

ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Автоматизированные системы управления»

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

*Методические рекомендации к лабораторным работам
для студентов направлений подготовки
09.03.01 «Информатика и вычислительная техника»
и 09.03.04 «Программная инженерия»
дневной формы обучения*



Могилев 2018

УДК 004.4
ББК 32.973.26
О 23

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Автоматизированные системы управления»
«19» июня 2018 г., протокол № 14

Составитель канд. техн. наук, доц. А. В. Кушнер

Рецензент канд. техн. наук, доц. С. В. Болотов

В методических рекомендациях кратко изложены теоретические сведения, необходимые для лабораторных работ. Рекомендации составлены в соответствии с рабочей программой по дисциплине «Обработка экспериментальных данных» для направлений подготовки 09.03.01 «Информатика и вычислительная техника» и 09.03.04 «Программная инженерия».

Учебно-методическое издание

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Ответственный за выпуск	А. И. Якимов
Технический редактор	А. А. Подошевко
Компьютерная верстка	Н. П. Полевничая

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 16 экз. Заказ №

Издатель и полиграфическое исполнение:
Государственное учреждение высшего профессионального образования
«Белорусско-Российский университет».

Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий

№ 1/156 от 24.01.2014.

Пр. Мира, 43, 212000, Могилев.

© ГУ ВПО «Белорусско-Российский
университет», 2018



Содержание

1 Моделирование экспериментальных данных различных типов	4
2 Динамическое моделирование экспериментальных данных различных типов	10
3 Операции свертки экспериментальных данных различных типов	12
4 Дискретная свертка	16
5 Изучение особенностей обработки экспериментальных данных различных типов	20
6 Изучение алгоритма быстрого преобразования Фурье и его возможностей	26
7 Алгоритмы статистической обработки экспериментальных данных различных типов	28
Список литературы	34



1 Моделирование экспериментальных данных различных типов

Цель работы:

- 1) реализация методов и алгоритмов моделирования на ЭВМ реализаций псевдослучайных последовательностей случайных непрерывных величин с заданным законом распределения;
- 2) реализация методов и алгоритмов статистической проверки полученных эмпирических распределений псевдослучайных чисел и проверка их качества.

1.1 Основные теоретические положения

Для получения последовательностей псевдослучайных чисел с заданным законом распределения чаще всего используют последовательности случайных величин, равномерно распределенных в интервале $[0, 1]$.

Пусть для величины z , имеющей заданное распределение, известна функция плотности распределения $f(z)$. Тогда, используя свойство функции распределения принимать значение от 0 до 1 при изменении аргумента от $-\infty$ до $+\infty$, каждому значению x_i случайной величины X , равномерно распределенной в интервале $[0, 1]$, ставим в однозначное соответствие значение z_i , величины Z :

$$x_i = \int_{-\infty}^{z_i} f(z) dz. \quad (1.1)$$

Формирование последовательности случайных чисел с показательным распределением.

Показательное распределение случайной величины имеет следующую плотность вероятности:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, \forall x \geq 0; \\ 0, \forall x < 0. \end{cases} \quad (1.2)$$

Это распределение зависит от одного параметра λ , а его функция распределения

$$F(x) = \begin{cases} 1 - \lambda e^{-\lambda x}, \forall x > 0; \\ 0, \forall x \leq 0. \end{cases} \quad (1.3)$$

Подставим выражение (1.3) в (1.1), тогда

$$x_i = \lambda \int_0^{z_i} e^{-\lambda x} dz; \quad (1.4)$$



$$x_i = 1 - e^{-\lambda z_i}; \quad (1.5)$$

$$z_i = -\frac{1}{\lambda} \ln(1 - x_i). \quad (1.6)$$

Так как $(1 - x_i)$ случайное равномерно распределенное в интервале $[0; 1]$ число, то

$$z_i = -\frac{1}{\lambda} \ln x_i. \quad (1.7)$$

Теоретические значения математического ожидания m , дисперсии D и среднего квадратического отклонения σ последовательности случайных чисел с показательным распределением:

$$m = \frac{1}{\lambda}; \quad D = \frac{1}{\lambda^2}; \quad \sigma = \sqrt{D}. \quad (1.8)$$

Формирование последовательности случайных чисел, равномерно распределенных в интервале $[a; b]$.

