

Введение

Сегодня в условиях постоянно растущей конкуренции между производителями продукции происходит непрерывный рост требований к качеству и достоверности обработки и интерпретации технологической информации, используемой для управления производственными процессами (проблема Statistical Quality Control). Во многих случаях задача обработки технологической информации относится к задачам оценивания функций распределений (ФР) и плотностей вероятностей (ПВ) некоторых случайных выборок данных (случайных последовательностей, СП), для решения которых разработано большое число различных параметрических и непараметрических методов.

Анализ технических систем, используемых в различных отраслях промышленности (угледобывающей, деревообрабатывающей, нефте- и газотранспортной, связи, приборостроительной и др.), показывает, что порождаемая ими технологическая информация с математической точки зрения представляет собой случайные последовательности с ограниченной областью рассеяния (СПООР) — например, время работоспособности приборов, диаметры бревен, концентрации углерода в углях, число пакетов, передаваемых в компьютерных сетях в течение выбранного временного интервала и др.

Приведем некоторые примеры подобных задач.

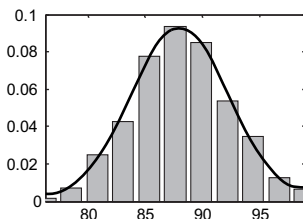
1. Определение параметров распределения процентного содержания углерода в коксующихся углях при торгово-промышленной маркировке угля, которое, исходя из объективных оценок, представляет собой случайную величину (СВ), область рассеяния которой представляет собой отрезок $[70; 100]$. Указанное свойство процентного содержания углерода в коксующихся углях обусловлено тем, что на каждом месторождении угли, добытые на его разных участках, имеют разную степень углефикации. В результате при добыче, отгрузке и транспортировке угля происходит его перемешивание. Затем реализуется многократный забор проб угля. Это, в свою очередь, приводит к тому, что процентное содержания углерода в коксующихся углях, оцениваемое по набору случайным образом извлекаемых проб из данной отгружаемой партии углей, оказывается параметром распределения некоторой СПООР. В итоге получают СПООР, состоящую из значений процентов содержания углерода в отобранных пробах угольной руды. Принимая во внимание, что содержание углерода в угле определяет его сортность и, следовательно, цену, понятно, что решение данной задачи имеет важное экономическое значение (рис. В.1).

2. Торгово-промышленная сортировка древесного сырья в деревообрабатывающей промышленности, в частности, сортировка бревен

Содержание углерода
в угольной руде



Параметры распределения
содержания углерода



Маркировка угля

Марки угля	Буквенное обозначение марок	Содержание углерода C_c , %
Бурые	Б	76 и менее
Длинно-пламенные	Д	76
Газовые	Г	83
Жирные	Ж	86
Коксовые	К	88
Отощенно-спекающиеся	ОС	89
Тощие	Т	90
Антрациты	А	91 и более

Рис. В.1. К объяснению технологии торгово-промышленной маркировки углей

по диаметрам, точность которой определяет эффективность использования древесины при оцилиндровке, и, следовательно, лесопильного производства в целом. Данная сортировка проводится на основе информации о значениях диаметров комлей деревьев данного штабеля (рис. В.2).

3. Оценивание прочностной надежности элементов и объектов нефтегазового оборудования с целью определения вероятности безотказной работы нефтепроводов, зависящей от напряжения в трубопроводе и предельного напряжения в трубопроводе, которые также являются СВ с ограниченной областью рассеяния (ограничены слева нулем). При наличии оценки вероятности безотказной работы оказывается возможным вычислить ресурс работы нефтепровода и оценить безопасность технических систем (рис. В.3).

4. Оценивание статистических свойств информационных потоков, передаваемых в телекоммуникационных сетях, важность которой обусловлено постоянно увеличением объемов трафика и типов устройств, генерирующих трафик (рис. В.4).

