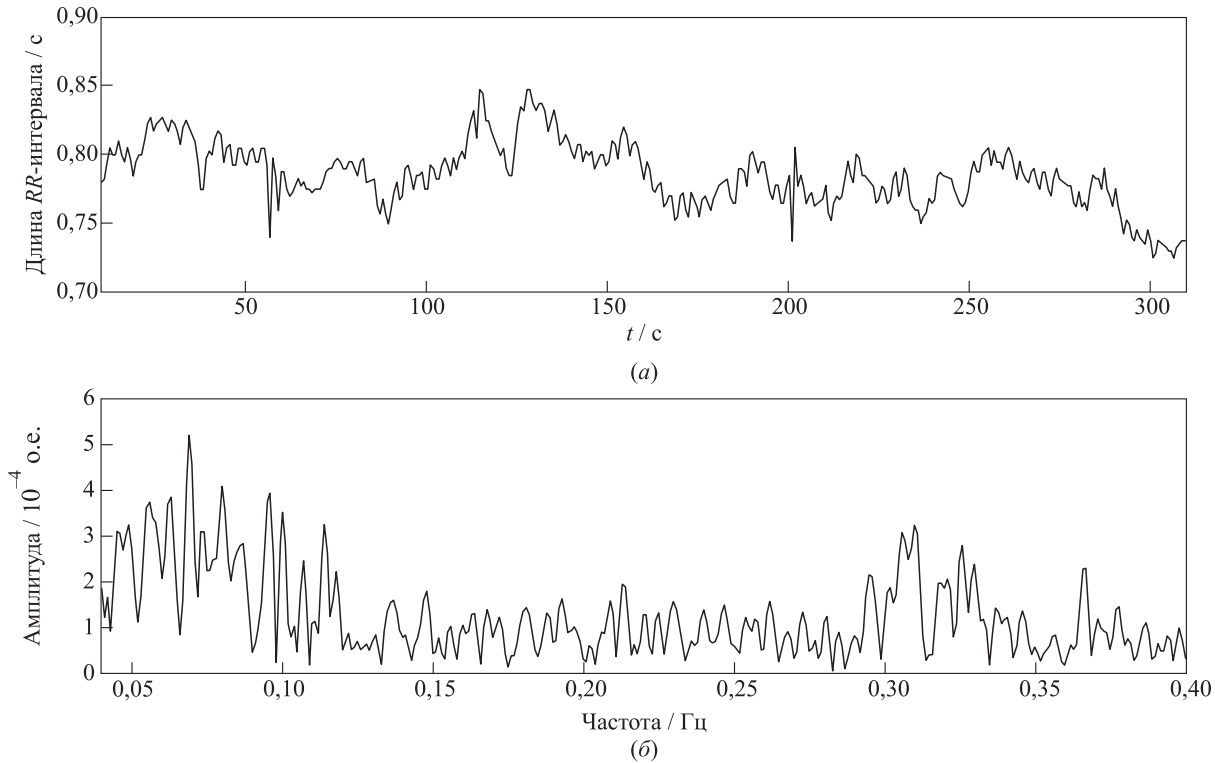
Для сравнения методов обработки ритмограмм можно использовать как реальные данные, так и математическую модель. Хотя модель всегда является лишь приближением, она обычно обладает вполне определенными параметрами, что позволяет оценивать методы непосредственно по точности вычисления этих параметров.

На практике для спектрального анализа обычно используются ритмограммы двух типов: короткие (5 мин) и длинные (24 ч). Частотные характери-

сти 5-минутных ритмограмм можно считать постоянными при выполнении ряда условий, тогда как измерения, полученные за 24 ч, таким свойством не обладают [1]. При этом длинные записи несут более полную спектральную информацию. Далее будут рассматриваться только короткие ритмограммы.

При работе с 5-минутными ритмограммами выделяют два диапазона частот: диапазон высоких частот (ВЧ) (0,15–0,4 Гц) и диапазон низких частот (НЧ) (0,04–0,15 Гц). Анализ данных показал [1, 2], что в этих диапазонах часто можно четко выделить два биоритма: волну Майера (Mayer wave) в НЧ области и дыхательную волну (Respiratory Sinus Arrhythmia wave) в ВЧ области. Волна Майера соответствует колебаниям кровяного давления, а дыхательная волна отражает влияние дыхания на сердечный ритм. На рис. 2 представлена ритмограмма из рис. 1 с эктопическими импульсами, удаленными вручную, и график амплитуды ее частотных компонент. Хорошо видны пики в НЧ и ВЧ областях.

Как видим, на спектре имеются два подъема с четкими максимумами, по одному в НЧ и ВЧ диапазонах. Это наталкивает на мысль о представлении ритмограммы в виде суммы двух синусоид с круговыми частотами, соответствующими пикам



**Рис. 2** Ритмограмма (а) и график амплитуды частотных компонент (б). Эктопические импульсы удалены вручную

спектра в НЧ и ВЧ областях. Таким образом, получаем представление:

$$RR(t) = A_1 \sin(\omega_1 t + \phi_1) + A_2 \sin(\omega_2 t + \phi_2) + \epsilon(t) + C. \quad (1)$$

Здесь  $RR(t)$  — ритмограмма;  $\omega_1 = 2\pi f_1$  и  $\omega_2 = 2\pi f_2$  — круговые частоты, соответствующие частотам  $f_1$  и  $f_2$ , где  $0,04 \leq f_1 \leq 0,15$  Гц,  $0,15 < f_2 \leq 0,4$  Гц;  $A_1$  и  $A_2$  — амплитуды;  $\phi_1$  и  $\phi_2$  — фазы синусоид;  $\epsilon(t)$  — шум, нормально распределенные случайные величины с нулевым средним и дисперсией  $\sigma^2$ ;  $C$  — некоторая константа, выполняющая роль сдвига графика ритмограммы относительно нуля.

Время меняется дискретно:

$$t \in \{t_1, t_2, \dots, t_N \mid 0 = t_1 < t_2 < \dots < t_N\}.$$

Следует отметить, что данная модель является весьма грубой, так как форма реальной ритмограммы несколько сложнее, чем сумма двух синусоид. Однако она вполне достаточна, ведь стоит задача оценки амплитуды синусовой части ритмограммы, а синусовая часть является довольно «гладкой» в смысле малых изменений в соседних точках, в то время как эктопические импульсы соответствуют скачкам на графике ритмограммы.

## 4 Построение доверительных интервалов

Идея предлагаемого метода состоит в следующем:

1. Оценить параметры  $\omega_1$ ,  $\omega_2$ ,  $\phi_1$ ,  $\phi_2$ ,  $A_1$  и  $A_2$  по спектру с помощью робастного варианта регрессии.
2. Оценить  $\sigma^2$ .
3. Построить доверительные интервалы для разностей  $RR(t_i) - RR(t_{i-k})$ ,  $i = 1, 2, \dots, N$ ,  $1 \leq k \leq i$ , с уровнем доверия не ниже  $1 - \gamma$ . В качестве  $\gamma$  можно взять, например,  $\gamma = 0,05$ .
4. Отсеять импульсы, не попадающие в доверительные интервалы для разностей процесса  $RR(t)$ .

Для разностей гармонической части  $RR_h(t) \equiv A_1 \sin(\omega_1 t + \phi_1) + A_2 \sin(\omega_2 t + \phi_2)$  будет выполнено:

$$\begin{aligned} |RR_h(t_{i+k}) - RR_h(t_i)| &\leq \\ &\leq 4A_1 \left| \sin \left( \frac{\omega_j(t_{i+k} - t_i)}{2} \right) \right| + \\ &+ 4A_2 \left| \sin \left( \frac{\omega_j(t_{i+k} - t_i)}{2} \right) \right| \equiv h_i^k. \quad (2) \end{aligned}$$

При построении доверительных интервалов предполагается, что дисперсия шума  $\sigma^2$  известна. Тогда нормальная случайная величина с нулевым средним и дисперсией  $\sigma^2$  с вероятностью  $1 - \gamma$  попадает в интервал  $(\sigma\Phi^{-1}(\gamma/2), \sigma\Phi^{-1}(1 - \gamma/2))$ , где  $\Phi(x)$  — функция распределения стандартной нормальной случайной величины. Причем длина этого интервала минимальна среди всех интервалов  $\mathcal{I}$  таких, что  $P(\mathcal{I}) = 1 - \gamma$  ( $P$  — распределение стандартной нормальной случайной величины) [4].

Доверительные интервалы для разностей  $RR(t)$  будут выглядеть так:

$$\left( -h_i^k + \sqrt{2} \cdot \sigma\Phi^{-1}\left(\frac{\gamma}{2}\right), h_i^k + \sqrt{2} \cdot \sigma\Phi^{-1}\left(1 - \frac{\gamma}{2}\right) \right). \quad (3)$$

Однако  $\sigma$  необходимо оценить. Для этого можно использовать разности  $\Delta RR_i = RR(t_i) - RR(t_{i-1})$  функции  $RR(t)$ , исключив некоторым образом из этих разностей влияние полезного сигнала. Предлагается делать это в соответствии с моделью (1): в качестве шума рассматривать значения

$$r_i = \Delta RR_i - (RR_h(t_i) - RR_h(t_{i-1})).$$

Коэффициенты  $RR_h$  оцениваются методом робастной регрессии (см. разд. 5).

В качестве оценки для  $\sqrt{2} \cdot \sigma$  используем нормированное абсолютное медианное отклонение от медианы  $\sigma^*$ :

$$\sigma^* = \frac{\text{med}|r_i - \text{med}r_i|}{\Phi^{-1}(3/4)}.$$

Эта оценка является наиболее робастной, ее пороговая точка (см. разд. 5) равна 0,5. Кроме того,  $\sigma^*$  будет состоятельной оценкой стандартного отклонения  $\sigma$  нормальной случайной величины [5]. Таким образом, осталось вычислить  $\omega_1, \omega_2, \phi_1, \phi_2, A_1$  и  $A_2$ .

## 5 Метод робастной регрессии

Оценку параметров предлагается провести робастным вариантом линейной регрессии. Под робастностью понимают нечувствительность к небольшим отклонениям от начальных предположений. Следует отметить, что наличие больших ошибок в измерениях, составляющих небольшую часть выборки, считается небольшим отклонением. В рассматриваемом случае такими ошибочными измерениями являются эктопические импульсы.

Классическое решение задачи линейной регрессии методом наименьших квадратов строится следующим образом [5]. Необходимо оценить  $p$  неизвестных параметров  $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$  по  $N$  наблюдениям  $y = (y_1, y_2, \dots, y_N)^T$ , причем имеет место связь:

$$y = X\theta + u,$$

где  $x_{ij}$  — некоторые известные постоянные коэффициенты,  $u_i$  — независимые одинаково распределенные случайные величины,  $i = 1, \dots, N$ ,  $j = 1, \dots, p$ .

Эта задача сводится к решению задачи:

$$X^T X \theta = X^T y. \quad (4)$$

Если матрица  $X$  имеет ранг  $p$ , то решение системы (4) будет выглядеть так:

$$\hat{\theta} = (X^T X)^{-1} X^T y.$$

Формула (4) не является робастной, так как квадратичная функция слишком чувствительна к ошибкам, особенно большим. Выход заключается в замене ее на некоторую менее быстро растущую функцию остатков  $\rho$  и решении следующей задачи:

$$\sum_{i=1}^N \rho \left( y_i - \sum_{j=1}^p x_{ij} \theta_j \right) \rightarrow \min. \quad (5)$$

Ограничения на  $\rho$ : выпуклость, немонотонность, наличие ограниченных производных достаточно высоких порядков. В частности,  $\psi = \rho'$  должна быть непрерывной ограниченной функцией [5]. В этом случае не появятся новые корни при переходе от задачи (5) к системе

$$\sum_{i=1}^N \psi \left( y_i - \sum_{k=1}^p x_{ik} \theta_k \right) x_{ij} = 0, \quad j = 1, \dots, p, \quad \psi = \rho'. \quad (6)$$

В качестве  $\rho$  предлагается использовать следующую функцию [6]:

$$\hat{\rho}(x) = \begin{cases} c \left( 1 - \cos \left( \frac{x}{c} \right) \right) & \text{при } |x| < c\pi, \\ 2c & \text{при } |x| \geq c\pi. \end{cases}$$

Ее производная равна

$$\hat{\psi}(x) = \begin{cases} \sin \left( \frac{x}{c} \right) & \text{при } |x| < c\pi, \\ 0 & \text{при } |x| \geq c\pi. \end{cases}$$

Вблизи нуля функция  $\hat{\rho}$  имеет квадратичный порядок роста, а начиная с некоторого значения — нулевой. В качестве  $c$  берется значение  $\hat{c} = 0,85$ ,

которое было подобрано опытным путем. Полезным свойством данной функции является то, что очень большие по модулю значения аргумента не оказывают влияния на сумму (6), так как значение функции  $\hat{\psi}$  в этих точках равно нулю. Это полностью согласуется с практикой, которая говорит, что очень длинные ( $\geq 2000$  мс) и очень короткие интервалы ( $\leq 300$  мс) однозначно являются эктопическими. Именно такие интервалы дают большие абсолютные значения разностей.

В робастном варианте регрессии возникает задача оценки масштаба, так как функции  $\rho$  и  $\psi$  практически всегда имеют разный порядок роста для разных значений аргумента. Соответственно, параметр масштаба  $s$  позволяет выполнить нормировку данных для эффективного деления на большие и маленькие значения. Вместо системы (6) нужно решать

$$\sum_{i=1}^N \psi \left( \frac{y_i - \sum_{k=1}^p x_{ik} \theta_k}{s} \right) x_{ij} = 0, \quad j = 1, \dots, p. \quad (7)$$

Значения  $s$  и  $\theta$  находятся итеративным путем, но так как функции  $\sum_j x_{ij} \theta_j$  являются линейными относительно  $\theta_j$ , то решение находится за один шаг. В качестве начального значения параметра масштаба  $\tilde{s}$  можно взять абсолютное медианное отклонение от медианы (АМО)

$$\tilde{s} = \text{med} |RR_i - \text{med}(RR_i)|,$$

которое обладает хорошими теоретическими свойствами, в частности АМО имеет пороговую точку  $\epsilon^* = 0,5$ . На интуитивном уровне понятие пороговой точки можно объяснить как максимальную долю паразитных импульсов, с которыми может справиться оценка.

**Определение 1 [5].** Асимптотическая пороговая точка  $\epsilon^*$  оценки  $T_n$  для распределения  $F_0$ , принятого в модели, определяется следующим образом:

$$\epsilon^* = \epsilon^*(T_n, F_0) \equiv \sup\{\epsilon | b(\epsilon) < b(1)\}.$$

Здесь

$$b(\epsilon) = \lim_{n \rightarrow \infty} \sup_{F \in \mathcal{P}_\epsilon} |M(F, T_n)|,$$

где  $M(F, T_n)$  — медиана распределения  $\mathcal{L}_F[T_n - T(F_0)]$ .

В качестве  $\mathcal{P}_\epsilon$  может выступать окрестность Леви:

$$\mathcal{P}_\epsilon(F_0) = \{F | \forall t F_0(t - \epsilon) - \epsilon \leq F(t) \leq F_0(T + \epsilon) + \epsilon\}$$

или окрестность загрязнения:

$$\mathcal{P}_\epsilon(F_0) = \{F | F = (1 - \epsilon)F_0 + \epsilon H\},$$

где  $H$  — некоторое (загрязняющее) распределение.

Максимально возможное значение пороговой точки равно 0,5, и для оценки  $\tilde{s}$  она этого значения достигает. Однако если известно, что доля эктопических выбросов в сигнале не превышает  $0 < \alpha < 0,5$ , то в качестве оценки масштаба можно взять оценку с пороговым значением не меньше  $\alpha$ , которая лучше оценит разброс полезных данных. В качестве такой оценки предлагается взять половину разности между квантилями порядка  $\alpha$  и  $1 - \alpha$ :

$$\hat{s} = \frac{(RR_{(1-\alpha)} - RR_{(\alpha)})}{2}.$$

**Предложение 1 [5].** Если  $\alpha > 0$ , то при условии, что мера  $\mu$  не имеет ненулевой массы в точках разрыва функции  $F_0^{-1}$ , функционал  $T(F) \equiv \int F^{-1}(s)\mu(ds)$  слабо непрерывен в  $F_0$ . Пороговая точка  $\epsilon^*$  удовлетворяет неравенству  $\epsilon^* \geq \alpha$ . Если  $\mu$  — положительная мера, то  $\epsilon^* = \alpha$  и равенство  $\alpha = 0$  влечет за собой разрывность  $T$ .

