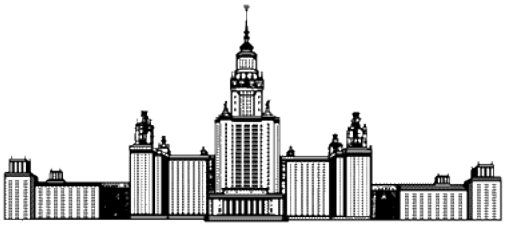
**Московский государственный университет имени М.В. Ломоносова**
**Факультет вычислительной математики и кибернетики**



Кафедра математических методов прогнозирования
Магистерская программа «Логические и комбинаторные методы анализа данных»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

# Проверка метрических характеристик коммуникационной среды вычислительного кластера

**Работу выполнил:**
студент 617 группы
Козлов Владимир Дмитриевич

**Научный руководитель:**
к. ф.-м. н, доцент
Майсурадзе Арчил Иериевич

Москва 2017

# Оглавление

# Глава 1

# Введение

Многие современные распределённые вычислительные системы состоят из тысяч и десятков тысяч процессоров. С увеличением числа процессоров в системе растёт сложность её коммуникационной среды, что приводит к увеличению времени, затрачиваемому на обмен информацией между вычислительными устройствами. Таким образом, эффективное использование современных многопроцессорных систем для выполнения параллельных вычислений опирается не только на характеристики составляющих её процессоров, но и на характеристики её коммуникационной среды. Одним из ключевых инструментов разработки параллельных программ для многопроцессорных систем является библиотечная реализация стандарта MPI (Message Passing Interface) [1]. При использовании технологии MPI программа разделяется на процессы, все взаимодействия между которыми производятся посредством обмена сообщениями. Учёт информации о задержках при передаче сообщений между вычислительными узлами необходим для эффективного использования ресурсов современных вычислительных кластеров.

Одной из областей применения анализа задержек является планирование распределения вычислительных ресурсов между подзадачами. Часто для решения задач планирования применяется статический подход, при котором распределение ресурсов между процессами производится ещё до запуска программы. Однако появление задач, размер и число подзадач в которых становится известно только в ходе выполнения, привело к развитию методов динамического планирования. При динамическом планировании для оптимизации общего времени решения задачи на кластере должны учитываться не только вычислительные мощности отдельных процессоров, но и время передачи сообщений между процессами. Таким образом, система динамического планирования должна иметь модель задержек в коммуникационной сети. Такая модель должна быть достаточно компактной и обеспечивать быстрый

доступ к информации о задержках.

Анализ задержек также может применяться при администрировании вычислительного кластера. Коммуникационная среда суперкомпьютера подвержена различным неполадкам и сбоям, как программным, так и аппаратным. Необычная картина задержек при передаче сообщений может сигнализировать о неисправности в системе. Для тестирования коммуникационной среды часто используются специальны программные продукты; типичным результатом работы таких программ являются некоторые статистики полученной в ходе тестов выборки задержек, например, среднее значение, медиана, максимальное и минимальное значения. Однако при таком описании теряется информация о структуре задержек. Таким образом, для тестирования коммуникационной среды требуется модель задержек, более полная, чем набор статистик.

Основной целью настоящей работы является построение модели задержек при передаче сообщений в коммуникационной среде вычислительного кластера для решения задач динамического планирования выполнения и тестирования коммуникационной среды. Как уже было сказано выше, такая модель должна предоставлять как можно более полное описание структуры задержек и при этом быть достаточно компактной, чтобы её можно было хранить и использовать в реальном времени. Величины задержек зависят от множества факторов, специфичных для разных систем и меняющихся со временем, учёт которых при моделировании задержек требует анализа программного и аппаратного обеспечения коммуникационной среды на каждом уровне сетевого протокола. Это делает построение модели «белого ящика» возможным только для простых систем и конкретных задач на них. В связи с этим в работе рассматривается подход к моделированию задержек методом «чёрного ящика», при котором неконтролируемые факторы рассматриваются как скрытые параметры модели, а задержки— как случайные величины с некоторыми распределениями. Из этого вытекают основные задачи работы:

1. подбор подходящего семейства распределений для описания задержек;
2. выбор способа оценки параметров распределения;
3. проверка применимости полученной модели для задачи диагностики кластера.

Работа устроена следующим образом. В главе 2 приводится формальная постановка решаемой задачи. Главы 3 и 4 посвящены выбору метода оценки параметров модели. В главе 5 проводится проверка метрических характеристик коммуникационных сред реальных суперкомпьютеров. В главе 6 делается заключение по проведённой работе.

# Глава 2

# Задача моделирования задержек

## 2.1 Модель вычислительного кластера

Объектом исследования в настоящей работе служат вычислительные кластеры. Дадим формальное определение этого класса многопроцессорных систем.

Все архитектуры вычислительных систем можно разбить на 4 класса на основании наличия параллелизма в потоках команд и данных [2]:

- Системы с одиночным потоком команд и одиночным потоком данных (SISD, Single Instruction stream over a Single Data stream): машины, в которых есть один поток команд, команды обрабатываются последовательно и каждая команда описывает одну операцию с одним потоком данных. К этому классу относятся классические последовательные машины.

- Системы со одиночным потоком команд и множественным потоком данных (SIMD, Single Instruction, Multiple Data): машины, в которых один поток команд включает инструкции по обработке нескольких потоков данных. Пример— векторные машины.

- Системы со множественным потоком команд и одиночным потоком данных (MISD, Multiple Instruction, Single Data): машины, в которых несколько потоков команд обрабатывают один поток данных. Класс считается пустым.

- Системы со множественным потоком команд и множественным потоком данных (MIMD, Multiple Instruction, Multiple Data): машины, в которых несколько устройств по обработке команд объединены в комплекс, и каждое работает над своим потоком команд и данных. В системах класса MIMD возникает

необходимость в передаче сообщений между вычислительными устройствами, соответственно, в настоящей работе рассматриваются именно они.

Все системы класса MIMD, в свою очередь, разделяются на два класса: системы с общей и распределённой памятью. В системах с общей памятью все вычислительные устройства работают в едином адресном пространстве. В системах с распределённой памятью адресное пространство разделено между процессорами. В таких системах все процессоры разбиваются на группы, работающие в одном адресном пространстве. Такие группы называются вычислительными узлами. Вычислительные узлы объединены некоторой коммуникационной средой, и получение информации с других узлов может проходить только через неё.

**Определение.** Вычислительный узел— группа процессоров, работающих в едином адресном пространстве. Как правило, представляет собой отдельную вычислительную машину, работающую под управлением своей операционной системы.

**Определение.** Коммуникационная сеть— множество вычислительных узлов и коммуникационных связей между ними.

**Определение.** Вычислительный кластер— набор вычислительных узлов, объединённых коммуникационной сетью, выполняющий вычисления и представляемый пользователю как единая система.

## 2.2 Модель задержек

Предметом исследования являются задержки сообщений MPI в коммуникационной сети вычислительного кластера.

**Определение.** Сообщение — последовательность байтов определённой длины, передающаяся от одного вычислительного узла к другому через коммуникационную сеть вычислительного кластера. Задержка при передаче сообщения $t_{s,a\to b}$ — интервал времени между отправкой сообщения длиной $s$ на узле $a$ и его приёмом на узле $b$.

В рамках настоящей работы не рассматривается зависимость задержек от структуры сообщений. Считается, что распределение задержек зависит только от узла— отправителя, узла— получателя и длины сообщения.

Проводившиеся ранее исследования задержек в локальных сетях и сети Интернет [3, 4, 5] показывают, что величины задержек хорошо описываются трёхпараметрическим гамма- или логнормальным распределением. Однако, как было показано в работе [6], в коммуникационной среде различных суперкомпьютеров наблюдаются наблюдаются одна или сразу несколько из следующих особенностей:

Таблица 2.1. Основные моментные характеристики трёхпараметрического логнормального распределения с параметрами $\gamma$, $\mu$, $\sigma$ [8] ($\beta = \exp(\mu)$, $\omega = \exp(\sigma^2)$)

| Характеристика | Формула |
|---|---|
| Математическое ожидание $\mathbb{E}$ | $\gamma + \beta\sqrt{\omega}$ |
| Дисперсия $\mathbb{D}$ | $\beta^2\omega(\omega-1)$ |
| Коэффициент асимметрии $\alpha_3$ | $\sqrt{\omega-1}(\omega+2)$ |
| Коэффициент эксцесса $\alpha_4$ | $\omega^4+2\omega^3+3\omega^2-6$ |

1. распределение является многомодальным;
2. в данных очень много повторов и мало уникальных значений;
3. наблюдается достаточно много задержек максимальной величины, что может говорить о тяжёлом хвосте распределения.

Дополнительной особенностью задачи является её размер. Количество вычислительных узлов на современных суперкомпьютерах измеряется десятками тысяч, что делает сбор и обработку больших объёмов данных о задержках в коммуникационной сети между всеми парами узлов чрезвычайно затратными.

Исходя из указанных выше особенностей, в работе [6] было предложено моделировать величины задержек смесью трёхпараметрических логнормальных распределений. Таким образом, задача оценки параметров модели распределения задержек может быть формализована как задача разделения конечной смеси трёхпараметрических логнормальных распределений.

**Определение.** Трёхпараметрическое логнормальное распределение (3LN распределение)— это абсолютно непрерывное одномерное распределение, функция плотности вероятности которого выражается формулой

$$p(x;\gamma,\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma(x-\gamma)} \exp\left(-\frac{(\ln(x-\gamma)-\mu)^2}{2\sigma^2}\right) & , x \geq \gamma, \\ 0 & , x < \gamma, \end{cases} \tag{2.1}$$

Случайная величина $X$ имеет 3LN распределение с параметрами $\gamma$, $\mu$ и $\sigma$, если случайная величина $\ln(X-\gamma)$ имеет нормальное распределение с параметрами $\mu$ и $\sigma$ [7]. Её функция распределения может быть записана в виде $F(x;\gamma,\mu,\sigma) = \Phi\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)$, где $\Phi(x)$ — функция распределения стандартного нормального закона [8]. Основные моментные характеристики распределения указаны в таб. 2.1.

Набор параметров $\gamma, \mu, \sigma$ будет обозначаться $\theta$. Будем обозначать случайную выборку длины $n$ $X^n = (X_1, \ldots, X_n)$, её реализацию— $x^n = (x_1, \ldots, x_n)$, $k$-ю

порядковую статистику и её реализацию— $X_{(k)}$ и $x_{(k)}$ соответственно.

## 2.3 Метрические характеристики среды

В работе [6] для тестирования коммуникационной среды были предложены так называемые обобщённые метрические характеристики:

1. $t_{s_1,a\to b} \leqslant t_{s_2,a\to b} \forall a,b$, если $s_1 \leqslant s_2$ (*монотонность*);
2. $t_{s_1+s_2,a\to b} \leqslant t_{s_1,a\to b} + t_{s_2,a\to b} \forall a,b,s_1,s_2$ (*неделимость*);
3. $t_{s,a\to c} \leqslant t_{s,a\to b} + t_{s,b\to c} \forall a,b,c,s$ (*неравенство треугольника*).

Эти условия естественным образом вытекают из природы задержек, и их нарушение может сигнализировать о неполадках в коммуникационной сети.