Для формирования псевдослучайных числовых последовательностей z_i , равномерно распределенных в интервале $[a; b]$ с плотностью распределения

$$f(z)' = \begin{cases} \frac{1}{b-a}, & \forall a \leq z \leq b; \\ 0, & \forall z < a; z > b; \end{cases} \quad (1.9)$$

достаточно случайное число x_i из интервала $[0; 1]$ привести к интервалу $[a; b]$ и сдвинуть на величину a :

$$z_i = (b - a)x_i + a. \quad (1.10)$$

При этом теоретическое значение математического ожидания m последовательности случайных чисел и их дисперсия D определяются по формулам:

$$m = \frac{a+b}{2}; \quad D = \frac{(b-a)^2}{12}. \quad (1.11)$$

Формирование последовательности случайных чисел, подчиняющихся распределению Пуассона (приближенный метод).

Во многих задачах (анализ вызовов на телефонной станции, излучение электронов из накаливаемого катода, анализ микробов в воздухе и т. д.) присутствуют случайные величины, распределенные по своеобразному закону, который называется законом Пуассона.



Случайная величина X распределена по закону Пуассона, если вероятность того, что она примет определенное значение m из ряда чисел $0, 1, 2, \dots, m, \dots$, выражается формулой

$$P_m = \frac{\alpha^m}{m!} \cdot e^{-\alpha},$$

где α – некоторая положительная величина, называемая параметром закона Пуассона (может быть и нецелой).

Ряд случайной величины X , распределенной по закону Пуассона, имеет вид, представленный в таблице 1.

Таблица 1

X_m	0	1	2	...	m	...
P_m	$e^{-\alpha}$	$\frac{\alpha}{1!} \cdot e^{-\alpha}$	$\frac{\alpha}{2!} \cdot e^{-\alpha}$...	$\frac{\alpha}{m!} \cdot e^{-\alpha}$...

На рисунке 1.1 приведены распределения случайной величины X по закону Пуассона, соответствующие различным значениям параметра.

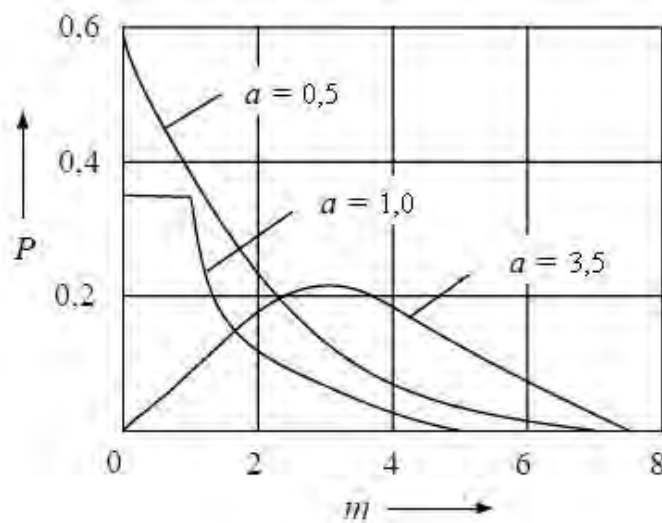


Рисунок 1.1 – Распределение случайной величины X по закону Пуассона

Для формирования такой последовательности может быть использована следующая процедура. Берется произведение равномерно распределенных в интервале $[0; 1]$ чисел x_j . Причем число сомножителей m выбирается таким, чтобы выполнялось неравенство

$$\prod_{i=1}^m x_i < e^{-\lambda}, \quad (1.12)$$

где m – параметр моделируемого распределения Пуассона.

Если условие (1.12) выполняется, то можно утверждать, что число $z_j = m - 1$ будет представлять случайное число, принадлежащее последовательности чисел, распределенных по закону Пуассона с параметром m .

Формирование последовательности случайных чисел, распределенных по нормальному закону.

Плотность распределения последовательности случайных чисел, распределенных по нормальному закону, имеет вид:

$$f(z') = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (1.13)$$

В силу центральной предельной теоремы случайная величина

$$z_j = \sum_{i=1}^N x_i, j = 1, \dots, k \quad (1.14)$$

при достаточно большом N будет иметь распределение, близкое к нормальному. Здесь в качестве x_i могут быть использованы равномерно распределенные псевдослучайные числа; N – количество равномерно распределенных чисел в сумме; k – количество моделируемых нормально распределенных чисел.

Если x_i – некоррелированные величины, то

$$m[z] = \sum_1^N m_{x_i} = N \frac{a+b}{2}; \quad D[z] = \sum_1^N D_{x_i} = N \frac{(a+b)^2}{12}. \quad (1.15)$$

Используя последние выражения для заданного N , можно определить границы $[a; b]$ такие, чтобы Z имела заданные значения параметров m и σ , решив систему уравнений:

$$\sigma = \frac{(b-a)\sqrt{N}}{2\sqrt{3}}; \quad m = \frac{N(a+b)}{2}, \quad (1.16)$$

откуда

$$a = \frac{m - \sigma\sqrt{N}}{N}; \quad b = \frac{m + \sigma\sqrt{3N}}{N}. \quad (1.17)$$

Для получения псевдослучайных чисел, равномерно распределенных на интервале $[a; b]$, можно использовать преобразование (1.10).