Стало быть, пороговое значение для оценки  $\hat{s}$  не меньше  $\alpha$ . Начальной оценкой для  $\theta$  является решение задачи (4), которое обозначим  $\theta^0$ . Шаг метода заключается в вычислении винзоризованных остатков  $r_i$ :

$$r_i = \psi \left( \frac{y_i - \sum_{k=1}^p x_{ik} \theta_k^0}{\hat{s}} \right) \hat{s}, \quad i = 1, \dots, N,$$

и решении относительно  $\tau$  системы

$$\mathbf{X}^T \mathbf{X} \tau = \mathbf{X}^T \mathbf{r}. \quad (8)$$

Если решение (8) обозначить  $\hat{\tau}$ , то решение  $\hat{\theta}$  системы (7) равно

$$\hat{\theta} = \theta^0 + \hat{\tau}. \quad (9)$$

В качестве факторов регрессии предлагается взять ортогонализированные единичный и тригонометрические векторы. Под единичным подразумевается вектор, полностью составленный из единиц. Такой вид линейной регрессии называется компенсированным дискретным преобразованием Фурье (КДПФ) (date-compensated discrete Fourier transform) [7]. Дело в том, что неравномерное квантование времени, а именно такое квантование имеет место при записи ритмограммы, нарушает ортогональность и нормировку тригонометрических векторов  $\mathbf{X}_2$ ,  $\mathbf{X}_3$  и единичного вектора  $\mathbf{X}_1$ :

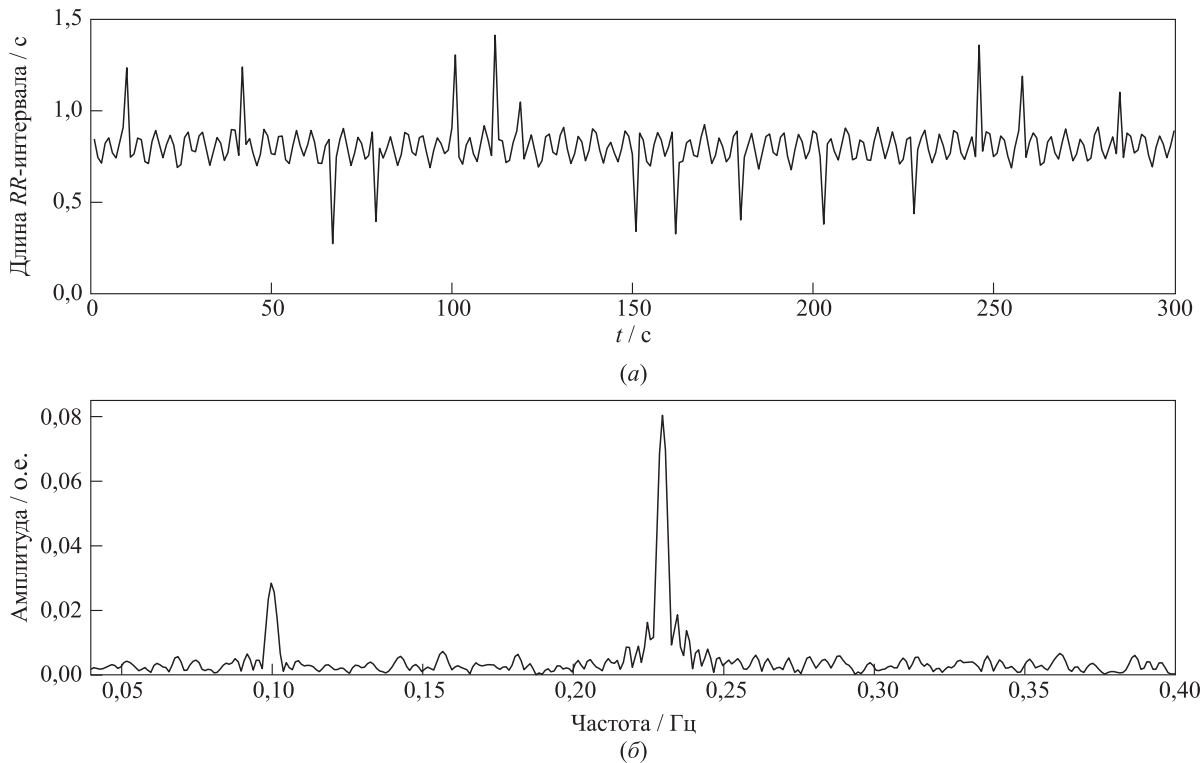


Рис. 3 Модельная функция (а) и график амплитуды частотных компонент (б)

$$\mathbf{X}_1 = (1, 1, \dots, 1); \quad (10)$$

$$\mathbf{X}_2 = (\cos(\omega t_1), \cos(\omega t_2), \dots, \cos(\omega t_N)); \quad (11)$$

$$\mathbf{X}_3 = (\sin(\omega t_1), \sin(\omega t_2), \dots, \sin(\omega t_N)). \quad (12)$$

В результате амплитуда частотных компонент оценивается неправильно, причем в отдельных случаях отличия от реальных значений амплитуд весьма существенные [8].

Эту проблему решает КДПФ. К трем векторам (10)–(12) применяется процедура ортогонализации Грама–Шмидта:

$$\mathbf{x}_1 = a_1 \mathbf{X}_1;$$

$$\mathbf{x}_2 = a_2 [\mathbf{X}_2 - \mathbf{x}_1(\mathbf{x}_1, \mathbf{X}_2)];$$

$$\mathbf{x}_3 = a_3 [\mathbf{X}_3 - \mathbf{x}_1(\mathbf{x}_1, \mathbf{X}_3) - \mathbf{x}_2(\mathbf{x}_2, \mathbf{X}_3)],$$

где круглыми скобками обозначено скалярное произведение. Коэффициенты  $a_1$ ,  $a_2$  и  $a_3$  находятся из условия ортонормированности векторов  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  и  $\mathbf{x}_3$ . Здесь частота  $\omega$  фиксирована и  $p = 3$ . Поэтому коэффициенты регрессии выглядят так:

$$\theta_1 = (\mathbf{z}, \mathbf{x}_1) = a_1(\mathbf{z}, \mathbf{X}_1);$$

$$\theta_2 = (\mathbf{z}, \mathbf{x}_2) = a_2 [(\mathbf{z}, \mathbf{X}_2) - a_1\theta_1(\mathbf{X}_1, \mathbf{X}_2)];$$

$$\theta_3 = (\mathbf{z}, \mathbf{x}_3) = a_3 [(\mathbf{z}, \mathbf{X}_3) - a_1\theta_1(\mathbf{X}_1, \mathbf{X}_3) - a_2\theta_2[(\mathbf{X}_2, \mathbf{X}_3) - a_1^2(\mathbf{X}_1, \mathbf{X}_2)(\mathbf{X}_1, \mathbf{X}_2)]] .$$

Подставляя вместо вектора  $\mathbf{z}$  вектор наблюдений  $\mathbf{y}$  и вектор винзоризованных остатков  $\mathbf{r}$ , получим, соответственно, векторы коэффициентов  $\theta^0$  и  $\hat{\tau}$ . Используя (9), находим значение амплитуды для частоты  $\omega$ :

$$A(\omega) = \sqrt{\frac{\hat{\theta}_2^2 + \hat{\theta}_3^2}{N/2}} .$$

Остается посчитать для некоторого набора частот  $\omega^l$ ,  $0,04 \leq \omega^l/(2\pi) \leq 0,4$  Гц ( $l$  — индекс, а не степень) значения соответствующих амплитуд и найти максимумы в НЧ и ВЧ:

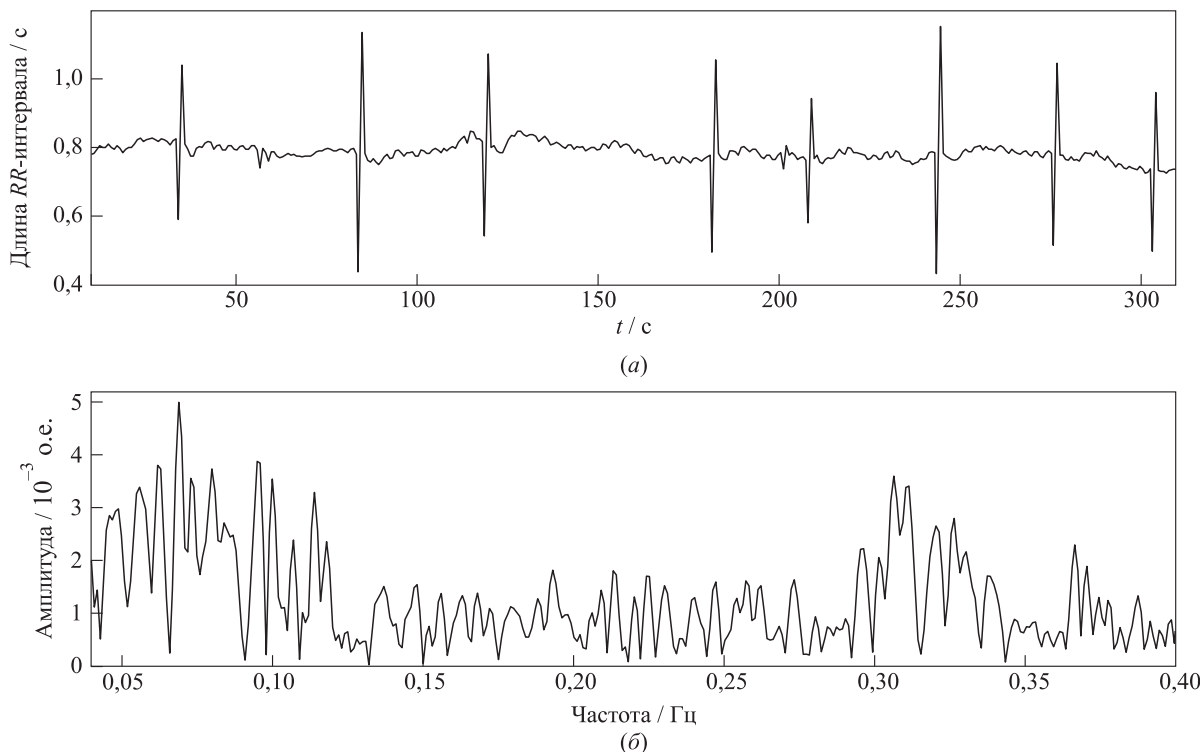
$$A_1 = \max_{0,04 \leq \omega^l \leq 0,15} \{A(\omega^l)\};$$

$$A_2 = \max_{0,15 < \omega^l \leq 0,4} \{A(\omega^l)\} .$$

Соответствующие им  $\omega_1$ ,  $\omega_2$ ,  $\phi_1$  и  $\phi_2$  будут искомыми для модели (1). Приводимые ниже результаты были получены для набора частот с шагом 0,001 Гц.

## 6 Отсев эктопических импульсов

Теперь, имея оценки всех параметров модели, можно приступить непосредственно к фильтрации эктопических выбросов. Фильтрация будет производиться во временной области на основе (2) и (3).



**Рис. 4** Исходная ритмограмма (а) и робастная оценка амплитуды частотных компонент (б)

Допустим, что  $\alpha = 0,05$ , т. е. доля несинусовых импульсов в сигнале не превышает 5%, что при пульсе 75 ударов в минуту означает не более 20 импульсов для 5-минутных записей. В качестве значения доверительного уровня возьмем  $1 - \gamma$ .

Первая точка ритмограммы должна относиться к синусовому ритму, это необходимо для правильной работы алгоритма. Предположим, что до момента времени  $t_{i-1}$  включительно ритмограмма отфильтрована, т. е. последний из предшествующих синусовых интервалов соответствует моменту времени  $t_{i-1}$ . Рассмотрим момент времени  $t_{i+k}$ ,  $k$  — целое число. Начинаем со значения  $k = 0$ .

Если разность  $RR(t_{i+k}) - RR(t_{i-1})$  попадает в доверительный интервал, построенный по формуле (3), то точка  $RR(t_{i+k})$  соответствует модели (1), она объявляется синусовой, обновляются значения  $i$  и  $k$ :  $i = i + k + 1$ ,  $k = 0$ . Выполняется переход к моменту времени  $t_i$  ( $i$  — обновленное).

Если же разность в интервал не попала, то точка  $RR(t_{i+k})$  удаляется из ритмограммы как эктопический импульс. Далее  $k$  увеличивается на единицу, и рассматривается момент  $t_{i+k}$  ( $k$  — обновленное).

Необходимо отметить, что обычно  $k \leq 3$ , так что существенного увеличения доверительного ин-

тервала и, как следствие, ухудшения работы метода ожидать не стоит.

## 7 Результаты

Выясним, как робастная регрессия оценивает параметры модельной функции. В качестве таковой использовалась функция

$$y(t) = 0,08 \sin(0,23 \cdot 2\pi t + 1,0) + 0,03 \sin(0,1 \cdot 2\pi t + 2,0) + \epsilon(t) + 0,8,$$

где  $\epsilon(t) \sim \mathcal{N}(0, 0,01^2)$ ,  $t = 1, 2, \dots, 300$ . Коэффициенты выбраны близкими к тем, что получаются при анализе реальных данных. К этой функции добавлено 15 паразитных импульсов. Функция и амплитуда частотных компонент, оцененная методом робастной регрессии, приведены на рис. 3. Частоты оценены без погрешности, пики находятся ровно в точках 0,1 Гц и 0,23 Гц, погрешности оценок амплитуд равны соответственно 1,6% и 0,3%.

Теперь рассмотрим результаты оценки спектральных характеристик реальных данных. На рис. 4 приведена ритмограмма с рис. 1 и 2, но предварительного ручного удаления эктопических импульсов, как на рис. 2, не проводилось. Оценка спектра методом робастной регрессии дает точ-

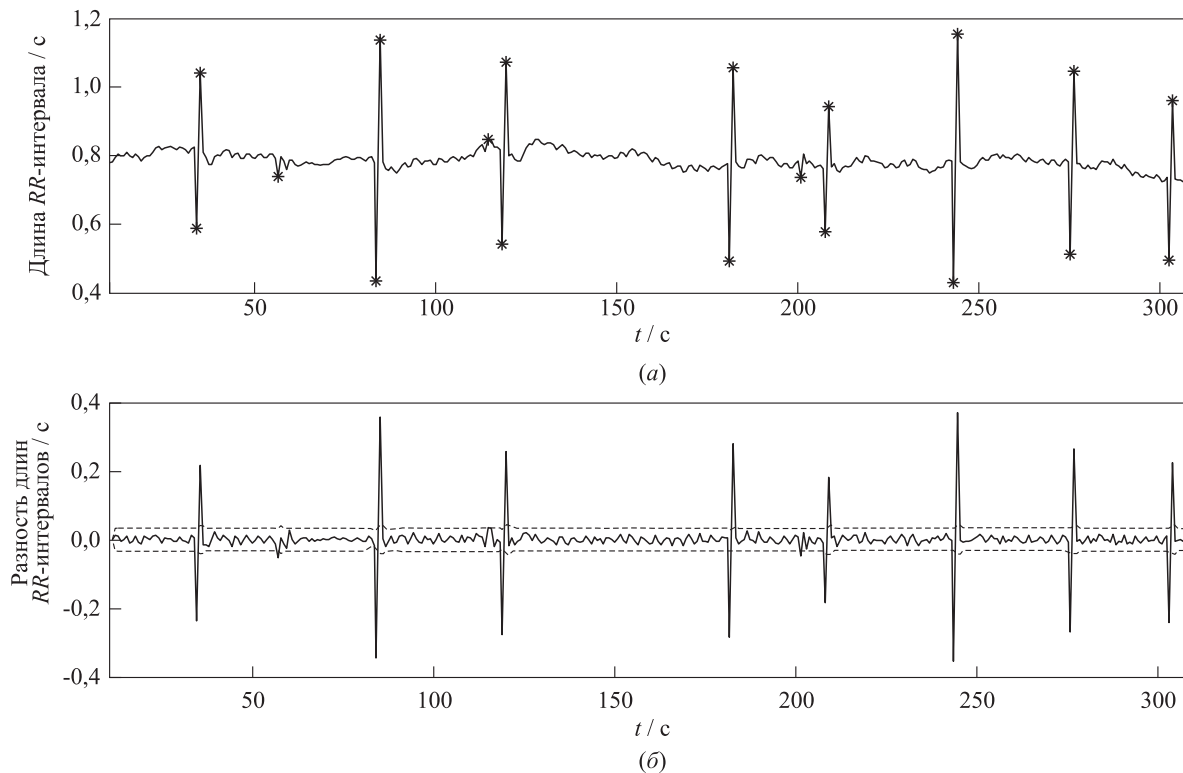


Рис. 5 Исходная ритмограмма (а) и разности и доверительные интервалы для разностей (б)

ную оценку низкочастотного пика, 0,079 Гц, погрешность оценки высокочастотного пика равна 1,3% (0,31 Гц — исходное значение, 0,306 Гц — оцененное). Погрешность оценки амплитуд пиков составляет 4,0% и 11,6% соответственно.

Фильтрация ритмограммы проводилась путем построения доверительных интервалов для разностей с использованием полученных значений частот и амплитуд. Доверительный уровень был выбран равным 95%, т. е.  $\gamma = 0,05$ . Результаты представлены на рис. 5. Пунктирные линии обозначают границы доверительных интервалов, а звездочками помечены те точки ритмограммы, которые были удалены в результате работы метода.

Как видим, все эктопические импульсы успешно удалены. Кроме того, были удалены три точки (56,58, 114,7 и 201,2 с), которые давали разности, немного выходящие за границы доверительного интервала. Такого эффекта можно избежать, задав меньшее значение  $\gamma$ , т. е. если построить интервалы с более высоким доверительным уровнем.

## 8 Заключение

В статье предложен новый метод отсева эктопических импульсов ритмограммы, основанный на собственной математической модели сигнала и ро-

бадном подходе к оценке параметров этой модели. Проведена демонстрация эффективной работы метода на реальных данных.

## Литература

1. *Malik M.* Heart rate variability // *European Heart J.*, 1996. Vol. 17. No. 3. P. 354–381.
2. *Azuaje F., Clifford G., McSharry P.* Advanced methods and tools for ECG data analysis. — Artech House Publishers, 2005. 384 p.
3. *Clifford G., McSharry P., Tarassenko L.* Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold // *Computers in Cardiology*, 2002. P. 129–132.
4. *Ивченко Г. И., Медведев Ю. И.* Математическая статистика. — М.: Высш. шк., 1984. 248 с.
5. *Хьюбер П.* Робастность в статистике. — М.: Мир, 1984. 304 с.
6. *Andrews D.* A robust method for multiple linear regression // *Technometrics*, 1974. Vol. 16. No. 4. P. 523–531.
7. *Ferraz-Mello S.* Estimation of periods from unequally spaced observations // *The Astronomical J.*, 1981. Vol. 86. No. 4. P. 619–624.
8. *Foster G.* The cleanest Fourier spectrum // *The Astronomical J.*, 1995. Vol. 109. No. 4. P. 1889–1902.

# ОБ АСИМПТОТИЧЕСКОМ РАСПРЕДЕЛЕНИИ МАКСИМАЛЬНОЙ ПОРЯДКОВОЙ СТАТИСТИКИ В ВЫБОРКЕ СЛУЧАЙНОГО ОБЪЕМА\*

В. И. Пагурова<sup>1</sup>

**Аннотация:** Исследуется асимптотическое распределение при  $n \rightarrow \infty$  нормированного максимума в предположении, что случайный объем выборки представим в виде суммы  $n$  независимых одинаково распределенных величин. Данная работа является обобщением работы [1], в которой объем выборки имеет распределение Пуассона с параметром  $n$ . Для однопараметрического семейства распределений, зависящего от неизвестного параметра сдвига, исследуется скорость сходимости распределения нормированного максимума к предельному закону. Рассматриваются классы распределений с экспоненциальными и степенными хвостами.