Согласно предложенной модели, задержки представляют собой случайные величины, распределённые как смесь 3LN распределений. Таким образом, проверка обобщённых метрических характеристик представляет собой проверку неравенства случайных величин. Соответственно, перед нами стоит задача стохастического упорядочивания. Одним из способов введения порядка в пространстве случайных величин является понятие *стохастического доминирования*:

**Определение.** Говорят, что наблюдается стохастическое доминирование первого порядка случайной величины $X$ над случайной величиной $Y$, если

$$\begin{aligned} &P(X > a) \geqslant P(Y > a) \forall a \in \mathbb{R}; \\ &\exists a : P(X > a) > P(Y > a). \end{aligned} \tag{2.2}$$

Такой способ введения порядка достаточно естественен. Условие 2.2 можно переписать в эквивалентном виде: $F_X(a) \leqslant F_Y(a) \forall a \in \mathbb{R}$. В настоящей работе рассматривается только стохастическое доминирование первого порядка, однако отметим, что можно ввести стохастическое доминирование больших порядков:

**Определение.** Говорят, что наблюдается стохастическое доминирование порядка $s$ случайной величины $X$ над случайной величиной $Y$, если

$$\begin{aligned} &D^s_X(x) \leqslant D^s_Y(x) \forall x \in \mathbb{R}; \\ &\exists x : D^s_X(x) < D^s_Y(x), \end{aligned} \tag{2.3}$$

где $D^s_A(x) = \int_{-\infty}^{x} D^{s-1}_A(t)dt$, $D^1_A(x) = F(x)$.

Задачей работы является построение модели задержек для проверки обобщённых метрических характеристик.

# Глава 3

# Оценка параметров 3LN распределения

Как уже было сказано выше, в главе 2, задача оценки параметров модели формализуется как задача разделения конечной смеси 3LN распределений. Однако, как показал обзор литературы, проблемы возникают даже при параметрическом восстановлении одного компонента такой смеси. В этой главе рассматриваются и анализируются различные способы оценки параметров 3LN распределения.

## 3.1 Метод максимума правдоподобия и его модификация

В методе максимума правдоподобия в качестве меры адекватности распределения $F(\cdot;\theta)$ данным $X^n$ используется функция правдоподобия $L(\theta)$, равная $p(x_1,\ldots,x_n;\theta)$ — совместной плотности вероятности объектов выборки. Полагается, что чем больше правдоподобие, тем лучше модель описывает данные ([9]).

Для семейства трёхпараметрических логнормальных распределений в предположении, что все элементы выборки — независимые одинаково распределённые случайные величины, логарифм функции правдоподобия записывается в виде

$$\ln L(\gamma,\mu,\sigma) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \sum_{i=1}^{n}\ln(x_i-\gamma) - \\ - \sum_{i=1}^{n}\left(\frac{(\ln(x_i-\gamma)-\mu)^2}{2\sigma^2}\right), \quad (3.1)$$

причём выражение имеет смысл только при $\gamma < x_{(1)}$. Необходимые условия экстре-

мума, полученные приравниванием частных производных (3.1) к нулю, выглядят следующим образом:

$$\left\{\begin{array}{rcccl} \frac{\partial \ln L}{\partial \gamma} & = & \sum_{i=1}^{n} \frac{1}{x_i-\gamma}\left(1+\frac{\ln(x_i-\gamma)-\mu}{\sigma^2}\right) & = & 0 \\ \frac{\partial \ln L}{\partial \mu} & = & \sum_{i=1}^{n} \frac{\ln(x_i-\gamma)-\mu}{\sigma^2} & = & 0 \\ \frac{\partial \ln L}{\partial \sigma} & = & \sum_{i=1}^{n} \frac{1}{\sigma}\left(-1+\frac{(\ln(x_i-\gamma)-\mu)^2}{\sigma^2}\right) & = & 0 \end{array}\right. \tag{3.2}$$

Метод максимального правдоподобия с успехом применяется во многих задачах статистики. Однако в случае оценки параметров трёхпараметрического логнормального распределения его применимость оказывается под вопросом. В работе [10] показано, что для любой выборки $X$ функция правдоподобия $L(\gamma, \mu, \sigma)$ не ограничена, и существуют траектории в пространстве параметров, сходящиеся к $(x_{(1)}, -\infty, +\infty)$, при движении вдоль которых $L(\gamma, \mu, \sigma)$ сходится к $+\infty$, при том что в самой точке $(x_{(1)}, -\infty, +\infty)$ функция правдоподобия равна $0$.

В работе [10] утверждается, что, несмотря на общую неограниченность функции правдоподобия $L(\gamma, \mu, \sigma)$, если элементы выборки принимают достаточно много различных значений, «вблизи» истинных значений параметров $(\gamma, \mu, \sigma)$ у неё зачастую наблюдается локальный максимум. Это приводит к идее использования так называемых локальных оценок максимального правдоподобия, соответствующих локальному максимуму функции правдоподобия. В работе [11] показано, что такие оценки обладают хорошими асимптотическими свойствами. Для поиска локального максимума предлагается использовать уравнения (3.2) ([12]): через второе и третье уравнения параметры $\mu$ и $\sigma^2$ выражаются как функции параметра $\gamma$

$$\begin{array}{rcl} \mu(\gamma) & = & \frac{1}{n}\sum_{i=1}^{n} \ln(x_i-\gamma), \\ \sigma^2(\gamma) & = & \frac{1}{n}\sum_{i=1}^{n} (\ln(x_i-\gamma)-\mu)^2, \end{array} \tag{3.3}$$

после чего решается уравнение $\lambda(\gamma) = \sum_{i=1}^{n} \frac{1}{x_i-\gamma}\left(1+\frac{\ln(x_i-\gamma)-\mu(\gamma)}{\sigma^2(\gamma)}\right) = 0$ для нахождения оценки локального максимума правдоподобия для параметра $\gamma$, через которую выражаются оценки параметров $\mu$ и $\sigma$.

В работе [13] предлагается модификация описанного выше метода с использования информации, содержащейся в порядковых статистиках выборки, состоящая в замене необходимого условия экстремума $\frac{\partial \ln L}{\partial \gamma} = 0$ в (3.2) на условие равенства математического ожидания $F(X_{(r,n)}; \gamma, \mu, \sigma)$ значению функции распределения в соответствующем элементе вариационного ряда $x_{(r)}$, то есть

$$\mathbb{E}_{X \sim F(\cdot;\gamma,\mu,\sigma)} F(X_{(r,n)}; \gamma, \mu, \sigma) = F(x_{(r)}; \gamma, \mu, \sigma). \tag{3.4}$$

Случайная величина $F(X;\gamma,\mu,\sigma)$ имеет равномерное на $[0,1]$ распределение, а $k$-я порядковая статистика равномерного распределения при длине выборки $n$ имеет бета-распределение с параметрами $(k, n-k+1)$, поэтому левая часть (3.4) равняется $\frac{r}{n+1}$.Так как $F(x;\gamma,\mu,\sigma) = \Phi(\frac{\ln(x-\gamma)-\mu}{\sigma})$, условие (3.4) переписывается в виде $\frac{\log(x_{(r)}-\gamma)-\mu}{\sigma} = \Phi^{-1}\left(\frac{r}{n+1}\right)$. Уравнение $\lambda(\gamma) = 0$ заменяется на $\theta(\gamma) = \log(x_{(r)}-\gamma) - \mu(\gamma) - \Phi^{-1}(\frac{r}{n+1})\sigma(\gamma) = 0$.

## 3.2 Общий метод моментов и его модификации

При оценке параметров с использованием метода моментов на распределение $F(\cdot;\theta)$ накладывается последовательность ограничений типа равенство, образующая систему уравнений вида $g_i(\theta) = h_i(X^n), i = \overline{1,k}$, где функции $g_i(\theta)$ характеризуют теоретическое распределение, а $h_i(X^n)$ являются их выборочными оценками, как правило, несмещёнными или хотя бы асимптотически несмещёнными ([9]).

Общий метод моментов применялся для оценки параметров трёхпараметрического логнормального распределения в связи с указанными выше проблемами, возникающими при использовании метода максимума правдоподобия [8, 13]. В работе [8] в качестве функций $g_i(\gamma,\mu,\sigma), i = 1,2,3$ используются математическое ожидание, дисперсия и коэффициент асимметрии 2.1, в качестве $h_i(X^n), i = 1,2,3$ используются их выборочные оценки [14]:

$$\begin{array}{l} \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \\ a_3 = \frac{m_3}{s^3} = \frac{\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i-\overline{x})^3}{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2)^{\frac{3}{2}}} \end{array} \tag{3.5}$$

Из 2.1 и (3.5) записывается система уравнений:

$$\left\{\begin{array}{rcl} \gamma + \beta\sqrt{\omega} - \overline{x} & = & 0 \\ \beta^2\omega(\omega-1) - s^2 & = & 0 \\ \sqrt{\omega-1}(\omega+2) - a_3 & = & 0 \end{array}\right. \tag{3.6}$$

Третье уравнение можно переписать как $\omega^3+3\omega^2-(4+a_3^2) = 0$. При любых значениях $a_3$ оно имеет действительный корень больше или равный единице, причём единственный, который вычисляется по формуле $\omega = 1 + \left(\sqrt[3]{\frac{(a_3+4)^2+a_3}{2}} - \sqrt[3]{\frac{(a_3+4)^2-a_3}{2}}\right)^2$. Поскольку $\omega = \exp(\sigma^2) > 1$, этот корень принимается в качестве оценки $\exp(\sigma^2)$, после чего оценки для $\sigma = \sqrt{\ln\omega}$, $\mu = \ln\beta$ и $\gamma$ вычисляются аналитически.

Использование показателя третьего порядка в методе моментов приводит к высокой ошибке выборочного наблюдения [8]. Для решения этой проблемы в работе [13] было предложено использовать модификацию метода моментов, основанную на использовании порядковых статистик, сходную с уже продемонстрированной выше модификацией для метода локального максимума правдоподобия. Третье уравнение в (3.6) заменяется на (3.4), или, что то же самое, $x_{(r)} = \gamma + \exp\left(\mu + \sigma\Phi\left(\frac{r}{n+1}\right)\right)$. Тогда система (3.6) превращается в следующую:

$$\begin{cases} \frac{s^2}{(\overline{x}-x_{(r)})^2} = \frac{\omega(\omega-1)}{\left(\sqrt{\omega}-\exp\left(\sqrt{\ln(\omega)}\Phi^{-1}\left(\frac{r}{n+1}\right)\right)\right)^2} = J(r,\omega) \\ \beta = \frac{s}{\sqrt{\omega(\omega-1)}} \\ \gamma = \overline{x} - \frac{s}{\sqrt{\omega-1}} \end{cases} \tag{3.7}$$

Если $\omega$ известно, оценки для $\beta$ и $\gamma$ находятся аналитически. Для решения первого уравнения относительно $\omega$ можно использовать любой метод одномерной оптимизации.