Используя аппарат функциональных преобразований, а также разложение функции в ряды специального вида, можно получить уточненную формулу для формирования последовательностей случайных чисел, распределенных по нормальному закону:



$$z = z' + \frac{1}{20 \cdot N} \left[(z')^3 - 3 \cdot z' \right].$$

Вычисление основных статистических характеристик полученных псевдослучайных числовых последовательностей.

Оценка математического ожидания m_p , дисперсии D_p и среднего квадратического отклонения σ_p числовой последовательности x_i выполняется по формулам:

$$m_p = \frac{1}{n} \sum_{i=1}^n x_i; \quad D_p = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2; \quad \sigma_p = \sqrt{D_p}, \quad (1.18)$$

где x_i – i -е значение псевдослучайной числовой последовательности.

Оценка результатов статистического моделирования реализаций случайных величин.

В силу случайных причин теоретические значения m , D и σ , полученные по формулам (1.8) и вычисленные на основе смоделированных экспериментальных реализаций по формулам (1.18), будут отличаться на величину ε , называемую точностью оценки:

$$(m - m_p) < \varepsilon; \quad (D - D_p) < \varepsilon. \quad (1.19)$$

Вероятность α того, что неравенства (1.19) выполняются, называют достоверностью оценки. Тогда

$$p(|m - m_p| < \varepsilon) = \alpha. \quad (1.20)$$

Согласно центральной предельной теореме, оценка среднего при больших значениях реализаций N будет иметь распределение, близкое к нормальному, с математическим ожиданием m и дисперсией σ^2/N . При этом точность оценки ε смоделированной нормально распределенной числовой последовательности при числе реализаций N может быть определена выражением

$$\varepsilon = \frac{t_\alpha \sigma_p}{\sqrt{N}}, \quad (1.21)$$

где t_α – квантиль порядка α (заданной вероятности α) для нормальной функции распределения с параметрами $m = 0$; $\sigma = 1$.

Квантилем порядка α одномерного распределения называется такое значение t_α случайной величины t , для которого $p(t < t_\alpha) = \alpha$.

Для нормального закона распределения случайной величины с параметрами m и σ вероятность $p(t < t_\alpha)$ определяется следующим образом:



$$p(t < t_\alpha) = \Phi\left(\frac{t_\alpha - \alpha}{\sigma}\right), \quad (1.22)$$

где Φ – табулированная нормальная функция распределения Лапласа.

При моделировании случайных величин важным является вопрос о соответствии полученной реализации заданному закону распределения.

Последняя проверка производится с помощью критериев согласия, из которых чаще всего используется критерий χ^2 Пирсона.

1.2 Порядок выполнения работы

1 Изучить методические рекомендации.

2 Выбрать исходные данные для выполнения работы согласно варианту, указанному преподавателем.

3 Разработать математическую модель и алгоритм генерации заданных псевдослучайных последовательностей, учитывающий возможность вычисления математического ожидания и среднеквадратического отклонения экспериментальной числовой последовательности.

4 Разработать программную реализацию алгоритма и провести расчетный эксперимент.

5 Оформить отчет.

Содержание отчета

1 Цель работы.

2 Постановка задачи и исходные данные.

3 Схема алгоритма и программа.

4 Результаты моделирования реализаций псевдослучайных числовых последовательностей и их основные статистические характеристики.

5 Выводы по оценке результатов моделирования реализаций псевдослучайных числовых последовательностей.

Контрольные вопросы

1 Какие методы моделирования равномерно распределенных псевдослучайных чисел Вы знаете?

2 Почему для получения последовательности случайных чисел с заданным законом распределения в качестве базовой последовательности используются равномерно распределенные случайные числа?

3 Как получить последовательность равномерно распределенных в интервале $[a; b]$ псевдослучайных чисел ?



2 Динамическое моделирование экспериментальных данных различных типов

Цель работы: смоделировать динамически поступающие экспериментальные данные.

2.1 Основные теоретические положения

Имитационное моделирование – это представление динамического поведения системы посредством продвижения ее от одного состояния к другому в соответствии с хорошо определенными операционными правилами.