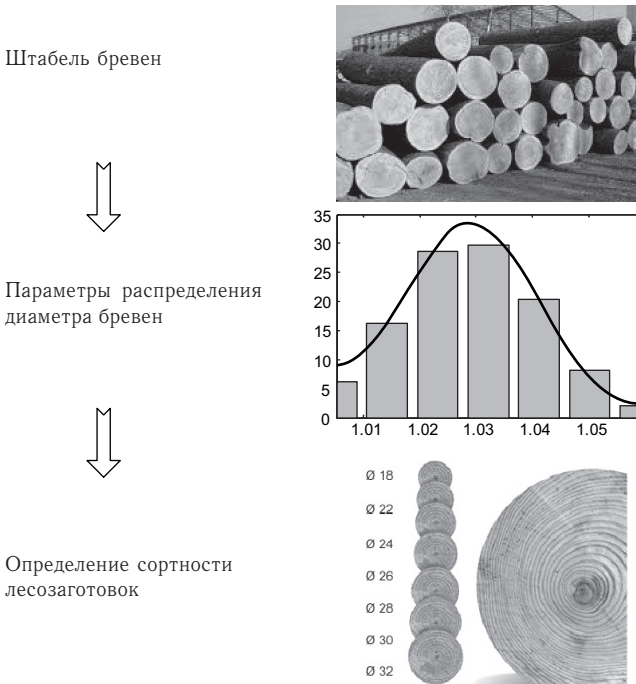


Рис. В.2. К объяснению технологии торгово-промышленной маркировки деревоматериалов

Статистические свойства Интернет-трафика оказываются необходимы на этапе проектирования телекоммуникационных сетей, в том числе нового поколения, где необходимо понимать особенности информационных потоков и использовать соответствующие алгоритмы управления потоками трафика (например, QoS-маршрутизация, управление очередями и планирования обслуживания пакетов, ограничение интенсивности потоков и др.).

Теория случайных блужданий броуновский частицы в ограниченной области рассеяния (ООР), порождающих соответствующие выборки СПООР, была построена в работах А. Эйнштейна и М. Смолуховского. В том числе они предложили математические модели ФР и ПР СПООР. Однако при оценке параметров данных моделей по экспериментальным данным возникает необходимость решения той или иной системы нелинейных уравнений с помощью какого-либо известного итерационного метода (метод Ньютона, симплекс-метод и т. д. [2, 5, 6, 13, 14, 27, 44, 71, 76, 78, 102, 112, 116]). Их сходимость, как известно [30, 84, 108–110], определяется удачным выбором начального приближения, сделать который в многомерном случае

- Напряжение
- Предельное напряжение



Параметры распределения
напряжения и предельного
напряжения



Вероятность безотказной
работы

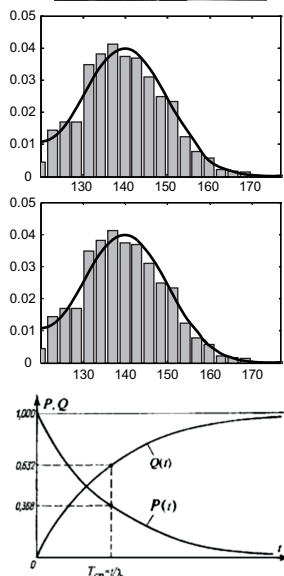


Рис. В.3. К постановке задачи оценивания вероятности безаварийной работы нефтепровода

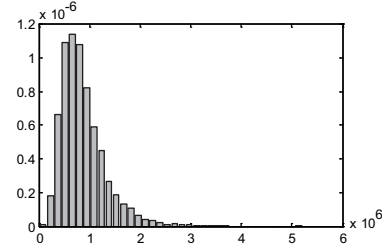
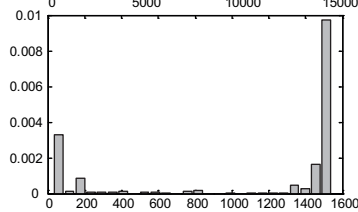
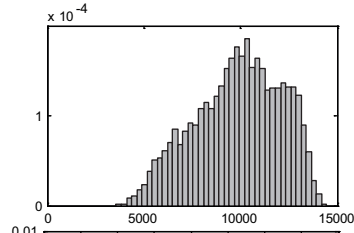
оказывается не всегда возможным. Данное обстоятельство, с нашей точки зрения, явилось одной из причин недостаточного широкого использования моделей Энштейна и Смолуховского в задачах обработки информации. Это подтверждается отсутствием программных реализаций методов и алгоритмов оценивания параметров распределений СПООР как в современных математических пакетах (MATLAB, Statistica, Statgraphics, Mathematica, Maple и др.), так и в математических библиотеках различных языков программирования (например, C++, Python, R и др.).

В тоже время сегодня разработаны методы решения экстремальных задач, основанные на случайном поиске (генетические алгоритмы, муравьиные алгоритмы и т. д.), сходимость которых к истинному решению, как утверждается, не зависит от начального приближения [16, 22, 85]. Однако не существует алгоритмов их использования в задаче аппроксимации эмпирических зависимостей и распределений СПООР.