В работе [15] приводится метод оценки параметров распределения трёхпараметрического логнормального распределения с использованием L-моментов. Теоретическим L-моментом порядка $r$ для распределения $F(x)$ называется величина

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\mathbb{E}X_{(r-k,r)}. \tag{3.8}$$

Выборочный L-момент порядка $r \leq n$ определяется как

$$l_r = \binom{n}{r}^{-1}\sum_{1\leq i_1\leq i_2\leq\cdots\leq i_r\leq n}\cdots\sum\frac{1}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}x_{(i_{r-k})}. \tag{3.9}$$

Эти статистики являются несмещёнными [16] оценками теоретических L-моментов. Для трёхпараметрического логнормального распределения можно выписать следующие аналитические формулы $\lambda_1$, $\lambda_2$ и $\tau_3 = \frac{\lambda_3}{\lambda_2}$ [16]:

$$\begin{aligned} \lambda_1 &= \gamma + \exp(\mu + \tfrac{\sigma^2}{2}), \\ \lambda_2 &= \exp(\mu + \tfrac{\sigma^2}{2})\,\mathrm{erf}(\tfrac{\sigma}{2}), \\ \tau_3 &= \frac{6}{\sqrt{\pi}}\frac{\int_0^{\frac{\sigma}{2}}\mathrm{erf}(\frac{x}{\sqrt{3}})\exp(-x^2)dx}{\mathrm{erf}(\frac{\sigma}{2})}. \end{aligned} \tag{3.10}$$

Отсюда можно выписать следующую систему уравнений:

$$\begin{cases} \gamma + \exp(\mu + \frac{\sigma^2}{2}) & = & l_1, \\ \exp(\mu + \frac{\sigma^2}{2}) \operatorname{erf}(\frac{\sigma}{2}) & = & l_2, \\ \frac{6}{\sqrt{\pi}} \frac{\int_0^{\frac{\sigma}{2}} \operatorname{erf}(\frac{x}{\sqrt{3}}) \exp(-x^2) dx}{\operatorname{erf}(\frac{\sigma}{2})} & = & \frac{l_3}{l_2}. \end{cases} \tag{3.11}$$

В работе [15] приводится приближённое решение этой системы:

$$\begin{array}{rcl} z & = & \sqrt{\frac{8}{3}} \Phi^{-1} \left( \frac{1+\frac{l_3}{l_2}}{2} \right), \\ \sigma & \approx & 0,999281z - 0,006118z^3 + 0,000127z^5, \\ \mu & = & \ln \left( \frac{l_2}{\operatorname{erf}(\frac{\sigma}{2})} \right) - \frac{\sigma^2}{2}, \\ \gamma & = & l_1 - \exp(\mu + \frac{\sigma^2}{2}). \end{array} \tag{3.12}$$

## 3.3 Метод минимизации расстояния

В методе минимизации расстояний мерой соответствия модели данным служит некоторым образом выбранное расстояние $d[\cdot, \cdot]$ между теоретическим и эмпирическим распределением данных. Полагается, что чем меньше расстояние, тем лучше модель описывает данные. Для непрерывных распределений расстояние обычно берётся между функцией распределения модели $F(x; \theta)$ и эмпирической функцией распределения $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[x > x_{(i)}]$ ([17]). Следует отметить, что термин «расстояние» используется условно: функционал $d$ может быть даже несимметричен, обычно от него требуется только неотрицательность и равенство нулю только в случае равенства распределений. Оценкой минимального расстояния $\theta_0$ называется $\theta_0 = \arg\min_{\theta \in \Theta} d[F(x; \theta), F_n(x)]$. Одним из наиболее привлекательных свойств оценок минимального расстояния является их робастность, то есть устойчивость к возмущениям в данных ([17]). В настоящей статье рассматриваются следующие функционалы расстояния: Колмогорова — Смирнова, Крамера — фон Мизеса, Андерсона — Дарлинга, Кёйпера и Уотсона. Выражения для них приводятся в таблице 3.1.

Применительно к задаче оценки параметров трёхпараметрического логнормального распределения в работах [6, 18] отмечалось, что методы минимизации расстояния, как правило, оказываются предпочтительнее других методов: они дают более точные оценки параметров, чем другие методы, в частности, метод максимального правдоподобия, и они не страдают от проблем со сходимостью оптимизационной процедуры. Несмотря на эти положительные свойства, тема оценок ми-

| Расстояние | $d[F(x), G(x)]$ |
|---|---|
| Колмогорова — Смирнова | $\sup_x \lvert F(x) - G(x) \rvert$ |
| Крамера — фон Мизеса | $\int_{-\infty}^{+\infty} (F(x) - G(x))^2 \, \mathrm{d}F(x)$ |
| Андерсона — Дарлинга | $\int_{-\infty}^{+\infty} \frac{(F(x)-G(x))^2}{F(x)(1-F(x))} \, \mathrm{d}F(x)$ |
| Кёйпера | $\sup_x (F(x) - G(x)) + \sup_x (G(x) - F(x))$ |
| Уотсона | $\int_{-\infty}^{+\infty} \left( G(x) - F(x) - \int_{-\infty}^{+\infty} (G(y) - F(y)) \, \mathrm{d}F(y) \right)^2 \mathrm{d}F(x)$ |

Таблица 3.1. Использующиеся в работе расстояния [14].

Таблица 3.2. Параметры распределений модельных данных.

| | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\theta_5$ | $\theta_6$ | $\theta_7$ |
|---|---|---|---|---|---|---|---|
| $\gamma$ | 3 | 10 | 16 | 10 | 10 | 10 | 10 |
| $\mu$ | 3 | 3 | 3 | 2 | 4 | 3 | 3 |
| $\sigma$ | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 | 0.1 | 0.35 |

нимального расстояния в применении к оценке параметров трёхпараметрического логнормального распределения достаточно слабо освещена в существующей литературе.

## 3.4 Методы оптимизации функционалов расстояния

Метод минимизации расстояния, в отличие от методов максимального правдоподобия и моментов, подразумевает решение задачи многомерной оптимизации. Зачастую решение оптимизационной задачи не выводится в аналитическом виде, поэтому для нахождения оценок параметров приходится использовать итерационные процедуры многомерной оптимизации. Характеристики метода оптимизации могут сильно разниться от задачи к задаче, поэтому для рассматриваемых в статье функционалов расстояний (таблица 3.1) необходимо подбирать свои методы оптимизации.

Поскольку расстояния Колмогорова — Смирнова и Кёйпера являются негладкими функциями, было решено рассматривать только безградиентные методы оптимизации. Было выбрано четыре метода оптимизации — методы Нелдера — Мида [19] и L-BFGS-B [20] с численной оценкой градиента, реализованные в библиотеке SciPy [21], и методы BOBYQA [22] и PRAXIS [23], реализованные в библиотеке NLopt [24]. Для этого для семи наборов параметров (см. таб. 3.2), выбранных на основании результатов анализа задержек при пересылке сообщений в локальной сети, проведённого в работе [4], было построено по $N = 200$ выборок из логнормального распределения с соответствующими параметрами длиной $N = 1000$. Затем для

каждой из этих выборок производилась оценка параметров методом минимизации каждого из расстояний каждым из методов. Результаты работы оптимизационных процедур для пары «набор параметров — функционал расстояния» сравнивались по следующим характеристикам:

- процент выборок, на которых оптимизационная процедура сошлась;
- достигнутое значение функционала расстояния между восстанавливаемым и эмпирическим распределением, усреднённое по выборкам (приемлемо, поскольку все выборки из одного распределения);
- среднее по всем выборкам время работы оптимизационной процедуры.

В таблице 3.3 указан процент числа выборок для каждого набора параметров, на которых алгоритм сошёлся, для каждого оптимизируемого расстояния и каждого метода оптимизации. На основании этих результатов из рассмотрения можно исключить методы Нелдера — Мида и L-BFGS-B как обладающие слишком плохой сходимостью.

Для методов оптимизации BOBYQA и PRAXIS, показавших стопроцентную сходимость для всех параметров и всех расстояний, исследовались характеристики сходимости — достигаемый минимум и время сходимости. Результаты представлены на рис. 3.1 — 3.5. Из них видно, что вне зависимости от параметров распределений и оптимизируемой функции метод PRAXIS всегда даёт немного лучший результат оптимизации, а метод BOBYQA всегда работает быстрее. Кроме того, можно видеть, что вне зависимости от минимизируемых функционалов расстояний характеристики метода оптимизации примерно одинаково зависят от значений параметров. В частности, для всех функционалов расстояний среднее достигнутое значение было минимально для набора параметров $\theta_4 = (10, 2, 0.23)$ и максимально для набора $\theta_3 = (16, 3, 0.23)$.

## 3.5 Тестирование

С целью сравнения описанных выше методов оценки параметров распределения мы провели их тестирование на модельных данных. Для этого мы взяли два набора параметров $\theta_3$ и $\theta_4$, так как оптимизация функционалов расстояния на них представляет для методов оптимизации наибольшую и наименьшую сложность соответственно. Модельные данные состояли из 100 выборок длиной 500 для каждого набора параметров. Для сравнения оценок, получаемых различными методами, мы использовали следующие характеристики:

| | | 3 | 10 | 16 | 10 | 10 | 10 | 10 | $\gamma$ |
|---|---|---|---|---|---|---|---|---|---|
| | | 3 | 3 | 3 | 2 | 4 | 3 | 3 | $\mu$ |
| | | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 | 0.1 | 0.35 | $\sigma$ |
| Расстояние | Метод оптимизации | | | | | | | | |
| Колмогорова — Смирнова | Нелдера — Мида | 91.5 | 84.5 | 77 | 82.5 | 82.5 | 84 | 90 | |
| | L-BFGS-B | 53 | 58 | 51 | 52 | 54.5 | 52.5 | 52 | |
| | BOBYQA | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| | PRAXIS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| Крамера — фон Мизеса | Нелдера — Мида | 100 | 100 | 99.5 | 100 | 99.5 | 75 | 100 | |
| | L-BFGS-B | 79 | 81 | 77 | 77.5 | 80.5 | 80.5 | 82 | |
| | BOBYQA | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| | PRAXIS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| Андерсона — Дарлинга | Нелдера — Мида | 100 | 100 | 100 | 100 | 1005 | 95 | 100 | |
| | L-BFGS-B | 93 | 91 | 92 | 94 | 91 | 95.5 | 96 | |
| | BOBYQA | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| | PRAXIS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| Кёйпера | Нелдера — Мида | 81.5 | 61 | 54.5 | 65 | 72 | 72.5 | 80.5 | |
| | L-BFGS-B | 45.5 | 40.5 | 43 | 37.5 | 41.5 | 45 | 52 | |
| | BOBYQA | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| | PRAXIS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| Уотсона | Нелдера — Мида | 97 | 96 | 94 | 96 | 96 | 63 | 100 | |
| | L-BFGS-B | 74.5 | 75 | 76.5 | 72 | 75.5 | 80.5 | 71 | |
| | BOBYQA | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |
| | PRAXIS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |

Таблица 3.3. Процент сходимости оптимизационных процедур на модельных данных.