**Ключевые слова:** случайно индексированный максимум; однопараметрическое семейство распределений; скорость сходимости

## 1 Введение

Пусть случайные величины  $X_1, \dots, X_n$  независимы и одинаково распределены (н.о.р.) с общей абсолютно непрерывной функцией распределения (ф.р.)  $F(x - \theta)$ ,  $\theta$  — неизвестный параметр,  $X_1^{(n)} \leq X_2^{(n)} \leq \dots \leq X_n^{(n)}$  — соответствующий вариационный ряд,  $N$  — неотрицательная целочисленная случайная величина, не зависящая от  $X_1, \dots, X_n$ , причем  $N$  имеет представление

$$N = \sum_{i=1}^n \xi_i, \quad (1)$$

где  $\xi_1, \dots, \xi_n$  — н.о.р. случайные величины,

$$\mathbf{E}\xi_1 = \alpha > 0, \quad 0 < \mathbf{E}\xi_1^2 < \infty. \quad (2)$$

Исследуется влияние оценки неизвестного параметра  $\theta$  на асимптотическое поведение разности  $X_N^{(N)} - \hat{\theta}_N$  при  $n \rightarrow \infty$ ,  $\hat{\theta}_N$  — состоятельная оценка параметра  $\theta$ , построенная по наблюдениям  $X_1, \dots, X_N$ . Частный случай данной задачи, когда  $N$  имеет распределение Пуассона с параметром  $n$ , был исследован в работе [1]. В качестве других примеров распределения  $N$ , удовлетворяющего условиям (1) и (2), можно указать биномиальное

распределение  $b(n, \alpha)$ ,  $0 < \alpha < 1$ , и отрицательное биномиальное распределение  $\bar{b}(n, p)$ ,  $\alpha = (1 - p)/p$ .

Сначала рассмотрим случай известного параметра  $\theta$ . Без нарушения общности положим  $\theta = 0$ . Введем величину  $y = y_n(t)$ , являющуюся решением уравнения

$$F(y) = 1 - \frac{t}{n}, \quad t > 0, \quad (3)$$

тогда

$$M_n(t) = \mathbf{P}\{X_N^{(N)} < y\} = \mathbf{E}_N \left(1 - \frac{t}{n}\right)^N = \left(\mathbf{E} \left(1 - \frac{t}{n}\right)^{\xi_1}\right)^n.$$

Рассмотрим

$$Z_n = \left(1 - \frac{t}{n}\right)^{\xi_1} = 1 - \frac{t\xi_1}{n} + \frac{\xi_1(\xi_1 - 1)(1 - \eta)^{\xi_1 - 2} t^2}{2n^2}, \quad 0 < \eta < \frac{t}{n}.$$

Учитывая условия (1) и (2), получим при  $n \rightarrow \infty$

$$\mathbf{E}Z_n = 1 - \frac{\alpha t}{n} + O\left(\frac{1}{n^2}\right)$$

\* Работа выполнена при поддержке РФФИ, грант 08-01-00567.

<sup>1</sup> Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, pagurova@yandex.ru



равномерно в любом конечном интервале  $t \in (0, t_0]$ .  
Поэтому при  $n \rightarrow \infty$

$$M_n(t) = \exp(-\alpha t) + O(n^{-1}).$$

## 2 Скорость сходимости

Обозначим через (А) класс распределений  $F(x)$ , для которых при  $x \rightarrow \infty$

$$\begin{aligned} 1 - F(x) &\sim ax^\gamma \exp(-bx^\beta); \\ F'(x) = f(x) &\sim ab\beta x^{\gamma+\beta-1} \exp(-bx^\beta); \\ |f'(x)| &\sim ab^2\beta^2 x^{\gamma+2(\beta-1)} \exp(-bx^\beta), \\ a, b, \beta &> 0, \quad -\infty < \gamma < \infty. \end{aligned}$$

Символом (В) обозначим класс распределений  $F(x)$ , для которых при  $x \rightarrow \infty$

$$\begin{aligned} 1 - F(x) &\sim ax^{-\gamma}, \quad f(x) \sim a\gamma x^{-\gamma-1}; \\ |f'(x)| &\sim a\gamma(\gamma+1)x^{-\gamma-2}, \quad a, \gamma > 0. \end{aligned}$$

Для величины  $y = y_n(t)$ , определяемой соотношением (3), при  $n \rightarrow \infty$  и фиксированном  $t > 0$  для класса (А) получаем

$$\begin{aligned} y &= O((\ln n)^{1/\beta}), \\ f(y) &= O(n^{-1}(\ln n)^{(\beta-1)/\beta}), \\ f'(y) &= O(n^{-1}(\ln n)^{2(\beta-1)/\beta}), \end{aligned}$$

для класса (В) получаем

$$\begin{aligned} y &= O(n^{1/\gamma}), \\ f(y) &= O(n^{-1-1/\gamma}), \\ f'(y) &= O(n^{-1-2/\gamma}). \end{aligned}$$

Распределения классов (А) и (В) играют важную роль при моделировании экстремальных событий (см., например, [2]). Далее рассматривается величина

$$L_n(t) = \mathbf{P}\{X_N^{(N)} - \hat{\theta}_N < y_n(t)\}, \quad (4)$$

где  $X_1, \dots, X_n$  — н.о.р. случайные величины с общей ф.р.  $F(x - \theta)$ ;  $N$  — неотрицательная целочисленная случайная величина, не зависящая от  $X_1, \dots, X_n$ , удовлетворяющая соотношениям (1) и (2);  $\hat{\theta}_N$  — оценка параметра  $\theta$ , построенная по наблюдениям  $X_1, \dots, X_N$ ;  $y_n(t)$  определяется соотношением (3).

**Теорема 1.** Пусть  $F(x)$  является непрерывной ф.р., распределение  $F$  симметрично относительно нуля,  $f(0) > 0$ ,  $f'(x)$  ограничена в окрестности нуля,

$\hat{\theta}_N = X_{[N/2]+1}^{(N)}$ , тогда при  $n \rightarrow \infty$  равномерно в любом конечном интервале  $t \in (0, t_0]$

$$L_n(t) = \exp(-\alpha t) + O(r_n),$$

где

$$r_n = \begin{cases} \max(n^{-1}, n^{-1}(\ln n)^{2(\beta-1)/\beta}) & \text{для класса (А),} \\ n^{-1} & \text{для класса (В).} \end{cases}$$

Доказательство будем проводить по плану доказательства теоремы 1 в [1]. Заметим, что

$$L_n(t) = \mathbf{E}_N J_N;$$

$$\begin{aligned} J_N &= \mathbf{P}\{X_N^{(N)} - X_{[N/2]+1}^{(N)} < y|N\} = \\ &= C_N \int_{-\infty}^{\infty} F^{[N/2]}(u)(F(u+y) - \\ &\quad - F(u))^{N-[N/2]-1} f(u) du = \\ &= C_N \int_{-\infty}^{\infty} H_N(u) f(u) du, \quad (5) \end{aligned}$$

где

$$\begin{aligned} H_N(u) &= F(u) (1 - \\ &\quad - F(u))^{N-2[N/2]} M^{[N/2]-1}(u) G^{N-[N/2]-1}(u); \quad (6) \end{aligned}$$

$$\begin{aligned} C_N &= \frac{N!}{[N/2]!(N - [N/2] - 1)!}; \\ G(u) &= \frac{F(u+y) - F(u)}{1 - F(u)}; \\ M(u) &= F(u)(1 - F(u)). \end{aligned}$$

Оценим

$$\begin{aligned} \mathbf{E}_N C_N \int_{-\infty}^{-n^{-1/4}} H_N(u) f(u) du &\leq \\ &\leq \mathbf{E}_N C_N 2^{-2[N/2]+2} \left( 1 - \frac{4f^2(0)}{\sqrt{n}} + \right. \\ &\quad \left. + o\left(\frac{1}{\sqrt{n}}\right) \right)^{[N/2]-1}, \quad (7) \end{aligned}$$

так как  $0 \leq G(u) \leq 1$ , а на интервале  $u \in (-\infty, -n^{-1/4})$  имеем

$$\begin{aligned} M(u) &\leq F(-n^{-1/4})(1 - F(-n^{-1/4})) = \\ &= \frac{1}{4} \left( 1 - \frac{4f^2(0)}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \right). \end{aligned}$$

Используя формулу Стирлинга, получим при  $n \rightarrow \infty$

$$C_n \left( \frac{\sqrt{2\pi}}{\sqrt{n} \cdot 2^n} \right) \rightarrow 1. \quad (8)$$

Так как  $N \xrightarrow{P} \infty$  при  $n \rightarrow \infty$ , отсюда получим

$$C_N \left( \frac{\sqrt{2\pi}}{\sqrt{N} \cdot 2^N} \right) \xrightarrow{P} 1,$$

а так как  $N = \alpha n + o_p(n)$  при  $n \rightarrow \infty$ , тогда правая часть соотношения (7) при  $n \rightarrow \infty$  имеет порядок  $o(r_n)$ . Аналогичным образом при  $n \rightarrow \infty$

$$\mathbf{E}_N C_N \int_{n^{-1/4}}^{\infty} H_N(u) f(u) du = o(r_n).$$

Таким образом,

$$\mathbf{E}_N J_N = \mathbf{E}_N C_N \int_{-n^{-1/4}}^{n^{-1/4}} H_N(u) f(u) du + o(r_n), \quad (9)$$

$H_N(u)$  определяется соотношением (6). Обозначим

$$k_N = N - \left\lfloor \frac{N}{2} \right\rfloor - 1, \quad (10)$$

тогда на интервале  $u \in (-n^{-1/4}, n^{-1/4})$  имеем

$$\begin{aligned} G^{k_N}(u) &= G^{k_N}(0) + k_N G^{k_N-1}(0) G'(u) u + \\ &+ (k_N(k_N-1) G^{k_N-2}(\xi) (G'(\xi))^2 + \\ &+ k_N G^{k_N-1}(\xi) G''(\xi)) \frac{u^2}{2}, \quad |\xi - u| < n^{-1/4}; \quad (11) \end{aligned}$$

$$G'(u) = \frac{f(u+y)(1-F(u)) - f(u)(1-F(u+y))}{(1-F(u))^2};$$

$$\begin{aligned} G''(u) &= ((1-F(u))^2 f'(u+y) + \\ &+ 2(1-F(u)) f(u) f(u+y) - \\ &- (1-F(u))(1-F(u+y)) f'(u) - \\ &- 2f^2(u)(1-F(u+y))(1-F(u))^{-3}. \end{aligned}$$

Учитывая соотношение (3), получим при  $n \rightarrow \infty$

$$A_n = \sup_{|u| \leq n^{-1/4}} |G'(u)| = O(\max(f(y), n^{-1})); \quad (12)$$

$$\begin{aligned} B_n &= \sup_{|u| \leq n^{-1/4}} |G''(u)| = \\ &= \begin{cases} O(\max(f'(y), n^{-1})) & \text{для класса (A),} \\ O(n^{-1}) & \text{для класса (B),} \end{cases} \quad (13) \end{aligned}$$

на интервале  $u \in (-n^{-1/4}, n^{-1/4})$  имеем

$$\begin{aligned} G(u) &= 1 - \frac{2t}{n} + o(n^{-1}), \\ G(0) &= 1 - \frac{2t}{n}, \quad G'(0) = O(\max(f(y), n^{-1})). \quad (14) \end{aligned}$$

Подставим представление (11) в выражение (5) и почленно проинтегрируем. Получим

$$\mathbf{E}_N J_N = \mathbf{E}_N G^{k_N}(0) + G'(0) \mathbf{E}_N K_{1N} + \mathbf{E}_N K_{2N}, \quad (15)$$

где

$$\begin{aligned} K_{1N} &= k_N G^{k_N-1}(0) C_N \int_{-\infty}^{\infty} F^{[N/2]}(1 - \\ &- F(u))^{N-[N/2]-1} u f(u) du; \end{aligned}$$

$$\begin{aligned} K_{2N} &\leq (A_n^2 k_N (k_N - 1) + \\ &+ k_N B_n) C_N \int_{-\infty}^{\infty} F^{[N/2]}(u) (1 - \\ &- F(u))^{N-[N/2]-1} u^2 f(u) du; \end{aligned}$$

$A_n$  и  $B_n$  определяются соотношениями (12) и (13).

Покажем, что при  $n \rightarrow \infty$

$$\begin{aligned} D_n &= k_n C_n \int_{-\infty}^{\infty} F^{[n/2]}(u) (1 - \\ &- F(u))^{n-[n/2]-1} u f(u) du = O(1). \end{aligned}$$

Тогда, учитывая, что

$$\frac{N}{\alpha n} \xrightarrow{P} 1; \quad G^{k_n-1}(0) \rightarrow \exp(-t) \quad (n \rightarrow \infty),$$

получим

$$G^{k_N-1}(0) \xrightarrow{P} \exp(-\alpha t), \quad D_N = O_p(1).$$

Имеем

$$\begin{aligned} D_n &= k_n C_n \int_0^{\infty} (F(u) (1 - \\ &- F(u))^{[n/2]-1} (2F(u) - 1) u f(u) du = \\ &= \frac{k_n C_n}{n} \int_0^{n^{-1/4}} (F(u) (1 - F(u))^{[n/2]}) du + o\left(\frac{1}{n}\right) = \\ &= \frac{k_n C_n C([n/2]!)^2}{n(n+1)!} \mathbf{P} \left\{ \frac{1}{2} \leq \beta_{[n/2]+1, [n/2]+1} \leq \right. \\ &\left. \leq \frac{1}{2} + \frac{f(0)}{n^{1/4}} + o\left(\frac{1}{n^{1/4}}\right) \right\} + o\left(\frac{1}{n}\right) = O(1). \end{aligned}$$

В последнем соотношении учтено поведение  $f(u)$  в окрестности нуля, соотношения (8)–(10) и формула Стирлинга,  $\beta_{k,n}$  означает случайную величину, имеющую бета-распределение с параметрами  $(k, n)$ . Таким образом, при  $n \rightarrow \infty$

$$\mathbf{E}_N K_{1N} = O(1).$$

Аналогичным образом при  $n \rightarrow \infty$

$$\mathbf{E}_N K_{2N} = O(r_n),$$

$r_n$  определено в формулировке теоремы. С учетом соотношений (14) и (15) получим доказательство теоремы.

**Теорема 2.** Пусть  $X_1, \dots, X_n$  — н.о.р. случайные величины с общей непрерывной ф.р.  $F(x - \theta)$ ,  $N$  означает неотрицательную целочисленную величину, не зависящую от  $X_1, \dots, X_n$  и удовлетворяющую условиям (1) и (2),  $F(0) = 0$ ,  $f(0) > 0$ ,  $f'(x)$  ограничена в окрестности нуля,  $L_n(t)$  определяется соотношением (4), где  $\hat{\theta}_N = X_1^{(N)}$ , тогда при  $n \rightarrow \infty$  равномерно по  $t$  в любом конечном интервале  $t \in (0, t_0]$

$$L_n(t) = \exp(-\alpha t) + O(r_n),$$

где

$$r_n = \begin{cases} \max(n^{-1}, n^{-1}(\ln n)^{(\beta-1)/\beta}) & \text{для класса (A)}, \\ n^{-1} & \text{для класса (B)}. \end{cases}$$

Доказательство. Как и прежде,

$$L_n(t) = \mathbf{E}_N J_N, \quad (16)$$

где

$$\begin{aligned} J_N &= \mathbf{P}\{X_N^{(N)} - X_1^{(N)} < y | N\} = \\ &= N \int_0^\infty (F(u+y) - F(u))^{N-1} f(u) du = \\ &= N \int_0^\infty G^{N-1}(u)(1 - F(u))^{N-1} f(u) du = \\ &= \left(1 - \frac{t}{n}\right)^{N-1} + (N-1) \int_0^\infty H_{1N}(u) du - \\ &\quad - (N-1) \int_0^\infty H_{2N}(u) du. \end{aligned} \quad (17)$$

Здесь

$$\begin{aligned} H_{1N}(u) &= (F(u+y) - F(u))^{N-2} (1 - F(u)) f(u+y); \\ H_{2N}(u) &= (F(u+y) - F(u))^{N-2} (1 - F(u+y)) f(u). \end{aligned}$$

Рассмотрим

$$\begin{aligned} K_{1N} &= (N-1) \int_{n^{-1/2}}^\infty H_{1N}(u) du < \\ &< (N-1)(1 - F(n^{-1/2})) \int_{n^{-1/2}}^\infty (F(u+y) - F(n^{-1/2}))^{N-2} f(u+y) du = \\ &= (1 - F(n^{-1/2}))^N - (1 - F(n^{-1/2}))(F(n^{-1/2} + y) - F(n^{-1/2}))^{N-1}. \end{aligned}$$

Так как

$$F(n^{-1/2}) = f(0)n^{-1/2} + o(n^{-1/2}),$$

$$\begin{aligned} F(n^{-1/2} + y) &= F(y) + f(\zeta)n^{-1/2} = \\ &= 1 - \frac{t}{n} + o\left(\frac{1}{n}\right), \quad |\zeta - y| < n^{-1/2}, \end{aligned}$$

то при  $n \rightarrow \infty$  получим

$$\mathbf{E}_N K_{1N} = o(r_n)$$

равномерно по  $t$  в любом конечном интервале  $t \in (0, t_0]$ ,  $r_n$  определено в формулировке теоремы. Рассмотрим

$$K_{2N} = (N-1) \int_0^{n^{-1/2}} H_{1N}(u) du.$$

На интервале  $u \in (0, n^{-1/2})$  имеем

$$\begin{aligned} F(u+y) - F(u) &= F(y) + (f(\eta) - f(\xi))u; \\ |\eta - y| &< n^{-1/2}; \quad 0 < \xi < n^{-1/2}; \\ f(u+y) &= O(f(y)). \end{aligned}$$

Тогда при достаточно больших  $n$  существует  $C_1 > 0$  такое, что  $f(\eta) - f(\xi) < -C_1$  и

$$\begin{aligned} K_{2N} &\leq (N-1) \int_0^{n^{-1/2}} (F(y) - C_1 u)^{N-2} du O(f(y)) = \\ &= O(f(y)) \frac{(F(y))^{N-1} - (F(y) - C_1 n^{-1/2})^{N-1}}{C_1}. \end{aligned}$$

Отсюда следует, что при  $n \rightarrow \infty$

$$\mathbf{E}_N K_{2N} = O(f(y)).$$

Далее,

$$(N-1) \int_0^\infty H_{2N}(u) du \leq (1-F(y)) = \frac{t}{n};$$

$$\mathbf{E}_N \left(1 - \frac{t}{n}\right)^{N-1} = \exp(-\alpha t) + O\left(\frac{1}{n}\right)$$

равномерно по  $t$  в любом конечном интервале  $t \in (0, t_0]$ . Учитывая соотношения (16) и (17), получим доказательство теоремы.