Изменения состояния системы могут происходить либо непрерывно, либо в дискретные моменты времени. Хотя процедуры описания динамического поведения дискретно и непрерывно изменяющихся моделей различны, основная концепция имитации системы – отображение изменений ее состояния с течением времени – остается той же. Практически одну и ту же систему можно представить в виде дискретно изменяющейся модели (далее называемой просто дискретной) либо непрерывно изменяющейся (непрерывной). Как правило, в имитационном моделировании время является основной независимой переменной. Другие переменные, включенные в имитационную модель, являются функциями времени, т. е. зависимыми переменными. При дискретной имитации зависимые переменные изменяются дискретно в определенные моменты имитационного времени, называемые моментами свершения событий. Переменная времени в имитационной модели может быть либо непрерывной, либо дискретной в зависимости от того, могут ли дискретные изменения зависимых переменных происходить в любые моменты времени или только в определенные моменты.

При дискретной имитации состояние системы может меняться только в моменты свершения событий. Так как состояние системы не изменяется между этими моментами, полный динамический портрет состояний системы может быть получен путем продвижения имитационного времени от одного события к другому.

Функционирование дискретной имитационной модели можно задать следующим образом: определяя изменения состояния системы, происходящие в момент свершения событий; описывая действия, в которых принимают участие элементы системы, или процесс, через который проходят элементы.

В непрерывной имитационной модели состояние системы представляется с помощью непрерывно изменяющихся зависимых переменных. Непрерывная имитационная модель создается путем задания уравнений для совокупности переменных состояния, динамическое поведение которых имитирует реальную систему.

Для того чтобы отличать непрерывно изменяющиеся переменные от дискретно изменяющихся, будем первые называть переменными состояния.

Модели непрерывных систем часто определяются в терминах производных переменных состояния. Это объясняется тем, что иногда легче задать выражение для определения скорости изменения переменной состояния, чем сделать это непосредственно для самой переменной. Уравнения такого вида, включающие

производные переменных состояния, называются дифференциальными уравнениями. Пусть, например, в процессе разработки модели составляется следующее дифференциальное уравнение для переменной состояния S по времени t :

$$\frac{dS(t)}{dt} = S^2(t) + t^2; S(0) = k.$$

Первое уравнение определяет скорость изменения S как функцию от S и t , второе – начальное условие для переменной состояния. Цель имитационного эксперимента определить реакцию переменной состояния в зависимости от имитационного времени.

В некоторых случаях возможно определение аналитического выражения для переменной состояния S , заданного уравнением для $\frac{dS}{dt}$. Однако на практике в большинстве случаев аналитическое выражение для S неизвестно. В результате необходимо получить реакцию путем интегрирования $\frac{dS}{dt}$ по времени, используя уравнение следующего вида:

$$S(t_2) = S(t_1) + \int_{t_1}^{t_2} \left(\frac{dS}{dt} \right) dt.$$

Каким образом выполняется интегрирование, зависит от того, использует ли разработчик аналоговый или цифровой компьютер. В 50-х и 60-х гг. аналоговые компьютеры были основным средством реализации непрерывных моделей. Аналоговые компьютеры представляют переменные состояния в модели с помощью электрических цепей. Динамическая структура системы моделируется с помощью таких элементов, как резисторы, конденсаторы и усилители. Основной недостаток аналоговых компьютеров состоит в том, что от характеристик этих элементов зависит точность результатов. Кроме того, в аналоговом компьютере мало логических контрольных функций и отсутствуют те возможности хранения данных, которые имеются в цифровом компьютере.

Ряд непрерывных имитационных языков был разработан для цифровых компьютеров. Несмотря на то, что цифровой компьютер является дискретным устройством, почти любая переменная, значение которой ограничивается только размером слова компьютера, может рассматриваться как непрерывная.

Компьютер с большой скоростью и точностью выполняет основные математические операции, такие как сложение, умножение и логическое тестирование. Выполнение же интегрирования требует применения числовых методов интегрирования. При использовании этих методов независимая переменная (обычно время) разделяется на части, называемые шагами. Значения переменных состояния, требующие интегрирования, получаются путем аппроксимации производных этих переменных по времени. Точность получаемых значений зависит от порядка аппроксимационного метода и размера шага: более высокую

точность дают аппроксимации высокого порядка и наименьшие размеры шагов. Так как аппроксимации высокого порядка и небольшие размеры шага требуют больше вычислений, то существует зависимость между точностью вычислений переменной состояния и затрачиваемым при этом машинным временем.

2.2 Порядок выполнения работы

1 Изучить методические рекомендации.

2 Выбрать исходные данные для выполнения работы согласно варианту, указанному преподавателем.

3 Разработать математическую модель и алгоритм генерации заданных псевдослучайных последовательностей изменяющихся динамически, учитывающий возможность вычисления математического ожидания и среднеквадратического отклонения экспериментальной числовой последовательности.