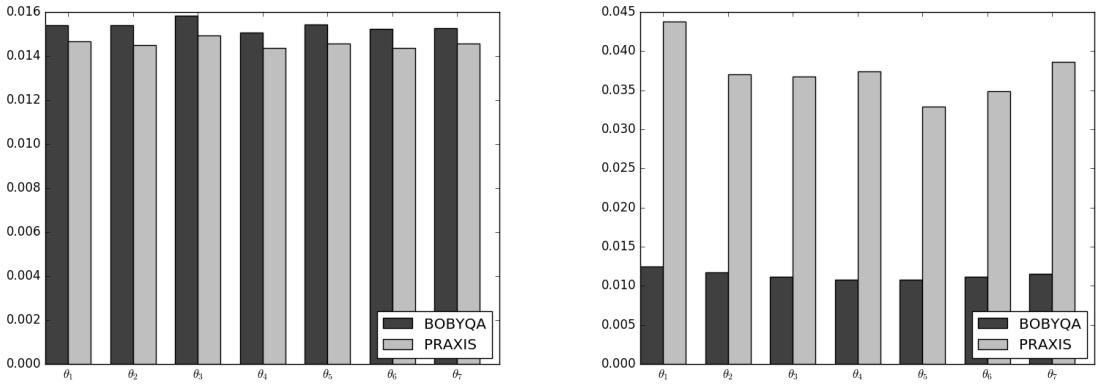



Рис. 3.1. Среднее достигнутое значение функционала (слева) и среднее время работы (справа) методов BOBYQA и PRAXIS для расстояния Колмогорова — Смирнова

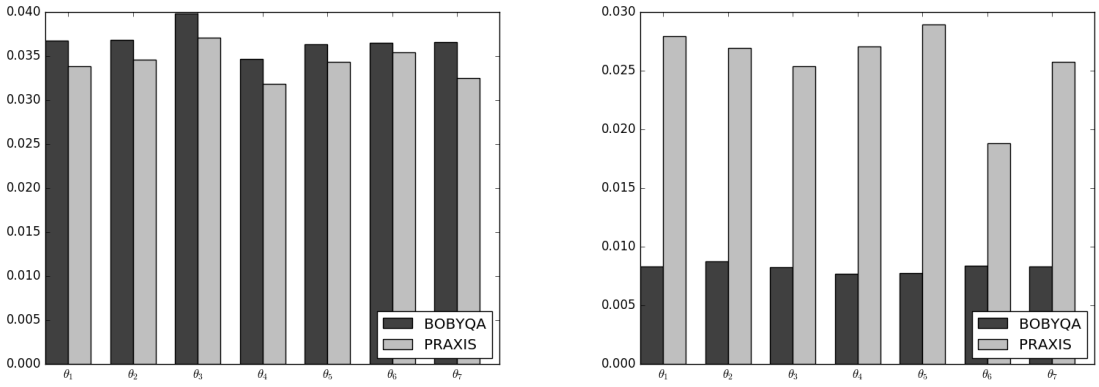



Рис. 3.2. Среднее достигнутое значение функционала и среднее время работы методов BOBYQA и PRAXIS для расстояния Крамера — фон Мизеса

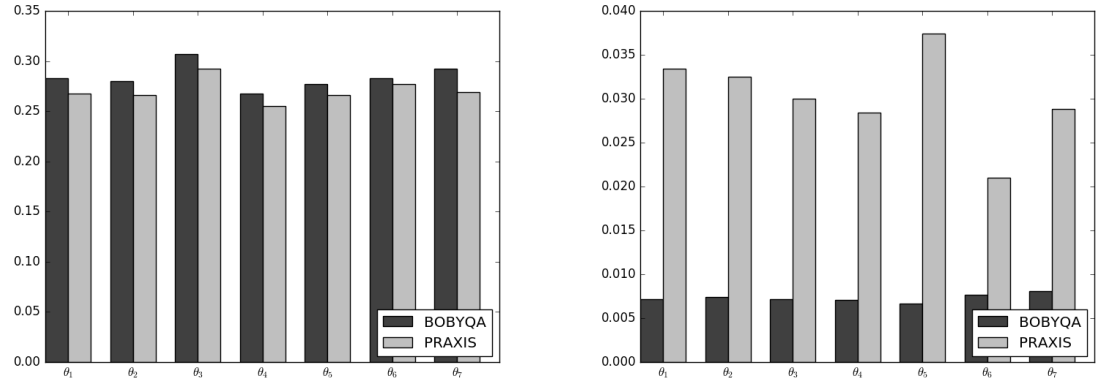



Рис. 3.3. Среднее достигнутое значение функционала и среднее время работы методов BOBYQA и PRAXIS для расстояния Андерсона — Дарлинга

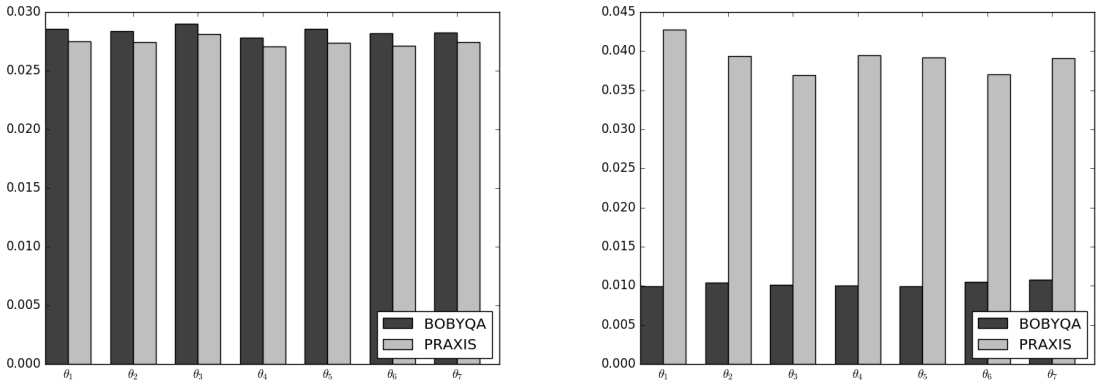



Рис. 3.4. Среднее достигнутое значение функционала и среднее время работы методов BOBYQA и PRAXIS для расстояния Кёйпера

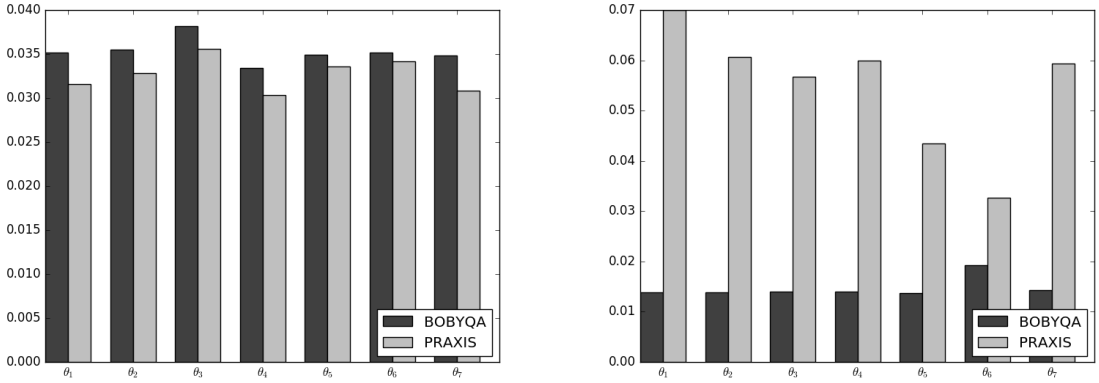



Рис. 3.5. Среднее достигнутое значение функционала и среднее время работы методов BOBYQA и PRAXIS для расстояния Уотсона

1. общая точность (accuracy): $\overline{(\hat{\theta} - \theta)^2}, \theta = \gamma, \mu, \sigma$;
2. среднее время работы метода.

Результаты тестирования приведены на рис. 3.6 и 3.7. По ним можно сделать несколько выводов.

1. Относительное качество почти всех методов (кроме модифицированного метода максимального правдоподобия и модифицированного метода моментов, оценки которых оказались очень неточными) одинаково для каждого из параметров $\gamma$, $\mu$ и $\sigma$. Это значит, что параметры не нужно оценивать различными методами.
2. Есть два лидера — метод максимума правдоподобия и метод L-моментов. Близко к ним стоит метод моментов. Применение методов минимального расстояния (которые использовали оценки, полученные методом моментов, как начальное приближение), как правило, не даёт серьёзного улучшения качества оценки — метод максимума правдоподобия всегда лучше.
3. Методы моментов и L-моментов дают практически такие же по качеству оценки, как и метод максимального правдоподобия, но делают это гораздо быстрее. Методы минимального расстояния оказались чрезвычайно медленными.

## 3.6 Заключение

Были проведены исследования существующих методов оценки параметров 3LN распределения. Было установлено, что, несмотря на отсутствие у функции

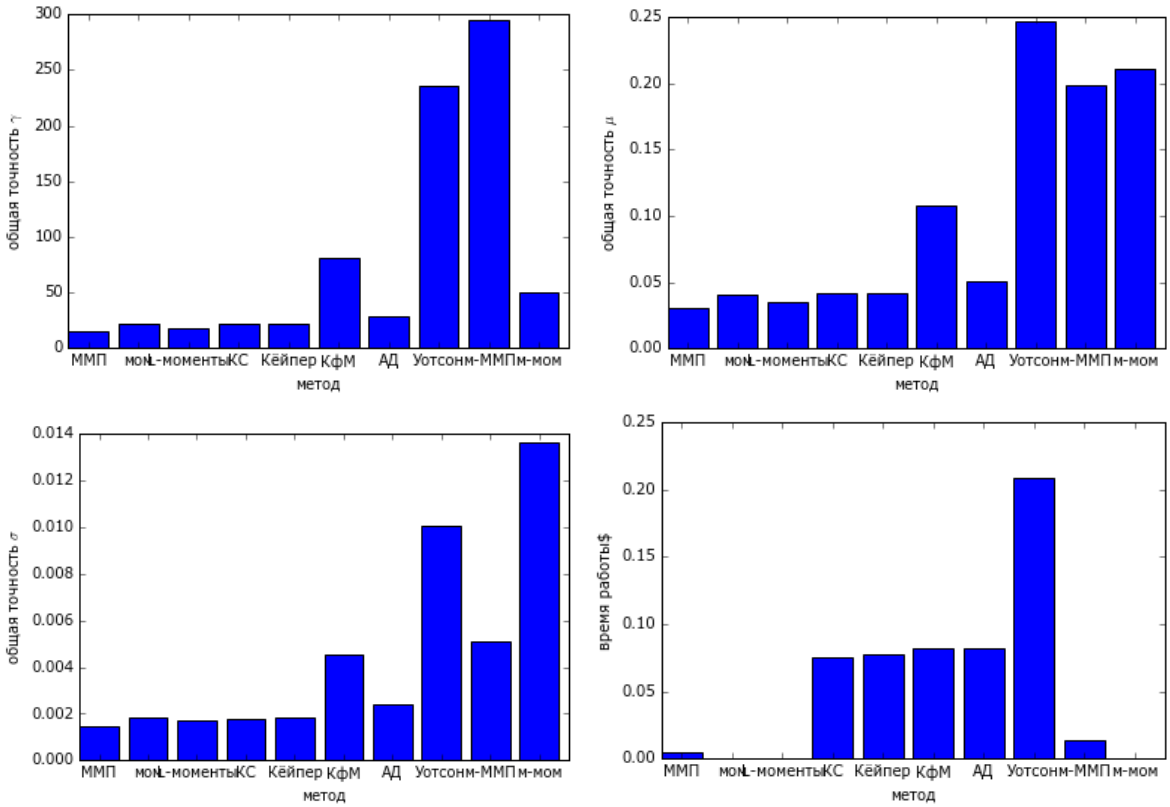



Рис. 3.6. Общая точность и время работы для методов. Набор параметров $\theta_3$.