**Теорема 2'.** В условиях теоремы 2 пусть  $N$  имеет распределение Пуассона с параметром  $\alpha n$ . Тогда при  $n \rightarrow \infty$  равномерно по  $t$  в любом конечном интервале  $t \in (0, t_0]$

$$L_n(t) = \exp(-\alpha t) + O(r_n),$$

где

$$r_n = \begin{cases} n^{-1}(\ln n)^{(\beta-1)/\beta} & \text{для класса (A),} \\ n^{-1-1/\gamma} & \text{для класса (B).} \end{cases}$$

Теорема 2' была доказана в работе [1]. Сравнение теорем 2 и 2' показывает, что пуассоновское распределение обеспечивает более быструю сходимость к предельному закону, по крайней мере, для

класса (B), чем в общем случае, рассматриваемом в теореме 2.

**Пример.** Пусть

$$F(x) = 1 - \frac{1}{\sqrt{x+1}}, \quad x \geq 0,$$

тогда при  $n \rightarrow \infty$  равномерно в любом конечном интервале  $t \in (0, t_0]$  имеем

$$L_n(t) = \exp(-\alpha t) + O\left(\frac{1}{n^3}\right),$$

если  $N$  имеет распределение Пуассона с параметром  $\alpha n$ ;

$$L_n(t) = \exp(-\alpha t) + O\left(\frac{1}{n}\right),$$

если  $N$  имеет отрицательное биномиальное распределение с параметрами  $(n, p)$ ,  $\alpha = (1-p)/p$ .

## Литература

1. Пагурова В.И. Об асимптотическом распределении случайно индексированного максимума // В сб. Статистические методы оценивания и проверки гипотез. – Пермь, 2005. С. 104–113.
2. Embrechts P., Kluppelberg K., Mikosch T. Modelling of extremal events for finance and insurance. — Berlin–New York: Springer, 1997.

# ОЦЕНИВАНИЕ РАСПРЕДЕЛЕНИЯ ЗАДЕРЖКИ В БИОЛОГИЧЕСКИХ ДИНАМИЧЕСКИХ СИСТЕМАХ НА ПРИМЕРЕ МОДЕЛИ, ОПИСЫВАЮЩЕЙ ВИЧ-ИНФЕКЦИЮ

А. Н. Ушакова<sup>1</sup>

**Аннотация:** Рассмотрена проблема оценивания распределения задержки в динамических системах, описывающих, в частности, взаимодействие вируса иммунодефицита человека (ВИЧ-инфекция) с иммунной системой. Используются параметрический метод, основанный на аппроксимации распределения задержки гамма-распределением, и непараметрический, основанный на регуляризации с выбором параметра регуляризации либо из предварительно оцененной погрешности измерений, либо из предварительно оцененного уровня гладкости распределения задержки.

**Ключевые слова:** системы с задержкой; распределение задержки; параметр регуляризации

## 1 Введение

Одной из важных задач в исследовании ВИЧ-инфекции является понимание явлений, происходящих в организме ВИЧ-инфицированного в течение длительного асимптоматического периода, начинающегося обычно через несколько недель (иногда месяцев) после заражения вирусом. Этот период может продолжаться многие годы и неизбежно заканчивается переходом в стадию СПИДа, когда иммунная система становится не в состоянии противостоять внешним и внутренним инфекциям. Для количественного описания указанных явлений необходимо математическое моделирование инфекционного процесса и противовирусного иммунитета. Модели, с одной стороны, должны достаточно точно описывать реальную ситуацию, а с другой — быть достаточно простыми, чтобы позволять оценивать содержащиеся в них параметры. В настоящее время имеется несколько монографий (среди которых отметим [1]) и большое число статей, посвященных этой проблеме.

Одна из наиболее известных простых моделей взаимодействия ВИЧ с иммунной системой при использовании антиретровирусной терапии (лекарств, которые подавляют вирус на различных стадиях развития, в частности, делают вновь появившиеся вирионы неспособными к инфекции) имеет вид

$$\begin{aligned} \frac{dI}{dt} &= kTV_I(t) - \delta I(t); \\ \frac{dV_I}{dt} &= (1 - \eta)pI(t) - cV_I(t); \\ \frac{dV_{NI}}{dt} &= \eta pI(t) - cV_{NI}(t), \end{aligned} \quad (1)$$

где  $I(t)$  — плотность зараженных клеток иммунной системы в крови пациента, а  $V_I$  и  $V_{NI}$  — концентрация соответственно способных и неспособных к инфекции вирионов [2]. Числа  $k$ ,  $T$ ,  $\delta$ ,  $\eta$ ,  $p$  и  $c$  представляют собой коэффициенты, характеризующие различные аспекты взаимодействия вируса с иммунной системой, а также эффект воздействия лекарства. Однако, как было отмечено в [3–5], необходимо учитывать задержку между моментом заражения вирусом клетки и гибелью клетки с выходом в кровь группы порожденных ею вирусов. Причем эта задержка может варьироваться в пределах 1–2 дней [4], поэтому модели с фиксированной (постоянной) задержкой не в полной мере адекватны. Наиболее подходящей является модель, в которой задержка представляет собой случайную величину.

С учетом этого, уравнение (1) в модели заменяется уравнением

$$\frac{dI}{dt} = kT \int_0^{\infty} V_I(t-t')f(t') dt' - \delta I(t), \quad (2)$$

в котором функцию  $f(t)$  — плотность распределения задержки — необходимо оценить по наблюдениям.

## 2 Оценивание в гамма-модели

Поскольку для рассматриваемых далее в работе методов связь между функциями  $I(t)$ ,  $dI(t)/dt$  и  $V_I(t)$  не играет роли, вместо уравнения (2) будем рассматривать общее уравнение свертки 1-го рода

<sup>1</sup>Норвежский научно-технологический университет (NTNU), Тронхейм, Норвегия, anastasi@math.ntnu.no

$$\int_0^{\infty} K(t-s)z(s) ds = u(t), \quad (3)$$

где  $K(t)$  и  $u(t)$  — наблюдаемые функции, а  $z(t)$  — оцениваемая функция (в нашем случае — плотность распределения задержки). Функции  $K(t)$  и  $u(t)$  предполагаются интегрируемыми, неотрицательными и стремящимися к нулю при  $t \rightarrow \pm\infty$ .

В биологических системах распределения задержек часто хорошо аппроксимируются гамма-распределениями [6]. В настоящем разделе будет предполагаться, что  $z(t)$  является гамма-плотностью, т. е. имеет вид

$$z(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, \quad t > 0, \quad \alpha > 0, \quad \beta > 0,$$

и, таким образом, задача сводится к оценке параметров  $\alpha$  и  $\beta$ . Функции  $K(t)$  и  $u(t)$  наблюдаются в моменты времени  $t_1, \dots, t_n$  со случайными ошибками.

В [4] предлагается оценивать параметры по методу наименьших квадратов. Однако в этом случае приходится минимизировать по  $\alpha$  и  $\beta$  выражение

$$S(\alpha, \beta) = \sum_{j=1}^n \left( \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} K(t_j - s) s^{\alpha-1} e^{-\beta s} ds - u(t_j) \right)^2,$$

что приводит к громоздким и сильно нелинейным уравнениям, очень чувствительным к ошибкам измерений и замене непрерывной модели дискретной. Ниже приводится простой в вычислительном плане метод, достаточно устойчивый к погрешностям измерений.

Пусть  $\mu_z$  и  $\sigma_z^2$  — соответственно математическое ожидание и дисперсия плотности распределения  $z(t)$ . Имеем

$$\alpha = \frac{\mu_z^2}{\sigma_z^2}; \quad \beta = \frac{\mu_z}{\sigma_z^2}.$$

Положим

$$\mu_K = \frac{\int_0^{\infty} tK(t) dt}{\int_0^{\infty} K(t) dt}; \quad \mu_u = \frac{\int_0^{\infty} tu(t) dt}{\int_0^{\infty} u(t) dt} \quad (4)$$

и

$$\sigma_K^2 = \frac{\int_0^{\infty} (t - \mu_K)^2 K(t) dt}{\int_0^{\infty} K(t) dt};$$

$$\sigma_u^2 = \frac{\int_0^{\infty} (t - \mu_u)^2 u(t) dt}{\int_0^{\infty} u(t) dt}.$$

Тогда

$$\mu_z = \mu_u - \mu_K; \quad \sigma_z^2 = \sigma_u^2 - \sigma_K^2$$

и, следовательно,

$$\alpha = \frac{(\mu_u - \mu_K)^2}{(\sigma_u^2 - \sigma_K^2)}; \quad \beta = \frac{\mu_u - \mu_K}{(\sigma_u^2 - \sigma_K^2)}. \quad (6)$$

Предположим, что точки  $t_1, \dots, t_n$  представляют собой равномерную сетку с шагом  $h$ , покрывающую интервал, вне которого функции  $K(t)$  и  $u(t)$  можно считать равными нулю. Параметры  $\mu_K, \mu_u, \sigma_K^2$  и  $\sigma_u^2$  можно оценить, заменяя интегралы в (4) и (5) соответствующими интегральными суммами, т. е. посредством

$$\hat{\mu}_K = \frac{\sum_{i=1}^n t_i K(t_i)}{\sum_{i=1}^n K(t_i)}; \quad \hat{\mu}_u = \frac{\sum_{i=1}^n t_i u(t_i)}{\sum_{i=1}^n u(t_i)}$$

и

$$\hat{\sigma}_K^2 = \frac{\sum_{i=1}^n (t_i - \hat{\mu}_K)^2 K(t_i)}{\sum_{i=1}^n K(t_i)};$$

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n (t_i - \hat{\mu}_u)^2 u(t_i)}{\sum_{i=1}^n u(t_i)}.$$

Заменяя в (6) параметры функций  $K(t)$  и  $u(t)$  их оценками, получим следующие оценки для  $\alpha$  и  $\beta$ :

$$\hat{\alpha} = \frac{(\hat{\mu}_u - \hat{\mu}_K)^2}{(\hat{\sigma}_u^2 - \hat{\sigma}_K^2)}; \quad \hat{\beta} = \frac{\hat{\mu}_u - \hat{\mu}_K}{(\hat{\sigma}_u^2 - \hat{\sigma}_K^2)}.$$

Численный эксперимент показывает, что оценки  $\hat{\alpha}$  и  $\hat{\beta}$  обладают довольно высокой точностью.

**Таблица 1** Среднеквадратическая ошибка исходных данных и оценок

$\alpha$	$\beta$	$\frac{\sigma_{\xi}}{\max K(t)}$	$\frac{\sigma_{\eta}}{\max u(t)}$	$\frac{\hat{\sigma}_{\alpha}}{\alpha}$	$\frac{\hat{\sigma}_{\beta}}{\beta}$
1	1	0,025	0,032	0,080	0,067
1	1	0,050	0,064	0,171	0,143
1	1	0,100	0,128	0,371	0,299
2	1	0,025	0,040	0,033	0,031
2	1	0,050	0,080	0,063	0,061
2	1	0,100	0,155	0,131	0,125
2	2	0,025	0,030	0,097	0,087
2	2	0,050	0,060	0,225	0,202
2	2	0,100	0,119	0,533	0,468
3	1	0,025	0,045	0,009	0,009
3	1	0,050	0,091	0,020	0,021
3	1	0,100	0,182	0,042	0,046
3	1	0,251	0,456	0,106	0,110
3	2	0,025	0,032	0,085	0,078
3	2	0,050	0,064	0,180	0,165
3	2	0,100	0,128	0,396	0,355

В табл. 1 приведены полученные с помощью моделирования среднеквадратические отклонения ошибок (вернее, среднеквадратические отклонения, поделенные, соответственно, на максимальное значение наблюдаемой функции либо на значение оцениваемой величины) исходных данных и оценок.

Предполагается, что

$$K(t_j) = K_0(t_j) + \xi_j; \quad u(t_j) = u_0(t_j) + \eta_j,$$

где  $K_0(t)$  и  $u_0(t)$  — точное ядро и точная правая часть уравнения (3), а  $\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_n$  — случайные ошибки. Случайные величины  $\xi_1, \dots, \xi_n$  так же, как и  $\eta_1, \dots, \eta_n$ , являются независимыми и одинаково распределенными;  $\xi_i$  и  $\eta_i$  имеют нормальные распределения с нулевыми математическими ожиданиями и дисперсиями  $\sigma_{\xi}^2$  и  $\sigma_{\eta}^2$  соответственно. Дисперсии оценок  $\hat{\alpha}$  и  $\hat{\beta}$  обозначены  $\sigma_{\alpha}^2$  и  $\sigma_{\beta}^2$ . Ядро  $K(t)$  является нормальной плотностью распределения

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-(t-4)^2/2}$$

(можно считать, что это распределение сосредоточено на положительной полупрямой).

Заметим, что в обратных задачах, подобных рассматриваемой, погрешность решения обычно имеет порядок  $\sqrt{\epsilon}$ , где  $\epsilon$  — погрешность исходных данных. Для предлагаемых оценок результат несколько лучше.

### 3 Непараметрическое оценивание распределения задержки

Использование параметрической модели удобно с точки зрения оценки распределения задержки, однако может привести к потере важных особенностей оцениваемой плотности. Например, все гамма-плотности являются одновершинными. В то же время наличие двух или более локальных максимумов, с одной стороны, вполне вероятно, а с другой стороны, может отражать существенные особенности изучаемого взаимодействия. В данном разделе будут кратко рассмотрены возможности использования непараметрического подхода.

Поскольку основное уравнение (3) является уравнением свертки 1-го рода, общая схема реализации непараметрического подхода сводится к применению хорошо развитой техники решения такого типа уравнений с использованием преобразований Фурье и регуляризации. Однако наиболее важный этап решения — выбор параметра регуляризации — предлагается производить с использованием специфики ситуации. В рассматриваемом случае параметрическая аппроксимация может привести к потере некоторых локальных особенностей плотности, но позволяет оценить уровень ее гладкости или погрешность измерений.

Обозначим преобразования Фурье (характеристические функции) функций  $K(t)$ ,  $z(t)$  и  $u(t)$  теми же буквами, но с аргументом  $\omega$ . Регуляризованное решение имеет вид

$$z_\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{K(-\omega)u(\omega)e^{-i\omega t}}{|K(\omega)|^2 + \delta M(\omega)} d\omega,$$

где  $M(\omega)$  — четная неотрицательная функция такая, что  $M(0) \geq 0$ ,  $M(\omega) > 0$  при  $\omega \neq 0$  и  $M(\omega) \geq c > 0$  при достаточно больших  $|\omega|$ , удовлетворяющая некоторым условиям регулярности (подробнее см. [7]), а  $\delta$  — параметр регуляризации (положительное число). Можно, например, положить

$$M(\omega) = 1 - e^{-c\omega^2},$$

где параметр  $c$  выбирается из условия, что функции  $\exp(-c\omega^2)$  и  $K(\omega)$  имеют, грубо говоря, приблизительно одинаковый размах.

Если бы было известно отклонение  $\gamma$  (в некоторой метрике  $\rho(\cdot, \cdot)$ ) наблюдаемой правой части уравнения (3)  $u(t)$  от точной  $u_T(t)$ :  $\rho(u, u_T) = \gamma$ , параметр регуляризации можно было бы выбирать из условия

$$\rho(K * z_\delta, u) = \gamma.$$

Поскольку это отклонение неизвестно, выбор параметра регуляризации предлагается осуществлять следующим образом. На первом этапе аппроксимируем  $z(t)$  гамма-плотностью и оцениваем ее параметры  $\alpha$  и  $\beta$ . Для этого используем метод, описанный в предыдущем разделе. Обозначим через  $\hat{z}(t)$  полученную гамма-плотность. Положим

$$\hat{u}(t) = \int_0^{\infty} K(t-s)\hat{z}(s) ds.$$

Ошибку (в  $L^2$ ) измерения правой части уравнения (3) оценим величиной

$$\Delta = \int_0^{\infty} (\hat{u}(t) - u(t))^2 dt.$$

В качестве параметра регуляризации  $\delta$  выбирается решение уравнения

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\delta^2 M^2(\omega) |u(\omega)|^2}{(|K(\omega)| + \delta M(\omega))^2} d\omega = \Delta,$$

поскольку левая часть этого уравнения равна

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} |K(\omega)z_\delta(\omega) - u(\omega)|^2 d\omega = \\ = \int_0^{\infty} (K * z(t) - u(t))^2 dt. \end{aligned}$$

Альтернативным путем использования параметрического старта в непараметрическом оценивании плотности распределения задержки является предварительное, на основе параметрической модели, оценивание какого-либо функционала от плотности, характеризующего уровень сглаженности, например полную вариацию. Параметр регуляризации в этом случае находится из условия равенства указанного функционала у решения уравнения свертки и оцененного из параметрической модели.

## Литература

1. Nowak M. A., May R. M. *Virus dynamics: Mathematical principles of immunology and virology*. — Oxford: Oxford University Press, 2000.
2. Perelson A. S., Neumann A. U., Markowitz M., Leonard J. M., Ho D. D. HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time // *Science*, 1996. Vol. 271. P. 1582–1586.
3. Herz A. V. M., Bonhoeffer S., Anderson R. M., May R. M., Nowak M. A. Viral dynamics in vivo: Limitations on estimates of intracellular delay and virus decay // *Proc. Natl. Acad. Sci. USA*, 1996. Vol. 93(14). P. 7247–7251.
4. Mittler J. E., Bernhard S., Neumann A. U., Perelson A. S. Influence of delayed viral production on viral dynamics in HIV-1 infected patients // *Mathematical Biosciences*, 1998. Vol. 152. P. 143–163.
5. Nelson P. W., Perelson A. S. Mathematical analysis of delay differential equation models of HIV-1 infection // *Mathematical Biosciences*, 1998. Vol. 179. P. 73–94.
6. MacDonald N. *Biological delay systems: Linear stability theory*. — Cambridge: Cambridge University, 1989.
7. Тихонов А. Н., Арсенин В. Я. *Методы решения некорректных задач*. 2-е изд. — М.: Наука, 2002.



# СУЩЕСТВОВАНИЕ СОСТОЯТЕЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ В ДИСКРЕТНЫХ СТАТИСТИЧЕСКИХ ЗАДАЧАХ ПРИ СЛОЖНОЙ НУЛЕВОЙ ГИПОТЕЗЕ\*

А. А. Грушо<sup>1</sup>, Е. Е. Тимонина<sup>2</sup>, В. М. Ченцов<sup>3</sup>

**Аннотация:** В работе рассматривается задача существования состоятельной последовательности критериев при проверке сложной гипотезы против сложной альтернативы в последовательности конечных пространств. В тех случаях, когда последовательность пространств представляет собой декартово произведение конечного множества и вероятностные меры на этих пространствах согласованы, удается найти достаточные условия существования состоятельной последовательности критериев в терминах топологических свойств множеств, покрывающих носитель доминирующей меры для класса предельных мер из нулевой гипотезы. При дополнительных условиях удается отказаться от требования доминирования класса предельных мер из нулевой гипотезы и равномерной ограниченности плотностей.