4 Разработать программную реализацию алгоритма и провести расчетный эксперимент.

5 Оформить отчет.

Содержание отчета

1 Цель работы.

2 Постановка задачи и исходные данные.

3 Схема алгоритма и программа.

4 Результаты моделирования реализаций псевдослучайных числовых последовательностей, изменяющихся динамически, и их основные статистические характеристики.

5 Выводы по оценке результатов моделирования реализаций псевдослучайных числовых последовательностей.

Контрольные вопросы

1 Что такое имитационное моделирование?

2 Чем характеризуются динамически изменяющиеся экспериментальные данные?



3 Операции свертки экспериментальных данных различных типов

Цель работы: изучить использование свертки экспериментальных данных различных типов.

3.1 Основные теоретические положения

Свертка – это математическая операция комбинирования двух сигналов (функций) для получения третьего сигнала (функции), которая может рассматриваться как модифицированная версия одного из первоначальных сигналов (функций). Это один из самых важных методов цифровой обработки данных.

Свертка является формальной математической операцией, такой же как умножение, сложение и интегрирование. Сложение берет два числа и производит третье число, в то время как свертка берет два сигнала и производит третий сигнал. Свертка используется во многих областях математики, например, статистике и вероятности. В линейных системах свертка применяется для описания соотношения между тремя сигналами: входным, выходным и импульсной характеристикой. На рисунке 3.1 показано обозначение свертки при использовании ее с линейными системами. Входной сигнал $x[n]$ поступает в линейную систему с импульсной характеристикой $h[n]$, в результате чего порождается выходной сигнал $y[n]$. В форме выражения это выглядит как $x[n]*h[n] = y[n]$. Сложение представляют знаком «+», умножение – знаком «х», свертку – знаком «*». К сожалению, большинство языков программирования используют знак «*» для индикации операции умножения. В компьютерных программах этот знак обозначает умножение, в то время как в выражениях – операцию свертки.

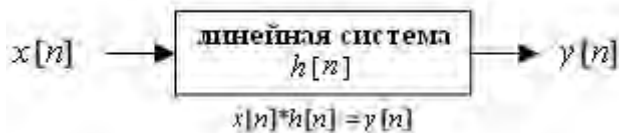


Рисунок 3.1 – Использование операции свертки в цифровой обработке сигналов

На рисунке 3.2 показано использование операции свертки для низкочастотной и высокочастотной фильтрации. Входной сигнал в примере является суммой двух составляющих: три периода синусоиды (представляющих высокую частоту) и медленно возрастающий уклон (состоящий из низких частот). На рисунке 3.2, *а* импульсная характеристика низкочастотного фильтра представляет собой гладкую дугу, пропускающую на выход только медленно изменяющиеся волны. В противоположность ему высокочастотный фильтр на рисунке 3.2, *б* пропускает только быстроизменяющиеся волны.

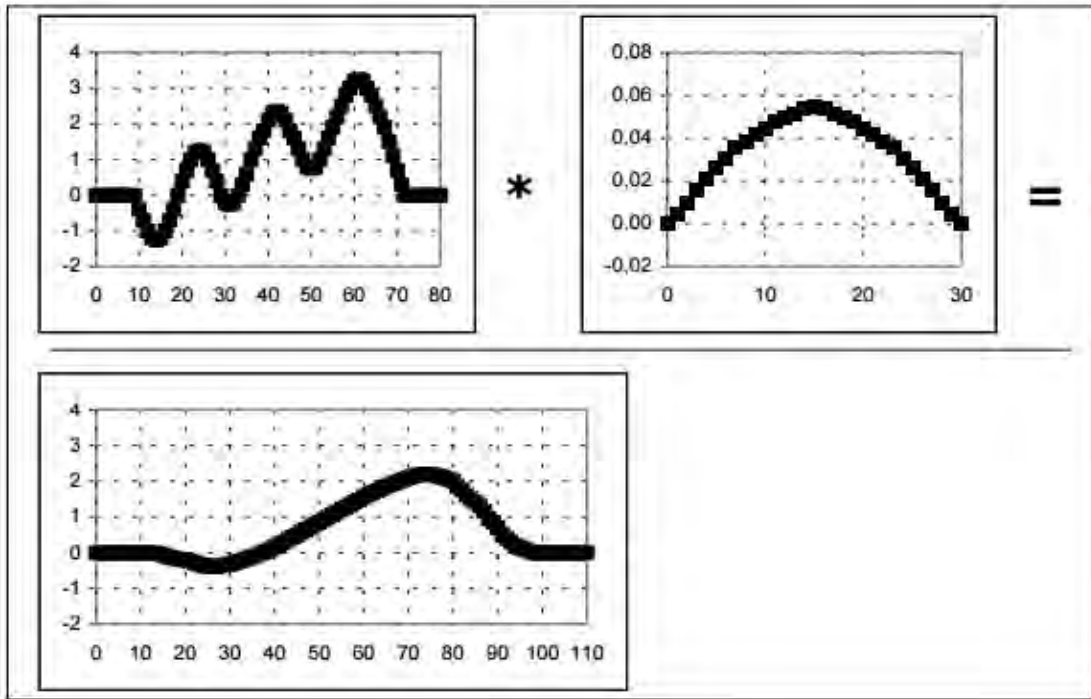
Теперь перейдем к математике. Принимая за основу механизм свертки, можно написать стандартное выражение для расчета свертки. Если $x[n]$ является N -точечным сигналом с индексом от 0 до $N - 1$, и $h[n]$ является M -точечным