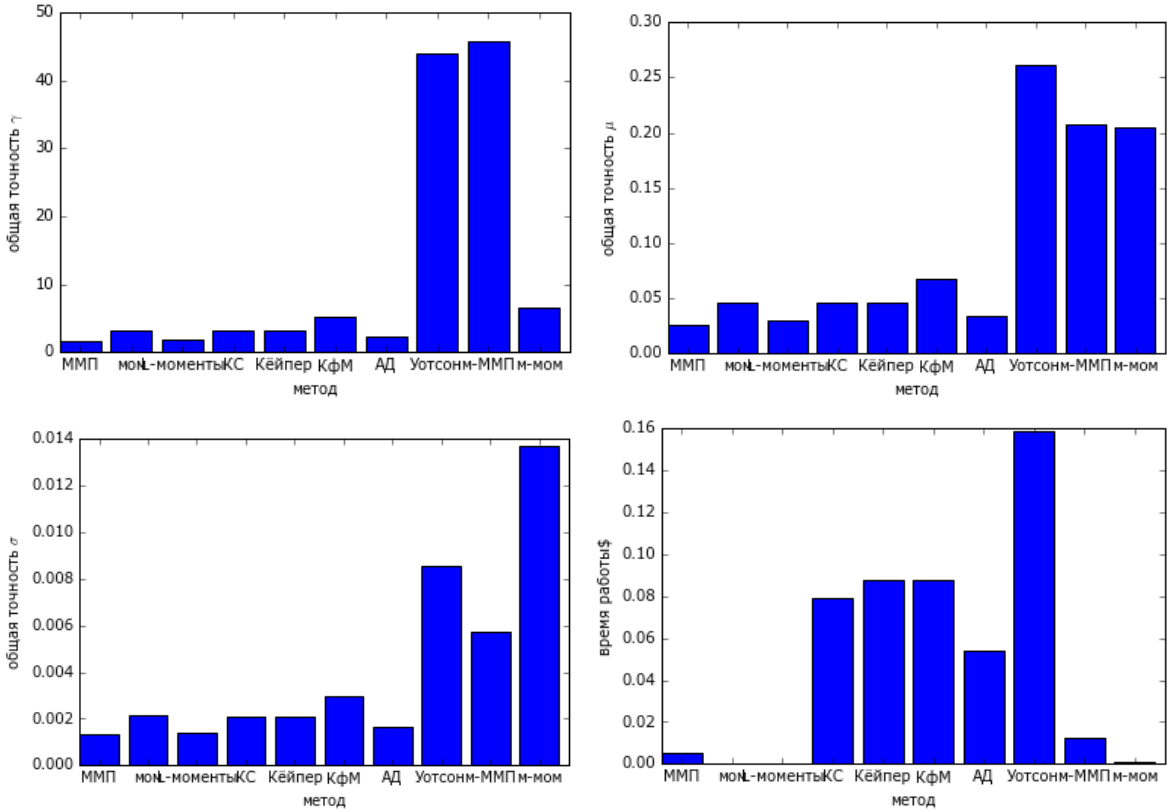



Рис. 3.7. Общая точность и время работы для методов. Набор параметров $\theta_4$.

правдоподобие глобального максимума, оценки максимального правдоподобия оказываются чрезвычайно качественными. Также было обнаружено, что метод моментов и метод L-моментов дают схожие по качестве оценки, но при этом работают быстрее.

# Глава 4

# Разделение смесей

**Определение [25].** Конечной смесью распределений параметрического семейства $\{F(\cdot;\theta)|\theta \in \Theta\}$ называется функция распределения вида $F(x;\overline{w},\overline{\theta}) = \sum_{j=1}^{k} w_j F(x;\theta_j)$, где $0 \leqslant w_j \leqslant 1$, $\sum_{j=1}^{k} w_j = 1$. Распределения $F(\cdot;\theta_j)$ называются компонентами смеси, величины $w_1,\ldots,w_j$ — весами этих компонентов.

Для плотности смеси распределений выполняется аналогичное равенство: $p(x;\overline{w},\overline{\theta}) = \sum_{j=1}^{k} w_j p(x;\theta_j)$, с теми же коэффициентами $w_j$ [25]. Смесь распределений задаёт распределение вероятности с параметрами $\overline{w} = (w_1,\ldots,w_k)$ и $\overline{\theta} = (\theta_1,\ldots,\theta_k)$. Задача параметрического восстановления смеси распределений по выборке носит специальное название — задача разделения смеси.

## 4.1 Идентифицируемость семейства распределений

Следует отметить, что для некоторых семейств распределений одна и та же смесь может задаваться разными наборами весов и параметров компонентов [25]. Семейства распределений, конечные смеси которых взаимно однозначно соответствуют наборам весов и параметров компонентов с точностью до перестановки компонентов и весов, добавления и удаления компонентов с нулевыми весами и перераспределения весов между компонентами с одинаковыми параметрами, называются конечно идентифицируемыми [25]. Если семейство распределений не является конечно идентифицируемым, задача разделения смеси может иметь принципиально не единственное решение.

**Теорема.** Семейство трёхпараметрических логнормальных распределений является конечно идентифицируемым.

*Доказательство.* Семейство распределений конечно идентифицируемо тогда и только тогда, когда множество функций распределения линейно независимо [26]. Пред-

положим, что существуют такие $c_1, \ldots, c_n$, $c_i \neq 0$, $\gamma_1, \ldots, \gamma_n$, $\mu_1, \ldots, \mu_n$, $\sigma_1, \ldots, \sigma_n$, все тройки параметров $(\gamma_i, \mu_i, \sigma_i)$ различны, что

$$\sum_{i=1}^{n} c_i F(x; \gamma_i, \mu_i, \sigma_i) = 0. \tag{4.1}$$

Пусть $\gamma'_1 < \cdots < \gamma'_m$ — все различные значения $\gamma_1, \ldots, \gamma_n$, $m \leqslant n$. Тогда (4.1) можно переписать в виде

$$\sum_{i=1}^{m} \sum_{j:\gamma_j=\gamma'_i} c_j F(x; \gamma'_i, \mu_j, \sigma_j) = 0. \tag{4.2}$$

Рассмотрим $\sum_{j:\gamma_j=\gamma'_1} c_j F(x; \gamma'_1, \mu_j, \sigma_j) = \sum_{\gamma_j=\gamma'_1} c_j \Phi\left(\frac{\ln(x-\gamma'_1)-\mu_j}{\sigma_j}\right)$. Это выражение должно быть равно $0$ на $[\gamma'_1, \gamma'_2]$. Покажем, что это возможно только при равенстве всех $c_j$ в этой сумме $0$. Для этого исследуем определитель Вронского системы функций плотности вероятности нормального распределения $\{\phi(x; \mu_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right), j = \overline{1,k}\}$ на равенство $0$.

Производная функции $\phi(x; \mu, \sigma)$ $i$-го порядка вычисляется по формуле $\phi^{(i)}(x; \mu, \sigma) = \phi(x; \mu, \sigma) s_i(x)$, где

$$\begin{aligned} &s_i(x) = s_{i-1}(x)\left(-\tfrac{x-\mu}{\sigma^2}\right) + s'_{i-1}(x), \\ &s_0(x) = 1. \end{aligned} \tag{4.3}$$

Из (4.3) следует, что $s_i(x) = \sum_{k=0}^{i} a_{ik}(\sigma)\left(-\frac{x-\mu}{\sigma^2}\right)^k$ — многочлен от $\left(-\frac{x-\mu}{\sigma^2}\right)$, $a_{ii}(\sigma) = 1$, $a_{i,i-1}(\sigma) = 0$.

Запишем вронскиан системы функций $\{\phi_j(x) = \phi(x; \mu_j, \sigma_j)\}$:

$$\begin{aligned} W[\phi_1, \ldots, \phi_k](x) = \\ &= \det \left| \phi_j(x) \left( \sum_{l=0}^{i} a_{il}(\sigma_j) \left(-\frac{x-\mu_j}{\sigma_j^2}\right)^l \right) \right|_{j=\overline{1,k}}^{i=\overline{0,k-1}} = \\ &= \prod_{j=1}^{k} \phi_j(x) \cdot \det \left| \sum_{l=0}^{i} a_{il}(\sigma_j) \left(-\frac{x-\mu_j}{\sigma_j^2}\right)^l \right|_{j=\overline{1,k}}^{i=\overline{0,k-1}} = \\ &= \prod_{j=1}^{k} \phi_j(x) \cdot \sum_{\alpha \in S(k)} (-1)^{N(\alpha)} \prod_{i=0}^{k-1} \left( \sum_{l=0}^{i} a_{il}(\sigma_{\alpha_i}) \left(-\frac{x-\mu_{\alpha_i}}{\sigma_{\alpha_i}^2}\right)^l \right), \end{aligned} \tag{4.4}$$

где $S(k)$ — симметрическая группа на $k$ элементах, $N(\alpha)$ — число инверсий в перестановке $\alpha = (\alpha_0, \ldots, \alpha_{k-1})$. Если раскрыть скобки во внутреннем произведении

сумм, получим

$$W(x) = \sum_{\alpha\in S(k)} (-1)^{N(\alpha)} \prod_{i=0}^{k-1} \left( \sum_{l=0}^{i} a_{il}(\sigma_{\alpha_i}) \left( -\frac{x-\mu_{\alpha_i}}{\sigma^2_{\alpha_i}} \right)^l \right) =$$

$$= \sum_{\alpha\in S(k)} (-1)^{N(\alpha)} \sum_{\substack{l_i=0,\dots,i \\ i=\overline{0,k-1}}} \prod_{r=0}^{k-1} a_{rl_r}(\sigma_{\alpha_r}) \left( -\frac{x-\mu_{\alpha_r}}{\sigma^2_{\alpha_r}} \right)^{l_r} \quad (4.5)$$

Поскольку $\prod_{j=1}^{k} \phi_j(x) > 0$, $W[\phi_1, \dots, \phi_k](x) = 0$ тогда и только тогда, когда $W(x) = 0$. Докажем, что $W(x) \not\equiv 0$.

$W(x)$ является многочленом от $x$. Рассмотрим слагаемое $V(x)$, соответствующее $l_i = i, i = \overline{0, k-1}$:

$$V(x) = \sum_{\alpha\in S(k)} (-1)^{N(\alpha)} \prod_{r=0}^{k-1} \left( -\frac{x-\mu_{\alpha_r}}{\sigma^2_{\alpha_r}} \right)^r. \quad (4.6)$$

Это определитель Вандермонда:

$$V(x) = \det \left| \left( -\tfrac{x-\mu_j}{\sigma_j^2} \right)^i \right|_{j=\overline{1,k}}^{i=\overline{0,k-1}} = \prod_{i<j} \left( \frac{x-\mu_i}{\sigma_i^2} - \frac{x-\mu_j}{\sigma_j^2} \right). \quad (4.7)$$

Если все $\sigma_j, j = \overline{1,k}$, различны, то $\deg V(x) = \frac{k(k-1)}{2}$, и это наибольшая степень в $W(x)$. Из этого следует, что $W(x) \not\equiv 0$. Если какие-либо $\sigma_{j_k}$ равны между собой, то степень $V(x)$ будет меньше $\frac{k(k-1)}{2}$, так как часть скобок обращается в константы. Покажем, что в этом случае $V(x)$ также имеет наибольшую степень в $W(x)$.