**Ключевые слова:** состоятельная последовательность критериев; сложная гипотеза против сложной альтернативы; конечные пространства; вероятностные меры; достаточные условия

По мере проникновения компьютерных и компьютеризированных систем во все сферы человеческой деятельности актуальность проблем компьютерной безопасности и защиты компьютерных систем становится все более актуальной. Очень важно понимать, что такое защищенность компьютерной системы и на чем должна быть основана уверенность, что требуемая защищенность имеется. Исторический опыт показывает, что формализацию как определения защищенности, так и обоснования защищенности следует искать математическими методами и в рамках математических моделей.

Авторами ранее [1, 2] построены примеры доказуемо защищенных компьютерных систем. Ими разработаны методы контроля информационных потоков в компьютерных системах. Проведены исследования свойств скрытых от контроля информационных потоков в распределенных компьютерных системах. Доказана возможность скрытой передачи информации через такие сильные средства защиты, как межсетевые экраны и криптография. Эти результаты нашли практическое подтверждение. С помощью доказательства утверждений о

несуществовании состоятельных последовательностей статистических критериев [3, 4] для выявления признаков искомых каналов удается доказать «невидимость» некоторых информационных потоков. Обратные утверждения, напротив, свидетельствуют о том, что при определенных условиях «невидимых» потоков не существует. Такие обратные утверждения доказываются в данной статье.

Пусть задано конечное множество  $X = \{x_1, \dots, x_m\}$ . Через  $X^\infty$  обозначим множество бесконечных последовательностей, где каждый элемент последовательности принадлежит  $X$ . Пусть  $\mathcal{A}$  —  $\sigma$ -алгебра, порожденная цилиндрическими подмножествами  $X^\infty$ . Рассмотрим два семейства вероятностных мер на измеримом пространстве  $\{X^\infty, \mathcal{A}\}$ :  $\{P_\lambda, \lambda \in \Lambda\}$  и  $\{P_\theta, \theta \in \Theta\}$ .

Обозначим через  $P_{\lambda,n}$  и  $P_{\theta,n}$  проекции введенных вероятностных мер на первые  $n$  координат последовательностей из  $X^\infty$ . Для каждого  $n$  рассмотрим задачу проверки сложной статистической гипотезы  $H_{0,n} : \{P_{\lambda,n}, \lambda \in \Lambda\}$  против сложной альтернативы  $H_{1,n} : \{P_{\theta,n}, \theta \in \Theta\}$ . Для каждого  $n$  критерий уровня значимости  $\alpha$  описывается кри-

\* Работа выполнена при поддержке РФФИ, проект 07-01-00484.

<sup>1</sup> Российский государственный гуманитарный университет, Институт информационных наук и технологий безопасности, Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, grusho@yandex.ru

<sup>2</sup> Российский государственный гуманитарный университет, Институт информационных наук и технологий безопасности, eltimon@yandex.ru

<sup>3</sup> Институт проблем информатики Российской академии наук, ipiran@ipiran.ru

тическим множеством  $S_n$ ,  $P_{\lambda,n}(S_n) \leq \alpha$ ,  $\lambda \in \Lambda$ , и мощностью критерия  $W_n(\theta) = P_{\theta,n}(S_n)$ .

Последовательность статистических критериев с критическими множествами  $S_n$  называется состоятельной [5], если для каждого  $\alpha \in (0; 1]$  мощность критерия  $W_n(\theta) \rightarrow 1$  для каждого  $\theta \in \Theta$ . При этом можно считать, что  $\alpha \rightarrow 0$  при  $n \rightarrow \infty$ .

Последовательность критериев с критическими множествами  $S_n$  называется слабо состоятельной, если  $P_{\lambda,n}(S_n) \rightarrow 0$  для каждого  $\lambda \in \Lambda$  и  $P_{\theta,n}(S_n) \rightarrow 1$  для каждого  $\theta \in \Theta$ .

Ясно, что состоятельная последовательность критериев является слабо состоятельной, но не наоборот. Вопрос о существовании состоятельной последовательности критериев для простой нулевой гипотезы изучался в [3]. Вопрос о несуществовании состоятельной последовательности критериев исследовался в [2, 4]. В данной работе предпринята попытка обобщить некоторые результаты этих работ на случай сложной нулевой гипотезы.

Будем предполагать, что на  $\{X^\infty, \mathcal{A}\}$  определена еще одна вероятностная мера  $P_0$ , которая доминирует семейство распределений  $\{P_\lambda, \lambda \in \Lambda\}$ . Тогда для каждого  $\lambda \in \Lambda$  существует плотность  $p_\lambda$  такая, что  $\forall B \in \mathcal{A}$

$$P_\lambda(B) = \int_B p_\lambda dP_0.$$

Будем считать, что существует  $C > 0$  такое, что для любого  $\lambda$  существует вариант плотности  $p_\lambda$  такой, что  $p_\lambda \leq C$ . Отсюда следует, что если  $A \in \mathcal{A}$  и  $P_0(A) = 1$ , то  $P_\lambda(A) = 1$ . Действительно,

$$P_\lambda(A) = \int_A P_\lambda dP_0 = \int_{X^\infty} p_\lambda dP_0 - \int_{\bar{A}} p_\lambda dP_0 = 1,$$

так как

$$\int_{\bar{A}} p_\lambda dP_0 \leq C \int_{\bar{A}} dP_0 = 0.$$

Для каждого  $\lambda \in \Lambda$ ,  $D_n \subseteq X^n$

$$\begin{aligned} P_{\lambda,n}(D_n) &= P_\lambda(D_n \times X^\infty) = \\ &= \int_{D_n \times X^\infty} p_\lambda dP_0 \leq CP_0(D_n \times X^\infty) = CP_{0,n}(D_n). \end{aligned}$$

Будем также рассматривать дискретную топологию в  $X$  и Тихоновское произведение в  $X^\infty$  [6].

**Теорема 1.** Пусть существует замкнутое в Тихоновском произведении множество  $A$  такое, что  $P_0(A) = 1$  и  $\forall \theta \in \Theta$  существует  $A(\theta) \in \mathcal{A}$  такое, что  $A \cap A(\theta) = \phi$  и  $P_\theta(A(\theta)) = 1$ . Тогда существует состоятельная последовательность критериев для проверки  $H_{0,n}$  против  $H_{1,n}$ .

**Доказательство.** В рассматриваемых условиях по теореме 1 из работы [3] существует состоятельная последовательность критериев для проверки  $H'_{0,n} : P_{0,n}$  против  $H_{1,n}$  с критическими множествами  $S_n$  такая, что  $P_{0,n}(S_n) \rightarrow 0$ ,  $n \rightarrow \infty$ , и  $\forall \theta \in \Theta$   $P_{\theta,n}(S_n) \rightarrow 1$ . Для каждого  $\alpha \in (0; 1]$  существует  $N$  такое, что для всех  $n \geq N$  имеем  $P_{0,n}(S_n) \leq \alpha$ . Тогда для всех  $\lambda \in \Lambda$

$$P_{\lambda,n}(S_n) \leq C\alpha.$$

Поэтому последовательность критериев с критическими множествами  $S_n$ ,  $n = 1, 2, \dots$  также является состоятельной для проверки  $H_{0,n}$  против  $H_{1,n}$ . Теорема доказана.

**Теорема 2.** Пусть существует открытое в Тихоновском произведении множество  $A$  такое, что  $P_0(A) = 1$  и  $\forall \theta \in \Theta$  существует  $A(\theta) \in \mathcal{A}$ , что  $A \cap A(\theta) = \phi$  и  $P_\theta(A(\theta)) = 1$ . Тогда существует состоятельная последовательность критериев для проверки  $H_{0,n}$  против  $H_{1,n}$ .

**Доказательство.** По теореме 2 из работы [3] в условиях рассматриваемой теоремы существует состоятельная последовательность критериев для проверки простой гипотезы  $H'_{0,n} : P_{0,n}$  против сложной альтернативы  $H_{1,n}$  с критическими множествами  $S_n$ . Далее, так же как в теореме 1, доказывается, что построенная последовательность критериев будет состоятельной для проверки  $H_{0,n}$  против  $H_{1,n}$ . Теорема доказана.

Откажемся теперь от требования доминирования мер  $P_\lambda$ ,  $\lambda \in \Lambda$ . Однако предположим, что существует замкнутое в Тихоновском произведении множество  $A$  такое, что для  $\forall \lambda \in \Lambda$  существует  $A_\lambda \in \mathcal{A}$  такое, что  $A_\lambda \subseteq A$  и  $P_\lambda(A_\lambda) = 1$ . Тогда можно доказать следующую теорему.

**Теорема 3.** Если существует замкнутое множество  $A_1$  такое, что  $A \cap A_1 = \phi$  и  $\forall \theta \in \Theta$   $P_\theta(A_1) = 1$ , то существует состоятельная последовательность критериев для проверки  $H_{0,n}$  против  $H_{1,n}$ .

**Доказательство.** Из замкнутости множеств  $A$  и  $A_1$  следует, что существуют [1] невозрастающие последовательности цилиндрических множеств  $I_n$ ,  $n = 1, 2, \dots$ , и  $I'_n$ ,  $n = 1, 2, \dots$  такие, что

$$\begin{aligned} I_n &= S_n \times X^\infty, \quad S_n \subseteq X^\infty \\ A &= \bigcap_{n=1}^{\infty} I'_n, \quad A_1 = \bigcap_{n=1}^{\infty} I_n. \end{aligned}$$

По условию  $A \cap A_1 = \phi$ . Тогда

$$\bigcap_{n=1}^{\infty} (I'_n \cap I_n) = \phi.$$

Любое пересечение цилиндрических множеств  $I'_n \cap I_n = I''_n$  — цилиндрическое множество или пустое множество. Последовательность  $I''_n$ ,  $n = 1, 2, \dots$  цилиндрических множеств монотонно не возрастает и  $\bigcap_{n=1}^{\infty} I''_n = \phi$ . Топологическое пространство  $X$  — компакт. Тогда из компактности Тихоновского произведения  $X^\infty$  следует [6], что существует  $N$  такое, что  $\bigcap_{n=1}^N I''_n = \phi$ . Из условия невозрастания  $I''_n$  следует, что  $I''_N = \phi$ .

Рассмотрим последовательность статистических критериев с критическими множествами  $S_n$ ,  $n = 1, 2, \dots$ . Из определения  $A$  следует, что  $A \subseteq I'_n$ . Тогда  $A \cap I_n = \phi$ , т. е.  $I_N \subseteq X^\infty \setminus A$ . Отсюда следует, что для каждого  $\lambda \in \Lambda$  при  $A_\lambda \in \mathcal{A}$  и  $P_\lambda(A_\lambda) = 1$

$$0 = P_\lambda(X^\infty \setminus A_\lambda) \geq P_\lambda(X^\infty \setminus A) \geq P_\lambda(I_N) = P_\lambda(S_N \times X^\infty) = P_{\lambda,N}(S_N).$$

Из условия теоремы так же, как в теореме 3, работы [3]

$$W_N(\theta) = P_{\theta,N}(S_N) = P_\theta(I_N) \geq P_\theta(A_1) = 1.$$

Теорема доказана.

Проверить замкнутость или открытость множеств в пространстве  $X^\infty$  можно конструктивно, представив соответствующее множество в виде монотонно невозрастающей или монотонно убывающей последовательности цилиндрических множеств. Сложнее проверяется свойство доминирования семейства вероятностных мер и существование равномерно ограниченных плотностей. В условиях теоремы 3 удается избежать такой проверки.

В работе [4] получено следующее достаточное условие несуществования состоятельной последовательности критериев в нашей модели при  $|\Lambda| = 1$ . Если для каждого  $A \in \mathcal{A}$  такого, что  $P_\lambda(A) = 1$ , существует  $\theta \in \Theta$  такое, что  $P_\theta > 0$ , то состоятельной последовательности критериев не существует.

Отметим, что в случае  $|\Lambda| = 1$  это условие также является необходимым. В самом деле, если состоятельной последовательности не существует, но существует  $A \in \mathcal{A}$  такое, что  $P_\lambda(A) = 1$ , и для каждого  $\theta \in \Theta$   $P_\theta(A) = 0$ , то рассмотрим критерий с критическим множеством  $X^\infty \setminus A$ . Этот критерий удовлетворяет следующим условиям:

$$P_\lambda(X^\infty \setminus A) = 0$$

и для каждого  $\theta \in \Theta$

$$P_\theta(X^\infty \setminus A) = 1.$$

Тогда последовательность критериев с одним и тем же критическим множеством  $X^\infty \setminus A$  является состоятельной, что противоречит предположению.

Для произвольного  $\Lambda$  справедлива следующая теорема.

**Теорема 4.** Если существует  $\lambda \in \Lambda$  такое, что для каждого множества  $A \in \mathcal{A}$   $P_\lambda(A) = 1$  и существует  $\theta \in \Theta$  такое, что  $P_\theta(A) > 0$ , то не существует состоятельной последовательности критериев проверки сложных гипотез  $H_{0,n} : \{P_{\lambda,n}, \lambda \in \Lambda\}$  против сложных альтернатив  $H_{1,n} : \{P_{\theta,n}, \theta \in \Theta\}$ .

**Доказательство.** Из условий теоремы и приведенных выше достаточных условий следует, что не существует состоятельной последовательности критериев для проверки простых гипотез  $H'_{0,n} : P_{\lambda,n}$  против сложных альтернатив  $H_{1,n} : \{P_{\theta,n}, \theta \in \Theta\}$ . Тогда, как это следует из определения, не существует слабо состоятельной последовательности критериев для проверки  $H_{0,n}$  против  $H_{1,n}$ . Это, в свою очередь, означает, что не существует состоятельной последовательности критериев для проверки  $H_{0,n}$  против  $H_{1,n}$ . Теорема доказана.

## Литература

1. *Grusho A., Kniazev A., Timonina E.* Detection of illegal information flow // Proceedings of 3rd International Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security, MMM-ACNS 2005. — St. Petersburg: Springer, 2005. LNCS 3685. P. 235–244.
2. *Grusho A., Grebnev N., Timonina E.* Covert channel invisibility theorem // Proceedings of 4th International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security, MMM-ACNS 2007. — St. Petersburg: Springer, 2007. P. 187–196.
3. *Грушо А. А., Тимонина Е. Е.* Некоторые связи между дискретными статистическими задачами и свойствами вероятностных мер на топологических пространствах // Дискретная математика, 2006. Т. 18. № 4. С. 128–135.
4. *Грушо А., Грушо Н., Тимонина Е.* Теоремы о несуществовании состоятельных последовательностей критериев в некоторых дискретных задачах // Дискретная математика, 2008 (в печати). Т. 20. № 2.
5. *Леман Е.* Проверка статистических гипотез. — М.: Наука, 1964.
6. *Прохоров Ю. В., Розанов Ю. А.* Теория вероятностей. — М.: Наука, 1973.

# СТОХАСТИЧЕСКИЕ РАЗЛОЖЕНИЯ НЕСМЕЩЕННЫХ ОЦЕНОК В СЛУЧАЕ ОДНОПАРАМЕТРИЧЕСКОГО ЭКСПОНЕНЦИАЛЬНОГО СЕМЕЙСТВА

В. В. Чичагов<sup>1</sup>

**Аннотация:** Получены асимптотические и стохастические разложения для всех несмещенных оценок, которые могут быть построены по повторной выборке, элементами которой являются независимые случайные величины, имеющие одно и то же распределение из однопараметрического экспоненциального семейства. Проведено сравнение разложений несмещенных оценок и оценок максимального правдоподобия. Найдено стохастическое разложение несмещенной оценки дисперсии несмещенной оценки.

**Ключевые слова:** несмещенная оценка; экспоненциальное семейство; стохастическое разложение; несмещенная оценка дисперсии

## 1 Введение

Точечные несмещенные оценки играют важную роль в современных научно-технических исследованиях. Техника получения несмещенных оценок хорошо разработана, решено значительное число теоретических и прикладных задач (см., например, [1, 2]). Продолжают появляться новые работы такой же направленности (например, [3–7]). Однако известно значительно меньше работ, в которых бы изучалось предельное поведение несмещенных оценок при большом объеме выборки и осуществлялось сравнение несмещенных оценок с оценками другого вида, в первую очередь с оценками максимального правдоподобия. Возможной причиной сложившейся ситуации можно считать появление работ [8, 9], в которых при весьма общих предположениях в случае экспоненциального семейства распределений была установлена асимптотическая эквивалентность несмещенных оценок и оценок максимального правдоподобия с точностью до слагаемого  $O_P(n^{-1})$ , где  $n$  — объем повторной выборки. В появившейся сравнительно недавно работе [10] получено более общее решение этой проблемы с использованием счетной системы ассоциированных с экспоненциальным семейством ортогональных полиномов, которые были предложены в работе [11]. Другой подход, непосредственно ориентированный на изучение предельного поведения несмещенных оценок, предложен в работах автора [12, 13]. Он основывается на применении локальных предельных теорем и разложения Эджворта [14–17] к исследованию плотно-

сти условного распределения, определяющей вид несмещенной оценки плотности [18–22]. В данной работе эта идея получила свое логическое завершение в случае однопараметрического экспоненциального семейства, которому принадлежит распределение наблюдаемой случайной величины. Применение этого подхода, в отличие от упомянутых выше методов, позволяет получать для несмещенных оценок асимптотические разложения высокого порядка. Полученные в работах [23–25] результаты, определяющие асимптотику условного распределения, хотя и близки в идейном плане, но оказались непригодными для решения поставленной проблемы, поскольку получены для случайных векторов, обе компоненты которых являются суммами случайных величин с большим числом слагаемых. В рассматриваемом случае первая компонента состоит из конечного числа слагаемых. Еще одна задача, которая решается в данной работе, состоит в исследовании предельного поведения несмещенной оценки дисперсии несмещенной оценки. Проблемы, связанные с нахождением несмещенных оценок дисперсии и их применением, рассматривались, например, в работах [26–28].

## 2 Модель наблюдений и основные предположения

Имеется  $X_1, \dots, X_n$  — независимая повторная выборка, элементы которой имеют то же распре-

<sup>1</sup>Пермский государственный университет, chvv50@mail.ru

деление, что и наблюдаемая случайная величина  $\xi$ , удовлетворяющая следующим предположениям.