сигналом с индексом от 0 до $M - 1$, то свертка этих сигналов $y[n] = x[n]*h[n]$ будет $N + M - 1$ точечным сигналом с индексом от 0 до $N + M - 2$, рассчитываемым как

$$y[i] = \sum_{j=0}^{M-1} h[j]x[i - j].$$

a)



б)

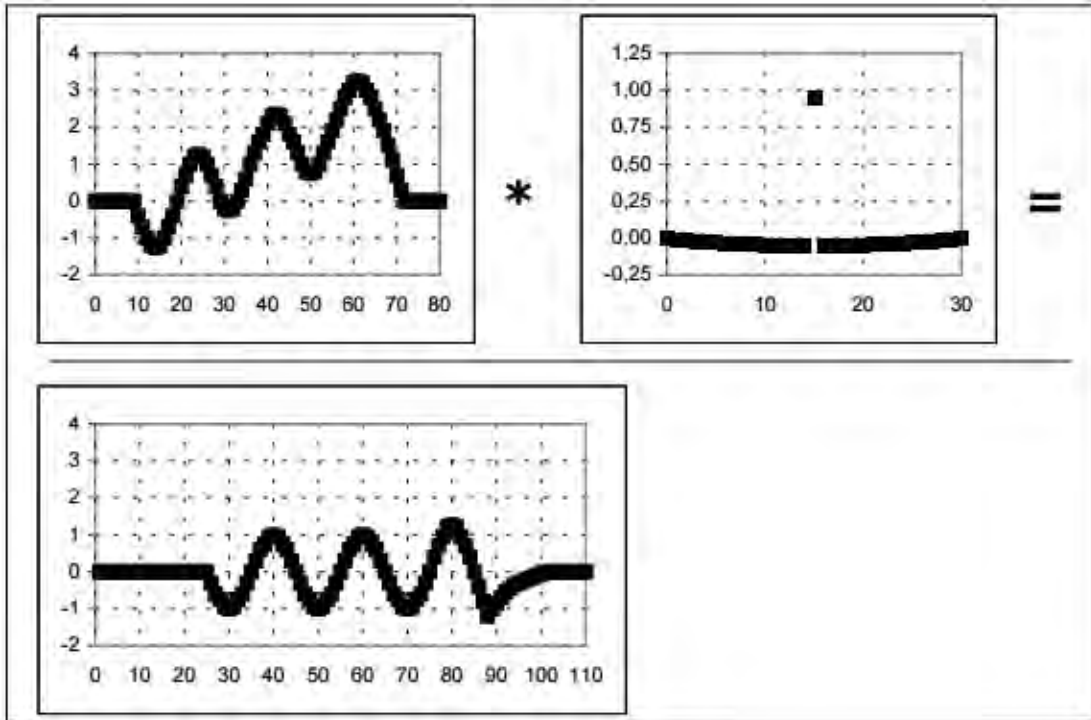


Рисунок 3.2 – Пример фильтрации с помощью операции свертки

Это выражение называется суммой свертки. Оно позволяет вычислить каждую точку выходного сигнала независимо от других. Индекс i определяет, какая выборка выходного сигнала будет вычислена и какой этому соответствует сдвиг позиции механизма свертки. В компьютерных программах, выполняющих операцию свертки, цикл использует этот индекс для прохождения каждой выборки выходного сигнала. Для вычисления одной из точек выходного сигнала внутри цикла механизма свертки используется индекс j . Поскольку j изменяется от 0 до $M - 1$, каждая выборка в импульсной характеристике $h[j]$ умножается на соответствующую выборку входного сигнала $x[i - j]$. Все произведения складываются для получения значения выборки выходного сигнала. Многие операции ЦОС основаны на выражении для расчета свертки. Символ n показывает, что некоторая переменная является индексом в массиве. Иногда выражения записываются в форме $y[n] = x[n] * h$.