Пусть $\sigma_{j_1} = \dots = \sigma_{j_s} = \sigma$. Обозначим подгруппу группы $S(k)$ перестановок на $\{j_1, \dots, j_s\}$ как $H$ и множество правых смежных классов $S(k)$ по $H$ как $G$. Любую перестановку $\alpha \in S(k)$ можно записать как $\alpha = \beta\alpha'$, $\beta \in H$, $\alpha' \in G$, причём $\alpha'(j) = \alpha(j)$ при $j \notin \alpha^{-1}(j_1, \dots, j_s)$. Тогда $W(x)$ представимо в виде:

$$W(x) = \sum_{\alpha\in S(k)} (-1)^{N(\alpha)} \sum_{\substack{l_i=0,\dots,i \\ i=\overline{0,k-1}}} \prod_{r=0}^{k-1} a_{rl_r}(\sigma_{\alpha_r}) \left( -\frac{x-\mu_{\alpha_r}}{\sigma^2_{\alpha_r}} \right)^{l_r} =$$

$$= \sum_{\alpha'\in G} (-1)^{N(\alpha')} \sum_{\substack{l_i=0,\dots,i \\ i=\overline{0,k-1}}} \sum_{\beta\in H} (-1)^{N(\beta)} \prod_{r=0}^{k-1} a_{rl_r}(\sigma_{\alpha_r}) \left( -\frac{x-\mu_{\alpha_r}}{\sigma^2_{\alpha_r}} \right)^{l_r} =$$

$$= \sum_{\alpha' \in G} (-1)^{N(\alpha')} \sum_{\substack{l_i = 0,\dots,i \\ i = \overline{0,k-1}}} \prod_{r \notin \alpha^{-1}(j_1,\dots,j_s)} a_{rl_r}(\sigma_{\alpha'_r}) \left( -\frac{x - \mu_{\alpha'_r}}{\sigma^2_{\alpha'_r}} \right)^{l_r} \times$$

$$\times \sum_{\beta \in H} (-1)^{N(\beta)} \prod_{r \in \alpha^{-1}(j_1,\dots,j_s)} a_{rl_r}(\sigma_{(\beta\alpha')_r}) \left( -\frac{x - \mu_{(\beta\alpha')_r}}{\sigma^2_{(\beta\alpha')_r}} \right)^{l_r}. \quad (4.8)$$

Поскольку $\sigma_{j_1} = \dots = \sigma_{j_s} = \sigma$, $\sigma_{(\beta\alpha')_j} = \sigma_{\alpha_j} = \sigma$ при $j \in \alpha^{-1}(j_1, \dots, j_s)$. Отсюда

$$\sum_{\beta \in H} (-1)^{N(\beta)} \prod_{r \in \alpha^{-1}(j_1,\dots,j_s)} a_{rl_r}(\sigma_{(\beta\alpha')_r}) \left( -\frac{x - \mu_{(\beta\alpha')_r}}{\sigma^2_{(\beta\alpha')_r}} \right)^{l_r} =$$

$$= \prod_{r \in \alpha^{-1}(j_1,\dots,j_s)} a_{rl_r}(\sigma) \cdot \sum_{\beta \in H} (-1)^{N(\beta)} \prod_{r \in \alpha^{-1}(j_1,\dots,j_s)} \left( -\frac{x - \mu_{(\beta\alpha')_r}}{\sigma^2} \right)^{l_r} =$$

$$= \prod_{r \in \alpha^{-1}(j_1,\dots,j_s)} \frac{a_{rl_r}(\sigma)}{(-\sigma^2)^{l_r}} \cdot \det \left| (x - \mu_j)^{l_i} \right|^{i \in \alpha^{-1}(j_1,\dots,j_s)}_{j \in \{j_1,\dots,j_s\}}. \quad (4.9)$$

Наконец, преобразуем $\det \left| (x - \mu_j)^{l_i} \right|^{i \in \alpha^{-1}(j_1,\dots,j_s)}_{j \in \{j_1,\dots,j_s\}}$:

$$\det \left| (x - \mu_j)^{l_i} \right|^{i = \overline{1,s}}_{j = \overline{1,s}} =$$

$$= \sum_{\beta \in S(s)} (-1)^{N(\beta)} \prod_{r=1}^{s} (x - \mu_{\beta_r})^{l_r} = \{L = l_1 + \dots + l_s\} =$$

$$= \sum_{\beta \in S(s)} (-1)^{N(\beta)} \sum_{r=0}^{L} (-1)^r \left( \sum_{\substack{0 \leqslant g_1 \leqslant l_1 \\ \dots \\ 0 \leqslant g_s \leqslant l_s \\ g_1 + \dots + g_s = r}} \prod_{t=1}^{s} \mu_{\beta_t}^{g_t} \right) x^{L-r} =$$

$$= \sum_{r=0}^{L} (-1)^r \left( \sum_{\substack{0 \leqslant g_1 \leqslant l_1 \\ \dots \\ 0 \leqslant g_s \leqslant l_s \\ g_1 + \dots + g_s = r}} \sum_{\beta \in S(s)} (-1)^{N(\beta)} \prod_{t=1}^{s} \mu_{\beta_t}^{g_t} \right) x^{L-r} =$$

$$= \sum_{r=0}^{L} (-1)^r \left( \sum_{\substack{0 \leqslant g_1 \leqslant l_1 \\ \dots \\ 0 \leqslant g_s \leqslant l_s \\ g_1 + \dots + g_s = r}} \det \left| \mu_j^{g_i} \right|^{i = \overline{1,s}}_{j = \overline{1,s}} \right) x^{L-r} \quad (4.10)$$

Если для каких-либо $u \neq v$ $g_u = g_v$, то $\det \left|\mu_j^{g_i}\right|_{j=\overline{1,s}}^{i=\overline{1,s}} = 0$, так как в матрице есть две совпадающие строки. Рассмотрим случай, когда все $g_i, i = \overline{1,s}$, различны. Без ограничения общности можно считать, что $g_1 < \cdots < g_s$. Но тогда $g_i \geqslant i-1, i = \overline{1,s}$, а значит, $g_1 + \cdots + g_s \geqslant \frac{s(s-1)}{2}$. Это значит, что при $0 \leqslant r < \frac{s(s-1)}{2}$ $\sum\limits_{\substack{0\leqslant g_1\leqslant l_1\\ \cdots\\ 0\leqslant g_s\leqslant l_s\\ g_1+\cdots+g_s=r}} \det \left|\mu_j^{g_i}\right|_{j=\overline{1,s}}^{i=\overline{1,s}} = 0$, то есть $\deg \det \left|(x-\mu_j)^{l_i}\right|_{j=\overline{1,s}}^{i=\overline{1,s}} \leq L - \frac{s(s-1)}{2}$. Это значит, что если у $s$ функций плотности $\phi_j(x)$ равны параметры $\sigma$, то степень каждого слагаемого в $W(x)$ (4.8) уменьшается по меньшей мере на $\frac{s(s-1)}{2}$. Если таких групп равных параметров $\sigma$ несколько, то каждая из них будет независимо снижать степень слагаемых в $W(x)$ указанным образом. Таким образом, $V(x)$ — всегда старший член, и, поскольку он не равен $0$, $W(x) = 0$ только в конечном числе точек (возможно, вообще нигде). Из этого следует, что $W[\phi_1, \ldots, \phi_k](x) \not\equiv 0$ ни на каком отрезке, значит, система $\{\phi_1(x), \ldots, \phi_k(x)\}$ не является линейно зависимой ни на каком отрезке, значит, система $\{\Phi_1(x), \ldots, \Phi_k(x)\}$ не является линейно зависимой ни на каком отрезке, поэтому $\{\Phi_1(\ln(x-\gamma)), \ldots, \Phi_k(\ln(x-\gamma))\}$ не является линейно зависимой ни на каком отрезке, значит, не существует $c_1, \ldots, c_k$, $\sum_{i=1}^k c_i^2 > 0$: $\sum_{j:\gamma_j=\gamma_1'} c_j F(x; \gamma_1', \mu_j, \sigma_j) = 0$. Таким образом, все $c_j$, соответствующие $\gamma_j = \gamma_1'$, равны $0$, и соответствующие компоненты, а значит, и $\gamma_1'$, можно не рассматривать. То же самое можно сказать про $\gamma_2', \ldots, \gamma_m'$, и получается, что $c_1 = \cdots = c_n = 0$. $\square$

## 4.2 Метод максимума правдоподобия. EM-алгоритм.

Логарифм функции правдоподобия параметров $\overline{w}, \overline{\theta}$ имеет вид

$$\ln L(w, \gamma, \mu, \sigma) = \sum_{i=1}^{n} \ln \sum_{j=1}^{k} w_k F(x_i, \theta_j). \tag{4.11}$$

Непосредственный поиск точки максимума этой функции весьма затруднителен, поэтому для решения задачи разделения смеси методом максимума правдоподобия зачастую используется EM-алгоритм. Каждая итерация EM-алгоритма записывается в виде [27]:

$$w_j^{m+1} = \frac{\sum_{i=1}^{n} q_{ij}}{n}, j = \overline{1,k} \tag{4.12}$$

$$\overline{\theta}^{m+1} = \arg\max_{\theta} \sum_{j=1}^{k} \sum_{i=1}^{n} q_{ij} \ln p(x_i; \theta_j) \tag{4.13}$$

где

$$q_{ij} = p(x_i \sim F(\cdot;\theta_j)|\overline{w}^m, \overline{\theta}^m) = \frac{w_j^m p(x_i;\theta_j^m)}{\sum_{l=1}^k w_l^m p(x_i;\theta_l^m)}. \tag{4.14}$$

Так как параметры компонентов смеси независимы, соотношение (4.13) можно представить в виде

$$\theta_j^{m+1} = \arg\max_\theta \sum_{i=1}^n q_{ij} \ln p(x_i;\theta), j \in \overline{1,k}. \tag{4.15}$$

Параметры каждого компонента смеси настраиваются независимо, и связь между ними осуществляется только через $q_{ij}$. В силу этого в дальнейшем индекс компонента смеси будет опускаться, если в нём нет необходимости.

Задачу оптимизации (4.15) можно рассматривать как задачу оценки параметров максимального правдоподобия для взвешенной выборки с набором весов $q$. Как было сказано выше, функция правдоподобия для трёхпараметрического логнормального распределения всегда является неограниченной сверху, и того же можно ожидать и для случая взвешенной выборки. Это приводит к необходимости использования иных методов оценки параметров компонентов распределения на каждой итерации EM-алгоритма. Для этого можно использовать обобщения методов, описанных в предыдущем разделе, на случай взвешенной выборки.

Метод локального максимума правдоподобия для случая взвешенной выборки можно получить из необходимых условий экстремума для (4.15):

$$\left\{\begin{array}{rcccl} \frac{\partial \ln L}{\partial \gamma} & = & \sum_{i=1}^n q_i \frac{1}{x_i-\gamma}\left(1+\frac{\ln(x_i-\gamma)-\mu}{\sigma^2}\right) & = & 0 \\ \frac{\partial \ln L}{\partial \mu} & = & \sum_{i=1}^n q_i \frac{\ln(x_i-\gamma)-\mu}{\sigma^2} & = & 0 \\ \frac{\partial \ln L}{\partial \sigma} & = & \sum_{i=1}^n q_i \frac{1}{\sigma}\left(-1+\frac{(\ln(x_i-\gamma)-\mu)^2}{\sigma^2}\right) & = & 0 \end{array}\right. \tag{4.16}$$

Из второго и третьего уравнений выводятся аналитические формулы для оценок $\mu$ и $\sigma$:

$$\begin{array}{rcl} \mu(\gamma;q) & = & \frac{\sum_{i=1}^n q_i \ln(x_i-\gamma}{\sum_{i=1}^n q_i}, \\ \sigma^2(\gamma;q) & = & \frac{\sum_{i=1}^n q_i (\ln(x_i-\gamma)-\mu)^2}{\sum_{i=1}^n q_i}, \end{array} \tag{4.17}$$

а из первого — условие на $\gamma$:

$$\lambda(\gamma;q) = \sum_{i=1}^n \frac{q_i}{x_i-\gamma}\left(1+\frac{\ln(x_i-\gamma)-\mu(\gamma;q)}{\sigma^2(\gamma;q)}\right). \tag{4.18}$$