Статистические свойства
магистрального трафика
(размер потоков, число
пакетов, размер пакетов)



Параметры распределений
свойств трафика



- Выявление аномалий
- Разработка алгоритмов управления трафиком для обеспечения QoS:
 - QoS-маршрутизация
 - Управление очередями
 - Ограничение интенсивности потоков

Рис. В.4. К постановке задачи анализа статистических свойств Интернет-трафика

В этой связи были проведены целенаправленные исследования, результаты которых обеспечили разработку научно-обоснованных алгоритмов оценивания параметров ФР и ПВ СПООР с помощью генетических алгоритмов (ГА), объединяющие параметрические и непараметрические методы оценивания, и их программной реализации в виде Toolbox ES&RP пакета MATLAB.

Монография имеет следующую структуру.

В *первой главе* проведен анализ свойств одномодальных и двумодальных ФР и ПР СПООР в соответствии с подходом, развитым в работах А. Эйнштейна и М. Смолуховского, базирующемся на использовании модели случайного блуждания броуновской частицы в ограниченной с двух сторон области, и показано, что использование

для аппроксимации ФР и ПР СПООР усеченного нормального закона в ряде случаев может приводить к возникновению значимых погрешностей соответствующих параметров ФР и ПР. Проведен анализ постановки задач оценивания параметров одномодальных и двумодальных ФР и ПР СПООР, результаты которого позволили выдвинуть гипотезу о целесообразности использования в задаче восстановления одномодальных и двумодальных ФР и ПР СПООР ГА. Сформулированы основные задачи исследования.

Во *второй главе* обоснована методика использования ГА для оценивания параметров одномодальных ФР и ПР СПООР и поставлена задача оценивания параметров одномодальных ФР и ПР СПООР в терминах ГА. Продемонстрирована ошибочность устоявшихся представлений о сходимости ГА к глобальному минимуму соответствующей целевой функции вне зависимости от выбора начальных значений параметров ГА и применительно к задаче оценивания ФР и ПР СПООР, обоснован выбор начальных значений, а также областей поиска следующих параметров ГА: границ области изменения генов хромосомы, размера популяции, способа формирования начальной популяции, функции приспособленности особи, количества особей, подвергаемых процедуре мутации, оператора мутации, оператора селекции, оператора кроссовера. Приведены аналитические результаты, а также результаты статистического моделирования, подтверждающие рекомендации по выбору начальных параметров ГА при их использовании в задаче оценивания одномодальных ФР и ПР СПООР, а также предложен соответствующий алгоритм оценивания их параметров.

В *третьей главе* поставлена задача оценивания параметров двумодальных ПР и ФР СПООР с помощью ГА и описана методика оценивания параметров двумодальных ПР и ФР СПООР. Для обоснования необходимости разработки специализированной методики проведен анализ параметров двумодальных ФР и ПР СПООР, полученные с помощью методики оценивания параметров одномодальных ФР и ПР СПООР, описанной в главе 2. Результаты анализа показали, что при аппроксимации двумодальной ПР и ФР СПООР для некоторых наборов параметров интегральная погрешность ПР оказывается на порядок больше, чем аналогичная величина, полученная при анализе точности оценок параметров одномодальных ФР и ПР СПООР. Для устранения отмеченного недостатка предложена методика оценивания ФР и ПР двумодальной СПООР, основанная на итерационном использовании ГА и аппроксимации Розенблатта–Парзена, и продемонстрирована ее работоспособность.

В *четвертой главе* приведено описание программной библиотеки (Toolbox ES&RP для пакета MATLAB), обеспечивающей прак-

тическое использование разработанных алгоритмов для оценивания параметров ФР и ПВ СПООР. В состав библиотеки входят 42 функции, в том числе позволяющие реализовать итерационный алгоритм оценивания параметров функций распределения и плотностей распределения двумодальных СВ с ограниченной областью рассеяния, основанный на совместном использовании метода мнимых источников и аппроксимации Розенблатта–Парзена. Для вычисления аппроксимации Розенблатта–Парзена предусмотрена возможность использования следующих ядерных функций: функции Коши, функции Епанчикова, функции Фишера, функции Лапласа, логистической функции, нормальной функции, квадратичной функции, треугольной функции, равномерной функции.

В Приложениях 3–6 приведены результаты применения разработанного алгоритмического, методического и программного обеспечения для анализа особенностей Интернет-трафика, передаваемого в магистральном Интернет-канале, полученные и описанные совместно с Д.А. Божалкиным.