3.2 Порядок выполнения работы

- 1 Изучить методические рекомендации.
- 2 Получить исходные данные для выполнения работы у преподавателя.
- 3 Разработать математическую модель и алгоритм фильтрации с использованием свертки.
- 4 Разработать процедуру реализации алгоритма и провести расчетный эксперимент.
- 5 Оформить отчет.

Содержание отчета

- 1 Цель работы.
- 2 Постановка задачи и исходные данные.
- 3 Схема алгоритма и листинг программной процедуры.
- 4 Результаты моделирования математической модели и алгоритм фильтрации с использованием свертки.
- 5 Выводы по оценке результатов моделирования.

Контрольные вопросы

- 1 Что такое свертка?
- 2 Для чего применяется свертка?



4 Дискретная свертка

Цель работы: изучить использование дискретной свертки для обработки экспериментальных данных.

4.1 Основные теоретические положения

Свертка – основной процесс в цифровой обработке сигналов. Поэтому важно уметь эффективно ее вычислять.

Уравнение дискретной свертки двух функций (сигналов) может быть получено непосредственно из интегрального уравнения свертки при замене интегрирования суммированием мгновенных значений функций с шагом Δt :

$$y(k\Delta t) = \Delta t \sum_n h(n\Delta t) \cdot s(k\Delta t - n\Delta t). \quad (4.1)$$

При выполнении дискретной свертки работа производится с цифровыми массивами, при этом шаг дискретизации для массивов по физическому аргументу свертки должен быть равным и принимается за 1, а в качестве аргумента используется нумерация отсчетов в массивах:

$$y(k) = \sum_n h(n) \cdot s(k-n) \equiv \sum_n h_n \cdot s_{k-n} \equiv y_k;$$

$$y(k) = h(n) \cdot s(k-n) \equiv s(k) \cdot h(n) \equiv s_k \cdot h_n$$

Для вычисления свертки массив одной из функций (s_k входного сигнала) располагается по ходу возрастания номеров. Массив второй функции (h_n более короткой, оператор свертки) строится параллельно первому массиву в обратном порядке (по ходу уменьшения номеров, в режиме обратного времени). Для вычисления y_k значение h_0 располагается против s_k , все значения s_{k-n} перемножаются с расположенными против них значениями h_n и суммируются. Результаты суммирования являются выходным значением функции y_k , после чего оператор h_n сдвигается на один номер k вперед (или функция s_k сдвигается ему навстречу) и вычисление повторяется для номера $k+1$ и т. д.

Пример выполнения дискретной свертки показан на рисунке 4.1.

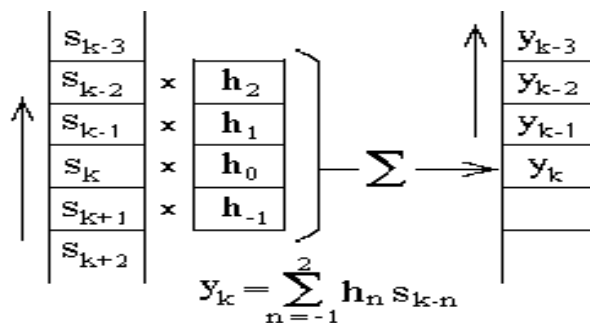


Рисунок 4.1 – Техника выполнения свертки дискретного сигнала

В начальный момент свертки при вычислении значений y_k оператор h_n , построенный в режиме обратного времени, «зависает» для значений $k - n$ при $n > k$ против отсутствующих отсчетов входной функции. «Зависание» исключают либо заданием начальных условий – дополнительных отсчетов, чаще всего нулевых или равных первому отсчету входной функции, либо началом свертки с отсчета входной функции $k = n$ с соответствующим сокращением интервала выходной функции. Для операторов со значениями n (вперед по времени) такой же момент может наступать и в конце входного массива.

Пример – Уравнение свертки: $y_k = \sum_{n=0}^2 b_n x_{k-n} = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2}$.

Значения оператора b_n : $b_0 = 5$, $b_1 = 3$, $b_2 = 2$. Входной сигнал: $x_k = \{0, 1, 0, 0, 0\}$. Начальные условия: $x_{-n} = 0$.