Метод моментов для случая взвешенной выборки можно получить, используя взвешенные аналоги выборочных моментов. Для выборочного среднего, выборочной дисперсии и выборочного коэффициента асимметрии используются следующие

формулы [28]:

$$\begin{aligned} \overline{x} &= \frac{\sum_{i=1}^n q_i x_i}{\sum_{i=1}^n q_i} \\ s^2 &= \frac{\sum_{i=1}^n q_i}{\left(\sum_{i=1}^n q_i\right)^2 - \sum_{i=1}^n q_i^2} \sum_{i=1}^n q_i (x_i - \overline{x})^2 \\ a_3 &= \frac{m_3}{s^3} = \frac{\frac{\left(\sum_{i=1}^n q_i\right)^2}{\left(\sum_{i=1}^n q_i\right)^3 - 3\sum_{i=1}^n q_i \sum_{i=1}^n q_i^2 + 2\sum_{i=1}^n q_i^3} \sum_{i=1}^n q_i (x_i - \overline{x})^3}{(s^2)^{\frac{3}{2}}} \end{aligned} \tag{4.19}$$

Метод L-моментов для взвешенной выборки можно получить, если рассматривать веса элементов $q_i$ как количества их вхождений в выборку. Для этого домножим их на $\frac{n}{\sum_{i=1}^n q_i}$, чтобы сумма была равна числу элементов в выборке. Заметим, что эта операция не повлияла бы на положение точки максимума взвешенной функции правдоподобия. Тогда можно рассмотреть обобщение выборочных L-моментов (3.9) для взвешенной выборки:

$$l_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_r \leq n} \sum \cdots \sum \frac{1}{r} \left( \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{(i_{r-k})} \right) \prod_{k=1}^{r} q_{i_k}. \tag{4.20}$$

Если бы $X$ была выборкой с повторами, и каждое значение повторялось бы в ней $q_i$ раз, ровно так и выглядело бы выражение для $l_r$. Можно показать, что, если ввести обозначения $v_i = \sum_{j \leqslant i} q_j$, $w_i = \sum_{j \geqslant i} q_j$, $t_i = v_{i-1} - w_{i+1}$, $l_1$, $l_2$ и $l_3$ можно выразить в виде

$$\begin{aligned} l_1 &= \frac{1}{n} \sum_{i=1}^n x_i q_i \\ l_2 &= \frac{1}{n(n-1)} x_i q_i t_i \\ l_3 &= \frac{2}{n(n-1)(n-2)} \sum_{i=1}^n x_i q_i \Big( t_i v_{i-2} - \sum_{j<i-1} q_j v_j - \\ &\quad - t_i w_{i+2} - \sum_{j>i+1} q_j w_j - q_{i-1} w_{i+1} - q_{i+1} v_{i-1} \Big) \end{aligned} \tag{4.21}$$

Здесь стоит особо отметить, что все эти оценки получаются за линейное от размера выборки время.

Методы минимизации расстояния между теоретической и эмпирической функциями распределения обобщаются на случай взвешенной выборки одним и тем же образом — посредством использования взвешенной выборочной функции распределения $F_n(x) = \frac{\sum_{i=1}^n q_i \mathbb{I}[x > x_i]}{\sum_{i=1}^n q_i}$.

В работе [29] предлагается ещё одна модификация EM-алгоритма для разделения смеси — так называемый «усечённый», или «сеточный», EM-алгоритм. Вместо

смеси вида $p(x;\overline{w},\overline{\theta}) = \sum_{j=1}^{k} w_j p(x;\theta_j)$ и задачи оптимизации

$$\max_{\overline{w},\overline{\theta}} \sum_{i=1}^{n} \ln p(x_i;\overline{w},\overline{\theta}) \tag{4.22}$$

рассматривается смесь вида $p(x;\overline{w}',\overline{\theta}') = \sum_{j=1}^{K} w'_j p(x;\theta'_j)$ с фиксированными параметрами компонентов $\theta'_j \in T$ и задача оптимизации

$$\max_{\overline{w}'} \sum_{i=1}^{n} \ln p(x_i;\overline{w}',\overline{\theta}'). \tag{4.23}$$

Такая задача имеет сходство с задачей кого-то там [30]. Функция $\ln L(w') = \sum_{i=1}^{n} \ln p(x_i;\overline{w}',\overline{\theta}')$ носит название сеточной, или усечённой, функции правдоподобия. Для решения задачи (4.23) предлагается использовать итерационный процесс, аналогичный соотношению (4.12):

$$q_{ij} = \frac{w'^m_j p(x_i;\theta'^m_j)}{\sum_{l=1}^{K} w'^m_l p(x_i;\theta'^m_l)} \tag{4.24}$$

$$w'^{m+1}_j = \frac{\sum_{i=1}^{n} q_{ij}}{n}, j = \overline{1,K} \tag{4.25}$$

Поскольку EM-алгоритм является монотонным [27], а оптимизируемая функция — вогнутой [29], такая оптимизационная процедура сходится к глобальному максимуму сеточной функции правдоподобия. Если элементы множества параметров $T$ лежат близко к истинным параметрам смеси, то можно рассчитывать на высокую точность аппроксимации истинного распределения данных полученной смесью с фиксированными компонентами. В частности, если в качестве множества $T$ взять сетку в пространстве параметров $\Theta$, то при достаточно маленьком шаге можно ожидать, что оптимизационная процедура сойдётся примерно к тем же компонентам и тем же весам, что образуют истинное распределение.

## 4.3 Метод минимизации расстояния

Для решения задачи разделения смеси также можно использовать методы минимизации расстояния, описанные в подразделе 3.4. Задача оптимизации при этом имеет следующие отличия от случая одного распределения:

1. задача оптимизации является условной с ограничениями типа равенство, поскольку сумма весов должна быть равна 1;

Таблица 4.1. Время работы алгоритмов

| Метод | EM | netEM | EM-мом. | EM-L-мом. | КС | КрфМ | АД | Кйп | Уот |
|---|---|---|---|---|---|---|---|---|---|
| Время, с | 0.5 | 41.19 | 0.26 | 0.14 | 5.05 | 20.57 | 27.9 | 3.97 | 16.53 |

2. размерность задачи непостоянна: количество компонент в смеси равняется $k$, то количество параметров распределения и, соответственно, размерность задачи— $4k$.

## 4.4 Тестирование

Тестирование методов проводилось на данных длиной 10000 о задержках при пересылке сообщений длиной 5000 байт от фиксированного отправителя фиксированному получателю на системе BlueGene/P. На каждом наборе данных запускались EM-алгоритм, его модификации с заменой метода оптимизации параметра, сеточный EM-алгоритм и все методы оптимизации расстояния. Результаты для одной из выборок приведены на рис. 4.1, 4.2 и в таб. 4.1. По результатам анализа можно сделать следующие выводы:

1. методы минимизации расстояний слишком медленные для оценки моделей;
2. сеточные методы дают слишком громоздкую модель, к тому же могут переобучаться;
3. модификации EM-алгоритма дают приемлемое решение, причём очень быстро.

## 4.5 Заключение

В главе проведён анализ методов разделения смесей 3LN. Выбраны методы, которые будут использоваться для построения модели при анализе выполнения обобщённых метрических характеристик. Также доказана теорема об идентифицируемости семейства распределений 3LN.

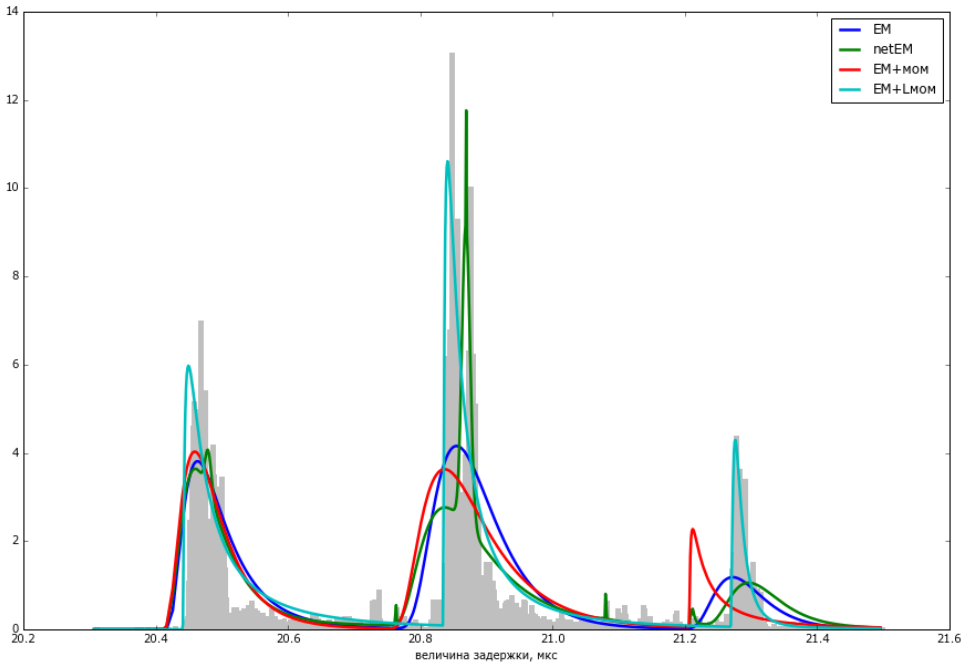



Рис. 4.1. Плотность смеси распределений, восстановленная по данным EM-алгоритмом и его модификациями

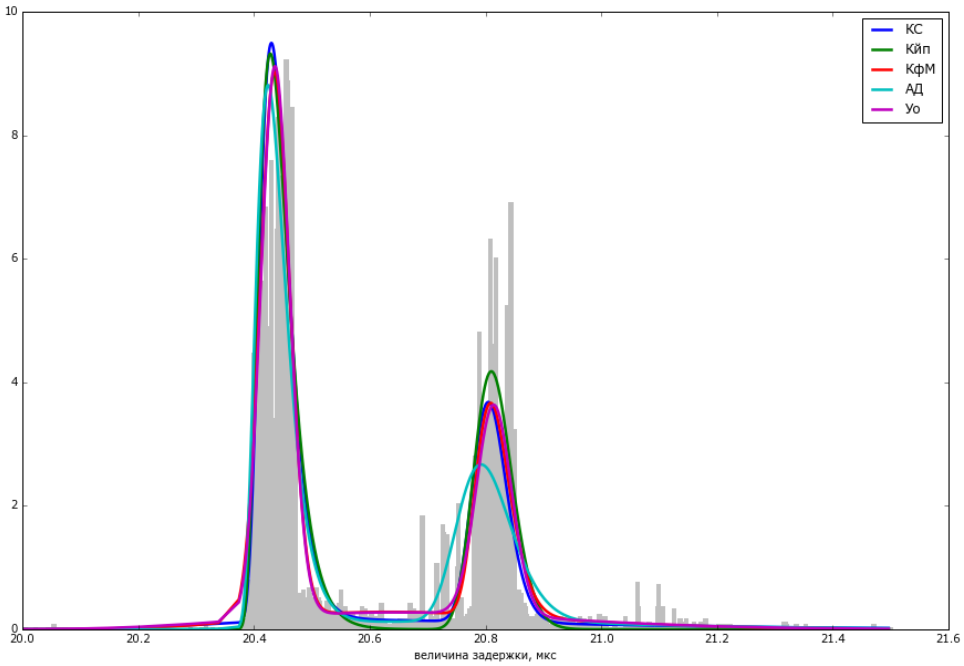



Рис. 4.2. Плотность смеси распределений, восстановленная по данным методами минимизации расстояний

# Глава 5

# Обобщённые метрические характеристики кластера

Как уже было сказано в главе 2, проверка выполнения метрических характеристик формализуется как проверка стохастического доминирования. Задачей работы является построение модели задержек для проверки стохастического доминирования, причём модель должна как можно лучше сохранять информацию о полной выборке. Проверка стохастического доминирования с использованием полных выборок формализуется как проверка статистической гипотезы.