(A<sub>1</sub>). Распределение вероятностей случайной величины  $\xi$  принадлежит однопараметрическому экспоненциальному семейству, определяемому выражением

$$f(x; \theta) = h(x) \exp\{\theta T(x) + v(\theta)\}, \quad x \in G \subset \mathbf{R}. \quad (1)$$

Здесь  $f(x; \theta)$  — плотность распределения случайной величины  $\xi$  относительно меры  $\mu(x)$ , являющейся либо мерой Лебега, если  $\xi$  имеет абсолютно непрерывное распределение, либо считающей мерой, когда  $\xi$  имеет решетчатое распределение;  $G$  — носитель распределения,  $\theta \in \Theta \subset \mathbf{R}$  — неизвестный канонический параметр распределения,  $h(x)$ ,  $T(x)$ ,  $v(\theta)$  — известные борелевские функции.

(A<sub>2</sub>). Если  $\mu(x)$  — мера Лебега, то для каждого  $\theta \in \Theta$  существует  $n_0 \in \mathbf{N}$  такое, что случайная величина

$$Z_n = \frac{S_n - na}{b\sqrt{n}},$$

где

$$S_n = \sum_{i=1}^n T(X_i);$$

$$a = \mathbf{E}[T(\xi)];$$

$$b^2 = \mathbf{V}[T(\xi)],$$

имеет непрерывную ограниченную плотность  $f_{Z_n}(x; \theta)$  для  $n \geq n_0$ . Если  $\mu(x)$  — считающая мера, то носитель  $G$  не содержится ни в какой подрешетке решетки  $\mathbf{Z}$ .

(A<sub>3</sub>). Носитель распределения  $G$  не зависит от параметра  $\theta$ .

(A<sub>4</sub>). Параметрическая область  $\Theta$  содержит некоторый интервал, принадлежащий  $\mathbf{R}$ .

### 3 Интегральное представление несмещенных оценок и несмещенно оцениваемых параметрических функций

Основной целью данной работы является исследование асимптотического поведения несмещенной оценки  $h_n(S_n)$  заданной параметрической функции  $g(\theta)$ . Оценка строится по независимой повторной выборке  $X_1, \dots, X_n$  объема  $n$ .

Условие несмещенности для этой оценки может быть записано в виде уравнения

$$g(\theta) = \int_{\mathbf{R}} h_n(t) f_n(t; \theta) \mu(dt), \quad n \geq l, \quad (2)$$

где  $l$  — минимальный объем выборки, для которого выполняется (2) и который является конечным числом, не зависящим от  $n$ ;  $f_n(t; \theta)$  — плотность распределения статистики  $S_n$ , которая является достаточной для параметра  $\theta$  в силу условий (A<sub>1</sub>) и (A<sub>3</sub>). Это означает, что любую параметрическую функцию  $g(\theta)$ , для которой существует несмещенная оценка, являющаяся функцией достаточной статистики  $S_n$ , можно представить в интегральной форме

$$g(\theta) = \int_{\mathbf{R}} h_l(t) f_l(t; \theta) \mu(dt). \quad (3)$$

Но в таком случае имеет место аналогичное представление и для несмещенной оценки параметрической функции (3):

$$h_n(S_n) = \int_{\mathbf{R}} h_l(t) \hat{f}_l(t|S_n) \mu(dt), \quad (4)$$

где  $\hat{f}_l(t|S_n)$  — несмещенная оценка плотности  $f_l(t; \theta)$ . Хорошо известно [1, 18, 20], что несмещенная оценка плотности определяется выражением

$$\hat{f}_l(t|S_n) = \frac{f_l(t; \theta_0) f_{n-l}(S_n - t; \theta_0)}{f_n(S_n; \theta_0)}, \quad (5)$$

где  $\theta_0$  — произвольное значение  $\theta$  из области  $\Theta$  (будем полагать  $\theta_0 = \theta$ ). В силу условия полноты (A<sub>4</sub>) экспоненциального семейства (1) представление (4) единственно. Из формул (3)–(5) следует, что ошибка несмещенной оценки  $h_n(S_n)$  определяется интегральным выражением

$$h_n(S_n) - g(\theta) = \int_{\mathbf{R}} h_l(t) f_l(t; \theta) \left[ \frac{f_{n-l}(S_n - t; \theta)}{f_n(S_n; \theta)} - 1 \right] \mu(dt), \quad (6)$$

поэтому предельное при  $n \rightarrow \infty$  поведение несмещенной оценки  $h_n(S_n)$  определяется отношением

$$U_1(S_n, t; \theta) = \frac{f_{n-l}(S_n - t; \theta)}{f_n(S_n; \theta)}. \quad (7)$$

Известно [26], что  $\hat{\mathbf{V}}[h_n(S_n)]$  — несмещенная оценка дисперсии несмещенной оценки  $\mathbf{V}[h_n(S_n)]$  может быть найдена по формуле

$$\hat{\mathbf{V}}[h_n(S_n)] = h_n^2(S_n) - \hat{g}^2(\theta|S_n), \quad (8)$$

где  $\widehat{g}^2(\theta|S_n)$  — несмещенная оценка параметрической функции  $g^2(\theta)$ . Воспользовавшись (8), нетрудно показать, что несмещенная оценка  $\widehat{V}[h_n(S_n)]$  может быть записана в интегральной форме:

$$\begin{aligned} \widehat{V}[h_n(S_n)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \widehat{f}_l(t_1|S_n) \widehat{f}_l(t_2|S_n) - \right. \\ &\quad \left. - \widehat{f}_{l,l}(t_1, t_2|S_n) \right] \prod_{j=1}^2 h_l(t_j) \mu(dt_j) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \frac{f_{n-l}(S_n - t_1; \theta)}{f_n(S_n; \theta)} \frac{f_{n-l}(S_n - t_2; \theta)}{f_n(S_n; \theta)} - \right. \\ &\quad \left. - \frac{f_{n-2l}(S_n - t_1 - t_2; \theta)}{f_n(S_n; \theta)} \right] \times \\ &\quad \times \prod_{j=1}^2 h_l(t_j) f_l(t_j; \theta) \mu(dt_j), \quad (9) \end{aligned}$$

где  $\widehat{f}_{l,l}(t_1, t_2|S_n)$  — несмещенная оценка плотности  $\prod_{j=1}^2 f_l(t_j; \theta)$ :

$$\begin{aligned} \widehat{f}_{l,l}(t_1, t_2|S_n) &= \\ &= \frac{f_l(t_1; \theta) f_l(t_2; \theta) f_{n-2l}(S_n - t_1 - t_2; \theta)}{f_n(S_n; \theta)}. \end{aligned}$$

Поэтому предельное при  $n \rightarrow \infty$  поведение несмещенной оценки  $\widehat{V}[h_n(S_n)]$  определяется выражением

$$\begin{aligned} U_2(S_n, t_1, t_2; \theta) &= \\ &= \frac{f_{n-l}(S_n - t_1; \theta)}{f_n(S_n; \theta)} \frac{f_{n-l}(S_n - t_2; \theta)}{f_n(S_n; \theta)} - \\ &\quad - \frac{f_{n-2l}(S_n - t_1 - t_2; \theta)}{f_n(S_n; \theta)}. \quad (10) \end{aligned}$$

#### 4 Стохастическое разложение несмещенной оценки плотности и интегрального функционала

Сначала изучается асимптотическое поведение компоненты (7), определяющей ошибку несмещенной оценки (6). Затем определяется стохастическое разложение несмещенной оценки. В заключительной части раздела проводится сравнение стохастических разложений несмещенной оценки и оценки максимального правдоподобия.

**Теорема 1.** Пусть выполнены предположения  $(A_1)$ – $(A_4)$ ,  $\Delta_l(t) = (t - la)/b$ . Тогда при  $s = na +$

$+ zb\sqrt{n}$ ,  $n \rightarrow \infty$ , равномерно по  $z$  при  $|z| \leq z_0$  для любого фиксированного  $z_0$  справедливо разложение

$$\begin{aligned} \frac{f_{n-l}(s-t; \theta)}{f_n(s; \theta)} &= 1 + \frac{z\Delta_l(t)}{n^{1/2}} - \frac{H_2(z)}{2n} [\rho_3 \Delta_l(t) + \\ &\quad + l - \Delta_l^2(t)] + \\ &\quad + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j(z) \Delta_l^j(t) + P_4(\Delta_l(t)) \cdot \mathbf{O}(n^{-2}), \end{aligned}$$

если  $\Delta_l(t) = \mathbf{o}(n^{1/2})$ , (11)

$$\frac{f_{n-l}(s-t; \theta)}{f_n(s; \theta)} = \mathbf{O}(1) \text{ при любом } t \in \mathbf{R}, \quad (12)$$

где

$$\begin{aligned} c_0(z) &= \frac{l\rho_3}{6} [2H_3(z) + 3z]; \\ c_1(z) &= H_3(z) \left[ \frac{\rho_3^2 - l}{2} - \frac{\rho_4}{6} \right] + \frac{\rho_3^2 z}{2}; \\ c_2(z) &= -\frac{\rho_3}{2} [z + H_3(z)]; \\ c_3(z) &= \frac{H_3(z)}{6}; \end{aligned}$$

$H_2(z) = z^2 - 1$ ;  $H_3(z) = z^3 - 3z$ ;  $\rho_3$  и  $\rho_4$  — нормированные кумулянты (соответственно коэффициенты асимметрии и эксцесса) распределения случайной величины  $\xi$ ;  $P_4(x)$  — некоторый полином степени 4.

Доказательство. Сначала рассмотрим случай абсолютно непрерывного распределения  $\xi$ . Воспользуемся тождеством

$$\frac{f_{n-l}(s-t; \theta)}{f_n(s; \theta)} = \sqrt{\frac{n}{n-l}} \cdot \frac{f_{Z_{n-l}}(u; \theta)}{f_{Z_n}(z; \theta)}, \quad (13)$$

где

$$u = \sqrt{\frac{n}{n-l}} \left( z - \frac{\Delta_l(t)}{\sqrt{n}} \right),$$

а также асимптотическим разложением плотности  $f_{Z_n}(z; \theta)$  с использованием приближения Эджворта, которое в случае принадлежности  $f(x; \theta)$  экспоненциальному семейству (см., например, (A.3) в [16]) имеет вид

$$\begin{aligned} f_{Z_n}(z; \theta) &= \phi(z) \left[ 1 + \sum_{j=1}^3 \frac{q_j(z)}{n^{j/2}} \right] + \mathbf{O}(n^{-2}) \equiv \\ &\equiv w_{3,n}(z) + \mathbf{O}(n^{-2}), \end{aligned}$$

где

$$\begin{aligned} \phi(z) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \\ q_1(z) &= \frac{\rho_3 H_3(z)}{6}, \\ q_2(z) &= \frac{\rho_4 H_4(z)}{24} + \frac{\rho_3^2 H_6(z)}{72}, \end{aligned}$$

$q_3(z)$  — некоторый полином степени 9,  $\{H_j(z)\}$  — полиномы Эрмита.

Учитывая, что

$$\begin{aligned} \sqrt{\frac{n}{n-l}} &= 1 + \frac{l}{2n} + \mathbf{O}(n^{-2}), \\ \frac{1}{\sqrt{n-l}} &= \frac{1}{\sqrt{n}} + \frac{l}{2n^{3/2}} + \mathbf{O}(n^{-2}), \\ u-z &= \frac{lz}{2n} - \left(1 + \frac{l}{2n}\right) \frac{\Delta_l(t)}{\sqrt{n}} + \\ &\quad + \mathbf{O}(n^{-2}) [1 + \Delta_l(t)], \end{aligned}$$

и используя свойство полиномов Эрмита  $H_r(z)\phi(z) = (-1)^r \phi^{(r)}(z)$ , осуществим разложение (7) при  $S_n = s$  и  $\Delta_l(t) = \mathbf{o}(n^{1/2})$  следующим образом:

$$\begin{aligned} \sqrt{\frac{n}{n-l}} \cdot \frac{f_{Z_{n-l}}(u; \theta)}{f_{Z_n}(z; \theta)} &= \sqrt{\frac{n}{n-l}} \times \\ &\times \frac{w_{3,n}(u) + \phi(u) l \rho_3 / (12n^{3/2}) H_3(u) + \mathbf{O}(n^{-2})}{w_{3,n}(z) + \mathbf{O}(n^{-2})} = \\ &= \left(1 + \frac{l}{2n}\right) \left[1 + \frac{1}{w_{3,n}(z)} \sum_{j=1}^3 w_{3,n}^{(j)}(z) \frac{(u-z)^j}{j!} + \right. \\ &\quad \left. + \frac{l \rho_3 H_3(z)}{12n^{3/2}} + \frac{P_4(\Delta_l(t))}{n^2}\right] + \mathbf{O}(n^{-2}). \quad (14) \end{aligned}$$

Путем элементарных преобразований получим выражения

$$(u-z)^2 = \frac{\Delta_l^2(t)}{n} - \frac{lz\Delta_l(t)}{n^{3/2}} + \mathbf{O}(n^{-2}) [1 + \Delta_l^2(t)];$$

$$(u-z)^3 = -\frac{\Delta_l^3(t)}{n^{3/2}} + \mathbf{O}(n^{-2}) [\Delta_l^3(t)];$$

$$\frac{w_{3,n}^{(1)}(z)}{w_{3,n}(z)} = \sum_{j=0}^2 \frac{c_{1j}(z)}{n^{j/2}} + \mathbf{O}(n^{-3/2});$$

$$\frac{w_{3,n}^{(2)}(z)}{w_{3,n}(z)} = \sum_{j=0}^1 \frac{c_{2j}(z)}{n^{j/2}} + \mathbf{O}(n^{-1});$$

$$\frac{w_{3,n}^{(3)}(z)}{w_{3,n}(z)} = H_3(z) + \mathbf{O}(n^{-1/2}),$$

где

$$\begin{aligned} c_{10}(z) &= z; \quad c_{11}(z) = -\frac{\rho_3 H_2(z)}{2}; \\ c_{12}(z) &= -\frac{\rho_4 H_3(z)}{6} + \frac{\rho_3^2}{2} [z + H_3(z)]; \\ c_{20}(z) &= H_2(z); \quad c_{21}(z) = -\rho_3 [z + H_3(z)]. \end{aligned}$$

Подставляя найденные выражения в (14), получим (11).

В соответствии с локальной предельной теоремой, примененной к последовательности случайных величин  $\{Z_{n-l}\}$ , справедливость (12) следует из равномерной ограниченности плотности  $f_{Z_{n-l}}(u; \theta)$  при достаточно больших  $n$ .

Случай решетчатого распределения  $\xi$  рассматривается аналогично случаю абсолютно непрерывного распределения. Основное отличие этого случая от предыдущего состоит в другой форме записи левой части (13):

$$\frac{f_{n-l}(s-t; \theta)}{f_n(s; \theta)} = \sqrt{\frac{n}{n-l}} \cdot \frac{b\sqrt{n-l} \cdot f_{n-l}(s-t; \theta)}{b\sqrt{n} \cdot f_n(s; \theta)},$$

которая позволяет воспользоваться приближением Эджворта для решетчатого распределения (см., например, теорему 6.10 из [17]).

Теорема 1 доказана.

Используя асимптотическое разложение (11), получим аналогичное разложение для значения ошибки несмещенной оценки (6).

**Теорема 2.** Пусть выполнены предположения  $(A_1)$ – $(A_4)$ . Пусть также справедливо условие:

$(A_5)$ . Существует  $0 < \nu < 0,5$  такое, что для  $j = \overline{0, 3}$  при  $n \rightarrow \infty$

$$\int_{B_{n,\nu}} h_l(t) \Delta_l^j(t) f_l(t; \theta) \mu(dt) = \mathbf{O}(n^{-2}),$$

$$B_{n,\nu} = \{t \in \mathbf{R} : |\Delta_l(t)| > n^\nu\};$$

$$D_j = \int_{-\infty}^{\infty} h_l(t) \Delta_l^j(t) f_l(t; \theta) \mu(dt) < \infty,$$

$$j = \overline{0, 4}. \quad (15)$$

Тогда при  $s = na + z b\sqrt{n}$ ,  $n \rightarrow \infty$ , равномерно по  $z$  при  $|z| \leq z_0$  для любого фиксированного  $z_0$  справедливо разложение

$$\begin{aligned} h_n(s) - g(\theta) &= \frac{zD_1}{\sqrt{n}} - \frac{H_2(z)}{2n} [lD_0 - D_2 + \rho_3 D_1] + \\ &\quad + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j(z) D_j + \mathbf{O}(n^{-2}). \end{aligned}$$

Доказательство. Положим

$$P_3(x) = \frac{zx}{\sqrt{n}} - \frac{H_2(z)}{2n} [\rho_3 x + l - x^2] + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j(z) x^3.$$

Используя теорему 1 и условие  $(A_5)$ , а также формулы (6), (11) и (15), получим при  $S_n = s$ ,  $n \rightarrow \infty$ , соотношение

$$\begin{aligned} h_n(s) - g(\theta) - \left\{ \frac{zD_1}{\sqrt{n}} - \frac{H_2(z)}{2n} [lD_0 - D_2 + \rho_3 D_1] + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j(z) D_j \right\} = \\ = \int_{\overline{B}_{n,\nu}} h_l(t) P_4(\Delta_l(t)) f_l(t; \theta) \mu(dt) \cdot \mathbf{O}(n^{-2}) + \\ + \int_{B_{n,\nu}} h_l(t) [\mathbf{O}(1) - P_3(\Delta_l(t))] f_l(t; \theta) \mu(dt) = \\ = \mathbf{O}(n^{-2}), \end{aligned}$$

которое справедливо равномерно относительно  $z$  при  $|z| \leq z_0$ .

Теорема 2 доказана.

Теперь сформулируем утверждение, которое позволяет обосновывать справедливость стохастического разложения, получаемого из асимптотического разложения методом подстановки.

**Лемма 1.** Пусть выполнены условия  $(A_1)$ – $(A_2)$ , а функция  $G(z; n)$  такова, что при  $n \rightarrow \infty$  равномерно относительно  $z$  при  $|z| \leq z_0$  для любого фиксированного  $z_0$  справедливо асимптотическое разложение

$$G(z; n) = \sum_{j=0}^r \frac{G_j(z)}{n^{j/2}} + \mathbf{O}(n^{-(r+1)/2}),$$

где  $\{G_j(z), j = \overline{0, r}\}$  — некоторые полиномы относительно  $z$ . Тогда при  $n \rightarrow \infty$  верно стохастическое разложение

$$G(Z_n; n) = \sum_{j=0}^r \frac{G_j(Z_n)}{n^{j/2}} + \mathbf{O}_P(n^{-(r+1)/2}),$$

где запись  $U_n = \mathbf{O}_P(n^{-(r+1)/2})$  означает, что выражение  $n^{(r+1)/2} U_n$  ограничено по вероятности.