## 5.1 Проверка гипотезы о доминировании

Для проверки стохастического доминирования в настоящей работе использовался критерий Баррета — Дональда [31]. Нулевой гипотезой служит наличие доминирования первого порядка величины $X$ над величиной $Y$, альтернативой — его отсутствие:

$$\begin{aligned} H_0: &\quad F_X(z) \leq F_Y(z) \forall z \\ H_1: &\quad \exists z : F_X(z) > F_Y(z) \end{aligned} \tag{5.1}$$

Статистика критерия имеет вид

$$T(X^N, Y^M) = \sqrt{\frac{NM}{N+M}} \sup_z (F_{XN}(z) - F_{XM}(z)), \tag{5.2}$$

где $M, N$ — величины выборок, $F_{XN}(z) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}[z > x_{(i)}]$, $F_{YM}(z) = \frac{1}{M} \sum_{j=1}^{M} \mathbb{I}[z > y_{(j)}]$ — эмпирические функции распределения выборок $X^N$ и $Y^M$ соответственно. Следует заметить, что, поскольку эмпирические функции распределения кусочно-постоянны, величину $(F_{XN}(z) - F_{XM}(z))$ достаточно вычислять в $z \in X^N \cup Y^M$.

Для аппроксимации распределения статистики (5.2) используется перестановочный метод. Из совокупности $X^N \cup Y^M$ $R$ раз генерируются выборки $\hat{X}_i^N$ и $\hat{Y}_i^M$, $i = \overline{1, R}$, и $p$-значение критерия приближённо вычисляется как

$$p = \frac{1}{R} \sum_{i=1}^{R} \mathbb{I}[T(\hat{X}_i^N, \hat{Y}_i^M) > T(X^N, Y^M)] \tag{5.3}$$

## 5.2 Использование модели задержек

Проверка стохастического доминирования с использованием модели задержек — это проверка выполнения неравенства между функциями распределения, задающимися соответствующей моделью. Для этого использовалось сравнение функций в точках сетки: $X$ доминирует $Y$, если $F_X(z) - F_Y(z) < \varepsilon \ \forall z = z_1, \ldots, z_k$. В работе использовалась равномерная сетка из 21 узла с крайними значениями, соответствующими максимуму и минимуму соответствующих выборок; $\varepsilon = 0.05$.

В случае проверки монотонности функции распределения представляют собой обычные смеси 3LN распределений; при проверке неравенства треугольника и неделимости функция распределения суммы задержек вычислялась по формуле свёртки $F_{X+Y}(z) = \int_{-\infty}^{+\infty} F_Y(z-t) p_X(t)\,\mathrm{d}t$. Заметим, что, поскольку 3LN распределение имеет минимальное значение, пределы интегрирования можно брать конечными.

## 5.3 Тестирование

Проверка выполнения метрических характеристик проводилась для суперкомпьютеров Ломоносов и BlueGene/P. Для выборок задержек сначала проводилась проверка статистической гипотезы о доминировании. После этого строились модели этих задержек, и решался вопрос о неравенстве соответствующих функций распределения. Для построения моделей использовался EM-алгоритм, а также его модификации с использованием моментов и L-моментов: как показало исследование, проведённое в главе 4, они достаточно точные и очень быстрые. Результаты сравнивались между собой. Ответ, полученный проверкой статистических гипотез, считался истинным. Точность проверки с использованием моделей приводится в таб. 5.1 и 5.3, время работы методов приведено в таб. 5.2 и 5.4. Видно, что результаты получаются достаточно точными (для суперкомпьютера BlueGene/P результаты проверки через полные выборки и через модели полностью совпали), причём методы проверки с использованием моделей в разы, а то и в десятки раз быстрее.

Таблица 5.1. Точность проверки с использованием моделей, с/к Ломоносов

| | Монотонность | Нерав-во треуг. | Неделимость |
|---|---|---|---|
| EM-алгоритм | 84% | 95% | 99% |
| EM + моменты | 81% | 97% | 96% |
| EM + L-моменты | 82% | 97% | 99% |

Таблица 5.2. Среднее время работы, секунды, с/к Ломоносов

| | Монотонность | Нерав-во треуг. | Неделимость |
|---|---|---|---|
| EM-алгоритм | 2.17 | 7.12 | 6.52 |
| EM + моменты | 2.19 | 7.75 | 6.58 |
| EM + L-моменты | 1.2 | 6.75 | 8.72 |
| Проверка гипотезы | 11.29 | 41.26 | 34.77 |

Таблица 5.3. Точность проверки с использованием моделей, с/к BlueGene/P

| | Монотонность | Нерав-во треуг. | Неделимость |
|---|---|---|---|
| EM-алгоритм | 100% | 100% | 100% |
| EM + моменты | 100% | 100% | 100% |
| EM + L-моменты | 100% | 100% | 100% |

Таблица 5.4. Среднее время работы, секунды, с/к BlueGene/P

| | Монотонность | Нерав-во треуг. | Неделимость |
|---|---|---|---|
| EM-алгоритм | 2.85 | 5.0 | 5.95 |
| EM + моменты | 1.34 | 3.81 | 3.79 |
| EM + L-моменты | 1.09 | 2.60 | 2.83 |
| Проверка гипотезы | 45.3 | 121.73 | 169.16 |

## 5.4 Выводы

По результатам работы показано, что моделирование задержек даёт достаточно точный и очень быстрый способ проверки выполнения обобщённых метрических условий.

# Глава 6

# Заключение

В работе рассмотрен способ моделирования задержек в коммуникационной сети вычислительного кластера для решения задач тестирования и динамического планирования. Предложены и исследованы несколько способов оценки параметров модели. С использованием рассматриваемой модели задержек проведена проверка выполнения обобщённых метрических характеристик коммуникационной сети многопроцессорных систем «Ломоносов» и BlueGene/P. Было показано, что использование модели позволяет при незначительных потерях качества значительно ускорить процесс тестирования по сравнению с использованием полных выборок.

# Список литературы


[1] Message Passing Forum. *MPI: A Message-Passing Interface Standard*. Тех. отч. Knoxville, TN, USA, 1994.

[2] В. В. Воеводин и Вл. В. Воеводин. *Параллельные вычисления*. СПб.: БХВ-Петербург, 2002.

[3] Andrew Corlett, DI Pullin и Stephen Sargood. «Statistics of one-way internet packet delays». В: *53 rd IETF* (2002).

[4] Mehmet Karakaş. «Determination of Network Delay Distribution over the Internet». Дис. ... док. Citeseer, 2003.

[5] Amarnath Mukherjee. «On the dynamics and significance of low frequency components of Internet load». В: *Technical Reports (CIS)* (1992), с. 300.

[6] A. Gorelov, A. Maysuradze и A. Salnikov. «Delay structure mining in computing cluster». Англ. В: *CEUR Workshop Proceedings*. Т. 1482. Aachen : M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen Germany Germany, 2015, с. 546—551.

[7] Norman L Johnson, Samuel Kotz и N Balakrishnan. «Continuous univariate distributions». В: т. 1. John Wiley & Sons, Inc., 1994. Гл. 14: Lognormal Distributions.

[8] A Clifford Cohen и Betty Jones Whitten. *Parameter estimation in reliability and life span models*. M. Dekker, 1988.

[9] С.А. Айвазян и В.С. Мхитарян. «Прикладная статистика. Основы эконометрики». В: т. 1: Теория вероятностей и прикладная статистика. ЮНИТИ-ДАНА, 2001. Гл. 7. Статистическое оценивание параметров, с. 231—281.

[10] Bruce M Hill. «The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic». В: *Journal of the American Statistical Association* 58.301 (1963), с. 72—84.

[11] H Leon Harter и Albert H Moore. «Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples». В: *Journal of the American statistical association* 61.315 (1966), с. 842—851.

[12] Fred Calitz. «Maximum likelihood estimation of the parameters of the three-parameter lognormal distribution — a reconsideration». В: *Australian Journal of Statistics* 15.3 (1973), с. 185—190.

[13] A Clifford Cohen и Betty Jones Whitten. «Estimation in the three-parameter lognormal distribution». В: *Journal of the American Statistical Association* 75.370 (1980), с. 399—404.

[14] А.И. Кобзарь. «Прикладная математическая статистика». В: М.: Физматлит, 2006. Гл. 3. Методы анализа законов распределения вероятностей случайных величин. С. 204—230.

[15] Diana Bílková. «Three-parametric lognormal distribution and estimating its parameters using the method of L-moments». В: *Reprodukce lidského kapitálu*. Дек. 2011.

[16] J. R. M. Hosking. «L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics». В: *Journal of the Royal Statistical Society. Series B (Methodological)* 52.1 (1990), с. 105—124. ISSN: 00359246. URL: `http://www.jstor.org/stable/2345653`.

[17] Ayanendranath Basu, Hiroyuki Shioya и Chanseok Park. *Statistical inference: the minimum distance approach*. CRC Press, 2011.

[18] JH Keffer. *Robust Minimum Distance Estimation of the Three Parameter Lognormal Distribution.* Тех. отч. DTIC Document, 1983.

[19] John A Nelder и Roger Mead. «A simplex method for function minimization». В: *The computer journal* 7.4 (1965), с. 308—313.

[20] Richard H Byrd и др. «A limited memory algorithm for bound constrained optimization». В: *SIAM Journal on Scientific Computing* 16.5 (1995), с. 1190—1208.

[21] Eric Jones, Travis Oliphant, Pearu Peterson и др. *SciPy: Open source scientific tools for Python*. [Online; accessed 2016-10-21]. 2001–. URL: `http://www.scipy.org/`.

[22] Michael JD Powell. «The BOBYQA algorithm for bound constrained optimization without derivatives». В: (2009).

[23] Richard P. Brent. *Algorithms for Minimization without Derivatives*. Dover Publications, Inc., 2002.

[24] Steven G. Johnson. *The NLopt nonlinear-optimization package*. [Online; accessed 2016-10-21]. URL: `http://ab-initio.mit.edu/nlopt`.

[25] С.А. Айвазян и др. «Прикладная статистика. Классификации и снижение размерности». В: М.: Финансы и статистика, 1989. Гл. 7. Статистическое оценивание параметров, с. 182—217.

[26] Sidney J Yakowitz и John D Spragins. «On the identifiability of finite mixtures». В: *The Annals of Mathematical Statistics* (1968), с. 209—214.

[27] В. Ю. Королев. *ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор*. ИПИ РАН Москва, 2007, с. 94.

[28] Lorenzo Rimoldini. «Weighted skewness and kurtosis unbiased by sample size and Gaussian uncertainties». В: *Astronomy and Computing* 5 (2014), с. 1—8.

[29] Виктор Юрьевич Королев и А Л Назаров. «Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия». В: *Автоматика и телемеханика* 3 (2010), с. 98—116.

[30] М.И. Шлезингер и Вацлав Главач. *Десять лекций по статистическому и структурному распознаванию*. Т. 546. 2004.

[31] Garry F Barrett и Stephen G Donald. «Consistent tests for stochastic dominance». В: *Econometrica* 71.1 (2003), с. 71—104.