Справедливость леммы 1 следует из центральной предельной теоремы для независимых одина-

ково распределенных случайных величин и неравенства

$$\begin{aligned} \mathbf{P} \left( \left| G(Z_n; n) - \sum_{j=0}^r \frac{G_j(Z_n)}{n^{j/2}} \right| < \frac{k_\varepsilon}{n^{(r+1)/2}} \right) \geq \\ \geq \mathbf{P} \left( |Z_n| \leq z_0; \left| \mathbf{O}(n^{-(r+1)/2}) \right| < \frac{k_\varepsilon}{n^{(r+1)/2}} \right) = \\ = \mathbf{P}(|Z_n| \leq z_0), \end{aligned}$$

которое верно при достаточно большом значении  $k_\varepsilon$ .

Из леммы 1 и теоремы 2 немедленно следует справедливость следующего утверждения.

**Следствие 1.** В условиях теоремы 2 для ошибки несмещенной оценки функции  $g(\theta)$  верно стохастическое разложение

$$\begin{aligned} h_n(S_n) - g(\theta) = \\ = \frac{Z_n D_1}{\sqrt{n}} - \frac{H_2(Z_n)}{2n} [lD_0 - D_2 + \rho_3 D_1] + \\ + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j(Z_n) D_j + \mathbf{O}_P(n^{-2}). \end{aligned}$$

Вычисление интегралов  $\{D_j, j = \overline{0, 3}\}$  (см. (15)) можно упростить с помощью следующего утверждения.

**Лемма 2.** Пусть выполнены предположения  $(A_1)$ – $(A_4)$ . Предположим также, что справедливо условие:  $(A_6)$ . Производная  $v'''(\theta)$  существует и непрерывна.

Тогда интегралы  $\{D_j, j = \overline{0, 3}\}$  и функция  $g(\theta)$  связаны соотношениями

$$\begin{aligned} D_0 &= g(\theta), \\ D_1 &= \frac{g'(\theta)}{b}, \\ D_2 &= l g(\theta) + \frac{g''(\theta)}{b^2}, \\ D_3 &= \frac{g'''(\theta)}{b^3} + l \rho_3 g(\theta) + 3l \frac{g'(\theta)}{b}. \end{aligned} \tag{16}$$

При этом  $i(\theta)$  — информация Фишера, содержащаяся в одном наблюдении  $X_1$ :

$$i(\theta) = b^2.$$

Доказательство. Известно (см., например, утверждение задачи 14 гл. 2 [29]), что

$$\begin{aligned} a &= \mathbf{E}[T(\xi)] = -v'(\theta), \\ b^2 &= \mathbf{V}[T(\xi)] = -v''(\theta). \end{aligned} \tag{17}$$



Формулы (16) получим, последовательно вычисляя производные  $g'(\theta)$ ,  $g''(\theta)$  и  $g'''(\theta)$  с помощью формул (15) и (17), исходя из соотношения

$$D_0 = g(\theta) = \int_{-\infty}^{\infty} h_l(t) f_l(t; \theta) \mu(dt).$$

Лемма 2 доказана.

Используя лемму 2, запишем стохастическое разложение ошибки несмещенной оценки, приведенное в следствии 1, в другой форме, используя только изначально известные функции  $g(\theta)$  и  $v(\theta)$ .

**Следствие 2.** Пусть выполнены предположения  $(A_1)$ – $(A_6)$ . Тогда при  $n \rightarrow \infty$  справедливо стохастическое разложение ошибки несмещенной оценки

$$\begin{aligned} h_n(S_n) - g(\theta) &= \\ &= \frac{Z_n g'(\theta)}{b\sqrt{n}} - \frac{H_2(Z_n)}{2n} \left[ \frac{\rho_3 g'(\theta)}{b} - \frac{g''(\theta)}{b^2} \right] + \\ &\quad + \frac{1}{n^{3/2}} \left\{ \frac{H_3(Z_n) g'''(\theta)}{6b^3} + \right. \\ &\quad + \frac{g'(\theta)}{b} \left[ \frac{\rho_3^2 (H_3(Z_n) + Z_n)}{2} - \frac{\rho_4 H_3(Z_n)}{6} \right] - \\ &\quad \left. - \frac{\rho_3 (H_3(Z_n) + Z_n) g''(\theta)}{2b^2} \right\} + \mathbf{O}_P(n^{-2}). \end{aligned}$$

Теперь приведем стохастическое разложение оценки максимального правдоподобия  $g(\theta)$  функции  $g(\theta)$ , которое может быть получено с помощью формулы Тейлора, лемм 1 и 2 путем элементарных преобразований.

**Лемма 3.** Пусть выполнены предположения  $(A_1)$ – $(A_4)$  и  $(A_6)$ . Тогда при  $n \rightarrow \infty$  справедливо стохастическое разложение ошибки оценки максимального правдоподобия

$$\begin{aligned} g(\tilde{\theta}) - g(\theta) &= \frac{Z_n g'(\theta)}{b\sqrt{n}} - \frac{Z_n^2}{2n} \left[ \frac{\rho_3 g'(\theta)}{b} - \frac{g''(\theta)}{b^2} \right] + \\ &+ \frac{Z_n^3}{6n^{3/2}} \left[ \frac{(3\rho_3^2 - \rho_4) g'(\theta)}{b} - \frac{3\rho_3 g''(\theta)}{b^2} + \frac{g'''(\theta)}{b^3} \right] + \\ &\quad + \mathbf{O}_P(n^{-2}), \end{aligned}$$

где  $\tilde{\theta} = S_n/n$ .

Следующие два утверждения позволяют сравнить между собой несмещенную оценку и оценку максимального правдоподобия функции.

**Следствие 3.** Пусть выполнены предположения  $(A_1)$ – $(A_6)$ . Тогда при  $n \rightarrow \infty$  для разности несмещенной

оценки и оценки максимального правдоподобия верно стохастическое разложение

$$\begin{aligned} h_n(S_n) - g(\tilde{\theta}) &= \frac{1}{2n} \left[ \frac{\rho_3 g'(\theta)}{b} - \frac{g''(\theta)}{b^2} \right] + \\ &+ \frac{Z_n}{n^{3/2}} \left\{ \left[ \frac{\rho_4}{2} - \rho_3^2 \right] \frac{g'(\theta)}{b} + \frac{\rho_3 g''(\theta)}{b^2} - \frac{g'''(\theta)}{2b^3} \right\} + \\ &\quad + \mathbf{O}_P(n^{-2}). \end{aligned}$$

**Следствие 4.** Пусть выполнены предположения  $(A_1)$ – $(A_6)$ . Тогда при  $n \rightarrow \infty$

$$\begin{aligned} n \left[ h_n(S_n) - g(\tilde{\theta}) \right] &\xrightarrow{P} \frac{1}{2} \left[ \frac{\rho_3 g'(\theta)}{b} - \frac{g''(\theta)}{b^2} \right] + \\ &\quad + \mathbf{O}_P(n^{-1/2}). \end{aligned}$$

Утверждения следствий 3 и 4 уточняют результаты работы [9], в которой была установлена асимптотическая эквивалентность несмещенной оценки и оценки максимального правдоподобия с точностью до слагаемого  $\mathbf{O}_P(n^{-1})$ .

## 5 Стохастическое разложение несмещенной оценки дисперсии

В данном разделе получено стохастическое разложение несмещенной оценки дисперсии несмещенной оценки  $\hat{V}[h_n(S_n)]$ , определяемой формулой (9). Оказалось, что главный член этого разложения совпадает с нижней границей неравенства Рао–Крамера для дисперсии несмещенной оценки  $\mathbf{V}[h_n(S_n)]$ . Еще одно стохастическое разложение, приводимое здесь же, определяет предельное поведение ошибки несмещенной оценки (6), нормированной с помощью несмещенной оценки дисперсии.

**Теорема 3.** Пусть выполнены предположения  $(A_1)$ – $(A_6)$ . Тогда при  $s = na + zb\sqrt{n}$ ,  $n \rightarrow \infty$ , равномерно по  $z$  при  $|z| \leq z_0$  для любого фиксированного  $z_0$  справедливо асимптотическое разложение

$$\begin{aligned} \hat{V}[h_n(s)] &= \frac{D_1^2}{n} \left\{ 1 + \frac{z}{\sqrt{n}} \left[ \frac{2(D_2 - lD_0)}{D_1} - \rho_3 \right] \right\} + \\ &\quad + \mathbf{O}(n^{-2}). \end{aligned}$$

**Доказательство.** Учитывая, что  $\Delta_{2l}(t_1 + t_2) = \Delta_l(t_1) + \Delta_l(t_2)$ , и используя (11), получим сначала асимптотическое разложение для (10) при  $S_n = s$

$$\begin{aligned}
 U_2(s, t_1, t_2; \theta) &= \prod_{i=1}^2 \left\{ 1 + \frac{z \Delta_l(t_i)}{n^{1/2}} - \right. \\
 &\quad \left. - \frac{H_2(z)}{2n} [\rho_3 \Delta_l(t_i) + l - \Delta_l^2(t_i)] + \right. \\
 &\quad \left. + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j(z) \Delta_l^j(t_i) + P_4(\Delta_l(t_i)) \cdot \mathbf{O}(n^{-2}) \right\} - \\
 &\quad - \left\{ 1 + \frac{z [\Delta_l(t_1) + \Delta_l(t_2)]}{n^{1/2}} - \right. \\
 &\quad \left. - \frac{H_2(z)}{2n} [\rho_3 (\Delta_l(t_1) + \Delta_l(t_2)) + 2l - \right. \\
 &\quad \left. - (\Delta_l(t_1) + \Delta_l(t_2))^2] + \right. \\
 &\quad \left. + \frac{1}{n^{3/2}} \sum_{j=0}^3 c_j^*(z) [\Delta_l(t_1) + \Delta_l(t_2)]^j + P_4(\Delta_l(t_1) + \right. \\
 &\quad \left. + \Delta_l(t_2)) \cdot \mathbf{O}(n^{-2}) \right\} = \\
 &= U_2^*(s, t_1, t_2; \theta) + \mathbf{O}(n^{-2}) P_4^*(\Delta_l(t_1), \Delta_l(t_2)),
 \end{aligned}$$

где

$$\begin{aligned}
 U_2^*(z, t_1, t_2; \theta) &= \frac{\Delta_l(t_1) \Delta_l(t_2)}{n} + \\
 &\quad + \frac{1}{n^{3/2}} \left\{ \sum_{j=0}^3 c_j(z) [\Delta_l^j(t_1) + \Delta_l^j(t_2)] - \right. \\
 &\quad \left. - \sum_{j=0}^3 c_j^*(z) [\Delta_l(t_1) + \Delta_l(t_2)]^j - \right. \\
 &\quad \left. - \frac{z H_2(z)}{2} [l (\Delta_l(t_1) + \Delta_l(t_2)) + \right. \\
 &\quad \left. + 2 \rho_3 \Delta_l(t_1) \Delta_l(t_2) - \Delta_l^2(t_1) \Delta_l(t_2) - \right. \\
 &\quad \left. - \Delta_l(t_1) \Delta_l^2(t_2)] \right\};
 \end{aligned}$$

$$c_0^*(z) = 2c_0(z); \quad c_1^*(z) = c_1(z) - \frac{lH_3(z)}{2};$$

$$c_2^*(z) = c_2(z); \quad c_3^*(z) = c_3(z);$$

$P_4^*(x, y)$  — некоторый многочлен степени 4.

Воспользовавшись (15), преобразуем интеграл

$$\begin{aligned}
 &\iint_{\mathbf{R}^2} U_2^*(z, t_1, t_2; \theta) \prod_{j=1}^2 h_l(t_j) f_l(t_j; \theta) \mu(dt_j) = \\
 &= \frac{D_1^2}{n} + \frac{1}{n^{3/2}} \left\{ \sum_{j=0}^3 [c_j(z) 2D_0 D_j] - 2c_0(z) D_0^2 - \right. \\
 &\quad \left. - \left[ c_1(z) - \frac{lH_3(z)}{2} \right] 2D_0 D_1 - \right.
 \end{aligned}$$

$$\begin{aligned}
 &\left. - c_2(z) [2D_0 D_2 + 2D_1^2] - c_3(z) [2D_0 D_3 + 6D_1 D_2] - \right. \\
 &\quad \left. - \frac{z H_2(z)}{2} [2l D_0 D_1 + 2\rho_3 D_1^2 - 2D_1 D_2] \right\} = \\
 &= \frac{D_1^2}{n} \left\{ 1 + \frac{z}{\sqrt{n}} \left[ -\rho_3 + \frac{2(D_2 - lD_0)}{D_1} \right] \right\}. \quad (18)
 \end{aligned}$$

Обозначим  $B_2 = \{(t_1, t_2) \in \mathbf{R}^2 : |\Delta_l(t_1)| < n^\nu, |\Delta_l(t_2)| < n^\nu\}$ ,  $\overline{B_2} = \mathbf{R}^2 \setminus B_2$ . Используя (9), (10), (18) и условия  $(A_5)$  и  $(A_6)$ , оценим разность

$$\begin{aligned}
 \widehat{\mathbf{V}}[h_n(s)] &- \frac{D_1^2}{n} \left\{ 1 + \frac{z}{\sqrt{n}} \left[ -\rho_3 + \frac{2(D_2 - lD_0)}{D_1} \right] \right\} = \\
 &= \iint_{B_2} [U_2^*(z, t_1, t_2; \theta) + P_4^*(\Delta_l(t_1), \Delta_l(t_2)) \times \\
 &\quad \times \mathbf{O}(n^{-2})] \prod_{j=1}^2 h_l(t_j) f_l(t_j; \theta) \mu(dt_j) + \\
 &\quad + \mathbf{O}(1) \iint_{\overline{B_2}} \prod_{j=1}^2 h_l(t_j) f_l(t_j; \theta) \mu(dt_j) - \\
 &\quad - \iint_{\mathbf{R}^2} U_2^*(z, t_1, t_2; \theta) \prod_{j=1}^2 h_l(t_j) f_l(t_j; \theta) \mu(dt_j) = \\
 &= \mathbf{O}(n^{-2}),
 \end{aligned}$$

что и завершает доказательство теоремы 3.

**Следствие 5.** Пусть выполнены предположения  $(A_1)$ – $(A_6)$ . Тогда при  $n \rightarrow \infty$

$$\begin{aligned}
 \widehat{\mathbf{V}}[h_n(S_n)] &= \frac{[g'(\theta)]^2}{nb^2} \left\{ 1 + \right. \\
 &\quad \left. + \frac{Z_n}{\sqrt{n}} \left[ \frac{2g''(\theta)}{bg'(\theta)} - \rho_3 \right] \right\} + \mathbf{O}_P(n^{-2}). \quad (19)
 \end{aligned}$$

**Примечание.** Как видно из (19), главный член разложения несмещенной оценки дисперсии несмещенной оценки  $\widehat{\mathbf{V}}[h_n(S_n)]$  совпадает с нижней границей неравенства Рао–Крамера для дисперсии несмещенной оценки дисперсии  $\mathbf{V}[h_n(S_n)]$ , поскольку в силу леммы 2 при сделанных предположениях информация Фишера, содержащаяся в одном наблюдении, равна  $b^2$ .

Наконец, используя утверждения следствий 2 и 5, нетрудно получить результат, который можно применить к построению асимптотических доверительных интервалов, основываясь на несмещенных оценках.

**Следствие 6.** Пусть выполнены предположения  $(A_1)–(A_6)$ ,  $g'(\theta) \neq 0$ . Тогда

$$\begin{aligned} \text{sign}[g'(\theta)] \frac{h_n(S_n) - g(\theta)}{\sqrt{\widehat{V}[h_n(S_n)]}} &= \\ &= Z_n + \frac{1}{2\sqrt{n}} \left[ \rho_3 - \frac{g''(\theta)}{bg'(\theta)} (Z_n^2 + 1) \right] + \mathbf{O}_P(n^{-1}) \end{aligned}$$

при  $n \rightarrow \infty$ .

## 6 Заключение

Найдены стохастические разложения, определяющие предельное поведение всех несмещенных оценок, которые могут быть построены на основе повторной выборки из распределения, принадлежащего однопараметрическому экспоненциальному семейству. Реализованный в работе подход к построению стохастических разложений может быть перенесен на случай многопараметрического экспоненциального семейства. Полученные результаты могут быть применены к решению разнообразных асимптотических параметрических задач интервального оценивания и проверки статистических гипотез, основываясь на несмещенных оценках.

## Литература

1. Воинов В. Г., Никулин М. С. Несмещенные оценки и их применения. — М.: Наука, 1989.
2. Voinov V. G., Nikulin M. S. Unbiased estimators and their applications. Vol. 2: Multivariate case. — Dordrecht, the Netherlands, 1996.
3. Pommeret D. A construction of the UMVU estimator for simple quadratic natural exponential families // J. Multivariate Analysis, 2003. Vol. 85. P. 217–233.
4. Blazquez F. L., Rubio D. G. Unbiased estimation in the multivariate natural exponential family with simple quadratic variance function // J. Multivariate Analysis, 2003. Vol. 86. P. 1–13.
5. Aghili A. On minimum variance unbiased estimators of exponential families // Int. Math. J., 2004. Vol. 4. P. 383–387.
6. Wang B. Unbiased estimations for the exponential distribution based on step-stress accelerated life-testing data // Applied Mathematics and Computation, 2006. Vol. 173. P. 1227–1237.
7. Liu A., Wu C., Yu K. F., Yuan W. Completeness and unbiased estimation of mean vector in the multivariate group sequential case // J. Multivariate Analysis, 2007. Vol. 98. P. 505–516.
8. Sharma D. Asymptotic equivalence for two estimators for an exponential family // Annals of Statistics, 1973. P. 973–980.
9. Portnoy S. Asymptotic efficiency of minimum variance unbiased estimators // Annals of Statistics, 1977. Vol. 5. No. 3. P. 522–529.
10. Lopez-Blazquez F., Salamanca-Mino B. Limit distributions of unbiased estimators in natural exponential families // Statistics, 2002. Vol. 36. No. 4. P. 329–338.
11. Morris C. N. Natural exponential families with quadratic variance functions: Statistical theory // Annals of Statistics, 1983. Vol. 11. No. 2. P. 515–529.
12. Чичагов В. В. Об асимптотической нормальности одного класса несмещенных оценок в случае абсолютно непрерывных распределений / Стат. методы оценивания и проверки гипотез. — Пермь: Изд-во Пермского университета, 2000. С. 71–79. [Перевод: Chichagov V. V. Concerning asymptotic normality of a class of unbiased estimators in the case of absolutely continuous distributions / Statistical methods of estimation and testing of hypotheses // J. Math. Sci. (N.Y.), 2004. Vol. 119. No. 3. P. 336–341.]
13. Чичагов В. В. Об асимптотическом поведении несмещенных оценок вероятностей для решетчатых распределений, достаточной статистикой которых является среднее / Стат. методы оценивания и проверки гипотез. — Пермь: Изд-во Пермского университета, 2002. С. 106–120.
14. Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972.
15. Bhattacharya R. N., Ghosh J. K. On the validity of the formal Edgeworth expansions // Annals of Statistics, 1978. Vol. 6. P. 434–451.
16. Barndorff-Nielsen O. E., Cox D. R. Edgeworth and Saddlepoint approximations with statistical applications // J. R. Statist. Soc. B, 1979. Vol. 41. P. 279–312.
17. Барндорф-Нильсен О., Кокс Д. Асимптотические методы в математической статистике. — М.: Мир, 1999.
18. Lieberman G. J., Resnikoff G. J. Sampling plans for inspection by variables // J. Amer. Statist. Assoc., 1955. Vol. 50. P. 457–516.
19. Patil G. P., Wani J. K. Minimum variance unbiased estimation of the distribution function admitting a sufficient statistics // Ann. Inst. Statist. Math., 1966. Vol. 18. P. 39–47.
20. Лумельский Я. П., Сапожников П. Н. Несмещенные оценки для плотностей распределений // Теория вероятностей и ее применение, 1969. Vol. XIV. № 2. С. 372–380.
21. Лумельский Я. П. Случайные блуждания, отвечающие обобщенным урновым схемам // ДАН СССР, 1973. Т. 209. № 6. С. 1281–1284.
22. Абусев Р. А., Лумельский Я. П. Несмещенные оценки и задачи классификации многомерных нормальных совокупностей // Теор. вероятностей и ее применение, 1980. Т. 25. № 2. С. 381–389.
23. Michel R. Asymptotic expansions for conditional distributions // J. Multivariate Anal., 1979. P. 393–400.

24. *Hipp C.* Asymptotic expansions for conditional distributions: The lattice case // *Probab. Math. Statist.*, 1984. Vol. 4. P. 207–219.
25. *Skovgaard I. M.* Saddlepoint expansions for conditional distributions // *J. of Applied Probability*, 1987. Vol. 24. No. 4. P. 875–887.
26. *Лумельский Я. П.* Несмещенные достаточные оценки вероятностей в случае многомерного нормального закона // *Вестник МГУ*, 1968. Vol. 6. С. 14–17. [Перевод: *Lumelskii Ya. P.* Unbiased sufficient estimators for probabilities in the case of a multivariate normal law // *Vestnik Moskov. Univ. Ser. Mat. Meh.*, 1968. Vol. 23. No. 6. P. 14–17; Selected translations in mathematical statistics and probability. Vol. 13. Amer. Math. Soc., Providence, R.I., 1973. P. 251–256.]
27. *Ившин В. В., Лумельский Я. П.* Статистические задачи оценивания в модели «нагрузка–прочность». — Пермь: Изд-во Пермского университета, 1995.
28. *Лумельский Я. П., Фейгин П. Д.* Несмещенные оценки дисперсии в параметрическом случае / *Стат. методы оценивания и проверки гипотез.* — Пермь: Изд-во Пермского университета, 2002. С. 38–51.
29. *Леман Э.* Проверка статистических гипотез. — М.: Наука, 1979.

### NETWORK METHODS OF SEPARATION OF MIXTURES OF PROBABILITY DISTRIBUTIONS AND THEIR APPLICATION TO THE DECOMPOSITION OF VOLATILITY INDEXES

V. Korolev<sup>1</sup>, E. Nepomnyashchiy<sup>2</sup>, A. Rybal'chenko<sup>3</sup>, and A. Vinogradova<sup>4</sup>

<sup>1</sup>M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@comtv.ru

<sup>2</sup>45th Central Science Research Institute, Russian Ministry of Defence

<sup>3</sup>M. V. Lomonosov Moscow State University, alex-rybalchenko@yandex.ru

<sup>4</sup>M. V. Lomonosov Moscow State University, a\_nuta@mail.ru

Methods are proposed for the statistical separation of mixtures of probability distributions based on the minimization of the discrepancy between the theoretical and empirical distribution functions. Main attention is paid to the minimization of sup- and L1-norms of the discrepancy. It is demonstrated that these problems can be reduced to problems of linear programming. The simplex-method is used for their numerical realization. The proposed methods are applied to the problem of decomposition of the volatility of financial indexes. Examples of the decomposition of the volatility of AMEX, CAC 40, NIKKEI, and NASDAQ indexes are presented.

**Keywords:** separation of mixtures of probability distributions; problem of linear programming; simplex-method; volatility

### SOME STATISTICAL PROBLEMS RELATED TO THE LAPLACE DISTRIBUTION

V. Bening<sup>1</sup> and V. Korolev<sup>2</sup>

<sup>1</sup>M. V. Lomonosov Moscow State University; IPI RAN, bening@yandex.ru

<sup>2</sup>M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@comtv.ru

Approach proposed in *Bening V., Korolev V. // Rus. J. Probability Theory and Its Applications, 2004. Vol. 49. No. 3. P. 419–435* is developed. Some grounds are given why it is natural to use the Laplace distribution in some problems of probability theory and mathematical statistics. As a statistical illustration, an application of the Laplace distribution to problems of the asymptotical testing of hypotheses is considered.

**Keywords:** Laplace distribution; testing of statistical hypotheses; power function; asymptotically most powerful tests

### SOME BOUNDS FOR CLOSED TO ABSORBING MARKOV MODELS

A. I. Zeifman<sup>1</sup>, A. V. Chegodaev<sup>2</sup>, and V. S. Shorgin<sup>3</sup>

<sup>1</sup>VSPU; IPI RAN; VSCC CEMI RAS, a\_zeifman@mail.ru

<sup>2</sup>VSPU, cheg\_al@mail.ru

<sup>3</sup>IPI RAN, vshorgin@ipiran.ru

Nonstationary continuous-time Markov models with “almost absorbing” zero state are considered. Such models describe some problems of queueing theory. The properties and bounds for the limit characteristics of such models are obtained. Simple continuous-time random walks as example of considered models are also studied.

**Keywords:** queueing networks; continuous-time Markov chains; ergodicity; birth and death process; simple random walk

## THE OPTIMIZATION OF THE SPATIAL LOCATION OF SERVICE STATIONS

T. Zakharova

M. V. Lomonosov Moscow State University, lsa@cs.msu.su

In the  $N$ -dimensional Euclidean space, service stations are being efficiently located according to the criterion of stationary average waiting time.

**Keywords:** service stations location; optimality criterion; waiting time

## ELIMINATION OF ECTOPIC BEATS FROM HEART TACHOGRAM USING ROBUST ESTIMATES

A. V. Markin<sup>1</sup> and O. V. Shestakov<sup>2</sup><sup>1</sup>M. V. Lomonosov Moscow State University, artem.v.markin@mail.ru<sup>2</sup>M. V. Lomonosov Moscow State University, oshestakov@cs.msu.su

A method of removing ectopic beats from tachogram is introduced. The method is based on authors' mathematical model. Parameters of the model are estimated using robust linear regression. Elimination of ectopic beats is made by the instrumentality of confidential intervals for RR-interval differences. Real data results are supplied and discussed.

**Keywords:** tachogram; robust estimates; linear regression; confidential intervals

## ON THE ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM ORDER STATISTIC IN A SAMPLE WITH RANDOM SIZE

V. Pagurova

<sup>1</sup>M. V. Lomonosov Moscow State University, pagurova@yandex.ru

The asymptotic distribution of the normalized maximum is investigated under the assumption that the random sample size is representable as a sum of  $n$  independent identically distributed random variables. This paper generalizes the results of *Pagurova V.* // Statistical methods of estimating and testing of hypotheses. — Perm, 2005. P. 104–113 where the sample size was Poisson-distributed with a parameter  $n$ . For a one-parameter family of distributions depending on an unknown location parameter, the rate of convergence of the distribution of the normalized maximum to the limit law is investigated. The classes of distributions with exponential and power-type tails are considered.

**Keywords:** randomly indexed maximum; one-parameter family of distributions; convergence rate

## ESTIMATION OF DELAY DISTRIBUTION IN BIOLOGICAL DYNAMICAL MODELS WITH A MODEL OF HIV INFECTION AS AN EXAMPLE

A. N. Ushakova

Norwegian University of Science and Technology, anastasi@math.ntnu.no

Two methods for estimating a delay distribution in biological dynamical systems are presented. The model of HIV infection serves as an example of such a system. The first method is based on parametric approach and approximation of the delay density by a gamma-density. The second method is nonparametric and is based on solution of a convolution equation with selection of the regularization parameter either from the measurement error estimated beforehand, or from the smoothness of the delay distribution.

**Keywords:** dynamical systems with delay; distribution of delay; regularization parameter

EXISTENCE OF CONSISTENT TEST SEQUENCES AT THE COMPLEX NULL HYPOTHESES  
IN DISCRETE STATISTICAL PROBLEMS

A. Grusho<sup>1</sup>, E. Timonina<sup>2</sup>, and V. Chentsov<sup>3</sup>

<sup>1</sup>Russian State Humanitarian University; M. V. Lomonosov Moscow State University, grusho@yandex.ru

<sup>2</sup>Russian State Humanitarian University, eltimon@yandex.ru

<sup>3</sup>IPI RAN, ipiran@ipiran.ru

The problem of existence of a consistent test sequence is considered for testing of complex hypothesis against complex alternatives in a sequence of finite spaces. When the sequence of spaces is generated by Cartesian product of a finite set and probability measures on these spaces are consistent, it is possible to find sufficient conditions of existence of consistent test sequence in terms of topological properties of the certain sets. Under additional conditions it is possible to refuse the requirement of domination of certain measures from a null hypothesis and uniform limitation of density.

**Keywords:** consistent test sequence; complex hypothesis against complex alternatives; finite spaces; probability measures; sufficient conditions

STOCHASTIC EXPANSIONS OF UNBIASED ESTIMATORS FOR THE CASE  
OF ONE-PARAMETER EXPONENTIAL FAMILY

V. Chichagov

Perm State University, chvv50@mail.ru

Asymptotic and stochastic expansions are obtained for all unbiased estimators which can be constructed from a repeated sample of independent identically distributed random variables having one and the same distribution from a one-parameter exponential family. A comparison is conducted of expansions of unbiased estimators with those of maximum likelihood estimators. A stochastic expansion is obtained for an unbiased estimator of the variance of an unbiased estimator.

**Keywords:** unbiased estimator; exponential family; stochastic expansion; unbiased estimator of variance

## Об авторах

**Бенинг Владимир Евгеньевич** (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова, старший научный сотрудник ИПИ РАН

**Виноградова Анна Валерьевна** (р. 1985) — студентка магистратуры факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Грушо Александр Александрович** (р. 1946) — доктор физико-математических наук, член-корреспондент Академии криптографии РФ; заведующий кафедрой компьютерной безопасности факультета защиты информации Института информационных наук и технологий безопасности Российского государственного гуманитарного университета; профессор кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Захарова Татьяна Валерьевна** (р. 1967) — кандидат физико-математических наук, младший научный сотрудник кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Зейфман Александр Израилевич** (р. 1954) — доктор физико-математических наук, профессор, декан факультета прикладной математики Вологодского государственного педагогического университета, старший научный сотрудник ИПИ РАН, ведущий научный сотрудник Вологодского научно-координационного центра ЦЭМИ РАН

**Королёв Виктор Юрьевич** (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова, ведущий научный сотрудник ИПИ РАН

**Маркин Артём Васильевич** (р. 1985) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Непомнящий Евгений Владимирович** (1947–2007) — доктор технических наук; работал ведущим научным сотрудником 45 ЦНИИ МО РФ

**Пагурова Вера Игнатьевна** (р. 1933) — кандидат физико-математических наук, старший научный сотрудник кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Рыбальченко Александр Глебович** (р. 1986) — студент кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Тимонина Елена Евгеньевна** (р. 1952) — доктор технических наук, профессор кафедры фундаментальной и прикладной математики Института информационных наук и технологий безопасности Российского государственного гуманитарного университета

**Ушакова Анастасия Николаевна** (р. 1981) — аспирантка Департамента математических наук Норвежского университета науки и технологии, Тронхейм, Норвегия

**Чегодаев Александр Вячеславович** (р. 1983) — аспирант факультета прикладной математики Вологодского государственного педагогического университета

**Ченцов Валерий Михайлович** (р. 1938) — доктор технических наук, старший научный сотрудник ИПИ РАН

**Чичагов Владимир Витальевич** (р. 1955) — кандидат физико-математических наук, доцент Пермского государственного университета

**Шестаков Олег Владимирович** (р. 1976) — кандидат физико-математических наук, ассистент кафедры математической статистики факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова

**Шоргин Всеволод Сергеевич** (р. 1978) — кандидат технических наук, старший научный сотрудник ИПИ РАН



## About Authors

**Bening Vladimir E.** (b. 1954) — Doctor of Science in physics and mathematics; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences

**Chegodaev Alexander V.** (b. 1983) — PhD student, Faculty of Applied Mathematics, Vologda State Pedagogical University

**Chentsov Valery M.** (b. 1938) — Doctor of Science in technology; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences

**Chichagov Vladimir V.** (b. 1955) — Candidate of Science (PhD) in physics and mathematics; associate professor, Perm State University

**Grusho Alexander A.** (b. 1946) — Doctor of Science in physics and mathematics; Corresponding member of the Russian Academy of Cryptography; Head of Department of Computer Security, Faculty of Information Protection, Institute of Information Sciences and Security Technologies, Russian State Humanitarian University; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Korolev Victor Y.** (b. 1954) — Doctor of Science in physics and mathematics; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; leading researcher, Institute of Informatics Problems, Russian Academy of Sciences

**Markin Artem V.** (b. 1985) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Nepomnyashchiy Evgeny V.** (1947–2007) — Doctor of Science in technology; worked as leading researcher in the 45th Central Science Research Institute, Russian Ministry of Defence

**Pagurova Vera I.** (b. 1933) — Candidate of Science (PhD) in physics and mathematics; senior researcher, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Rybal'chenko Alexander G.** (b. 1986) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Shestakov Oleg V.** (b. 1976) — Candidate of Science (PhD) in physics and mathematics; assistant professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Shorgin Vsevolod S.** (b. 1978) — Candidate of Science (PhD) in technology; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences

**Timonina Elena E.** (b. 1952) — Doctor of Science in technology; professor, Department of Fundamental and Applied Mathematics, Institute of Information Sciences and Security Technologies, Russian State Humanitarian University

**Ushakova Anastassia N.** (b. 1981) — PhD student, Department of Mathematical Sciences, Norwegian University of Science and Technology

**Vinogradova Anna V.** (b. 1985) — Master student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Zakharova Tatyana V.** (b. 1967) — Candidate of Science (PhD) in physics and mathematics; junior researcher, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

**Zeifman Alexander I.** (b. 1954) — Doctor of Science in physics and mathematics; professor, Dean of the Faculty of Applied Mathematics, Vologda State Pedagogical University; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences; leading researcher, Vologda Science-Coordination Center, Central Economy and Mathematics Institute, Russian Academy of Sciences

# Правила подготовки рукописей статей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений. Журнал издается на русском языке. Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях. Публикация не должна нарушать закон об авторских правах. Направляя свою рукопись в редакцию, авторы автоматически передают учредителям и редколлегии неисключительные права на издание данной статьи на русском языке и на ее распространение в России и за рубежом. При этом за авторами сохраняются все права как собственников данной рукописи. В связи с этим авторами должно быть представлено в редакцию письмо в следующей форме: Соглашение о передаче права на публикацию:

*«Мы, нижеподписавшиеся, авторы рукописи « \_\_\_\_\_ », передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций. Подписи авторов: (ф. и. о., дата, адрес)».*

Редколлегия вправе запросить у авторов экспертное заключение о возможности опубликования представленной статьи в открытой печати.

2. Статья подписывается всеми авторами. На отдельном листе представляются данные автора (или всех авторов): фамилия, полное имя и отчество, телефон, факс, e-mail, почтовый адрес. Если работа выполнена несколькими авторами, указывается фамилия одного из них, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет самостоятельную экспертизу присланных статей. Возвращение рукописи на доработку не означает, что статья уже принята к печати. Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.
4. Решение редакционной коллегии о принятии статьи к печати или ее отклонении сообщается авторам. Редколлегия не обязуется направлять рецензию авторам отклоненной статьи.
5. Корректурa статей высылается авторам для просмотра. Редакция просит авторов присылать свои замечания в кратчайшие сроки.
6. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки. Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3. Основной текст: стиль — «Обычный»; шрифт Times New Roman, размер 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине. Рекомендуемый объем рукописи — не свыше 25 страниц указанного формата. Ознакомиться с шаблонами, содержащими примеры оформления, можно по адресу в Интернете: <http://www.ipiran.ru/journal/template.doc>.
7. К рукописи, предоставляемой в 2-х экземплярах, обязательно прилагается электронная версия статьи (как правило, в форматах MS WORD (.doc) или LaTeX (.tex), а также — дополнительно — в формате .pdf) на дискете, лазерном диске или по электронной почте. Сокращения слов, кроме стандартных, не применяются. Все страницы рукописи должны быть пронумерованы.
8. Статья должна содержать следующую информацию на русском и английском языках: название, Ф.И.О. авторов, места работы авторов и их электронные адреса, аннотация (не более 100 слов), ключевые слова. Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в порядке их первого упоминания. Все фамилии авторов, заглавия статей, названия книг, конференций и т. п. даются на языке оригинала, если этот язык использует кириллический или латинский алфавит.
9. Присланные в редакцию материалы авторам не возвращаются.
10. При отправке файлов по электронной почте просим придерживаться следующих правил:
  - указывать в поле subject (тема) название журнала и фамилию автора;
  - использовать attach (присоединение);
  - в случае больших объемов информации возможно использование общеизвестных архиваторов (ZIP, RAR);
  - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
11. Журнал «Информатика и её применения» является некоммерческим изданием, и гонорар авторам не выплачивается.

**Адрес редакции:** Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05 E-mail: [rust@ipiran.ru](mailto:rust@ipiran.ru)