Расчет выходного сигнала

$$y_0 = 5x_0 + 3x_{-1} + 2x_{-2} = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 = 0;$$

$$y_1 = 5x_1 + 3x_0 + 2x_{-1} = 5 \cdot 1 + 3 \cdot 0 + 2 \cdot 0 = 5;$$

$$y_2 = 5x_2 + 3x_1 + 2x_0 = 5 \cdot 0 + 3 \cdot 1 + 2 \cdot 0 = 3;$$

$$y_3 = 5x_3 + 3x_2 + 2x_1 = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 1 = 2;$$

$$y_4 = 5x_4 + 3x_3 + 2x_2 = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 = 0;$$

$$y_5 = 5x_5 + 3x_4 + 2x_3 = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 = 0.$$

Выходной сигнал: $y_k = \{0, 5, 3, 2, 0\}$

Следует отметить, что свертка функции оператора с единичным входным сигналом представляет собой повторение функции оператора свертки на выходе.

Прямое вычисление свертки требует $K \cdot N$ умножений, где K – длина исходного сигнала, а N – длина ядра свертки. Как длина сигнала, так и длина ядра свертки может достигать нескольких тысяч точек, и число умножений становится огромным.

Для дискретной свертки действительны все свойства и теоремы интегральной свертки. В частности, свертка функций в координатной области отображается произведением их спектров в частотной области, а умножение в координатной области эквивалентно свертке в частотной области. Это значит, что для выполнения свертки двух сигналов можно перевести их в частотную область, перемножить их спектры и перевести результат обратно во временную область, т. е. действовать по следующей схеме:

$$s(k) \Leftrightarrow S(\omega), h(n) \Leftrightarrow H(\omega); \quad Y(\omega) = S(\omega) \cdot H(\omega); \quad Y(\omega) \Leftrightarrow y(k).$$

С появлением алгоритмов быстрого преобразования Фурье (БПФ), позволяющих быстро вычислять преобразования Фурье, вычисление свертки через частотную область стало широко использоваться. При значительных размерах сигналов и длины ядра свертки такой подход позволяет в сотни раз сократить



время вычисления свертки.

Выполнение произведения спектров может производиться только при одинаковой их длине, и оператор $h(n)$ перед дискретным преобразованием Фурье (ДПФ) необходимо дополнять нулями до размера функции $s(k)$.

Второй фактор, который следует принимать во внимание, – это цикличность свертки при ее выполнении в спектральной области, обусловленная периодизацией дискретных функций. Перемножаемые спектры являются спектрами периодических функций, и результат на концевых интервалах может не совпадать с дискретной линейной сверткой, где условия продления интервалов (начальные условия) задаются, а не повторяют главный период.

На рисунке 4.2 приведены результаты свертки сигнала s_k , заданного на интервале $k = (0 - 50)$, с функцией $h_n = a \cdot \exp(-a \cdot n)$, $a = 0,1$. Свертка, выполненная через ДПФ, в левой части интервала резко отличается от линейной свертки. Характер искажения становится понятным, если дополнить главный интервал с левой стороны его периодическим продолжением (на рисунке показана часть левого бокового периода, свертка с которым заходит в главный период). Для операторов h_n со значениями n (вперед по положению) аналогичные искажения появятся и в правой части главного периода. Для устранения таких искажений сигнальная функция должна продлеваться нулями на размер оператора $h(n)$, что исключит наложение боковых периодов главной трассы функции. Следовательно, перед переводом функций $s(k)$ и $h(n)$ в спектральную область их размер должен продлеваться нулями до длины $K + N$ при односторонних операторах $h(n)$ или до длины $K + 2N$ при двусторонних операторах $h(n)$.

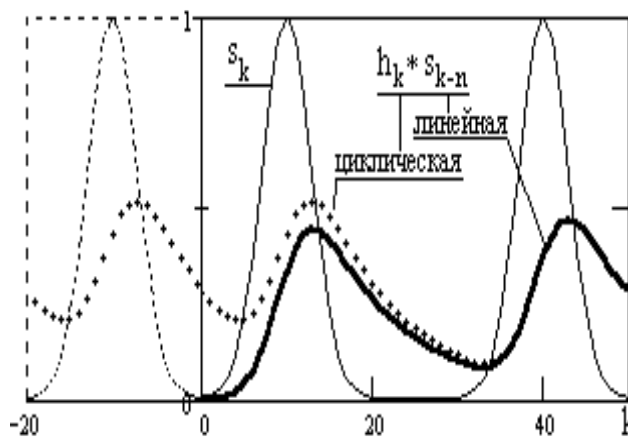


Рисунок 4.2 – Результат свертки сигнала

При выполнении свертки через БПФ ощутимое повышение скорости вычислений появляется только при большой длине функций и операторов (например, $M > 1000$, $N > 100$). Следует также обращать внимание на разрядность результатов, т. к. перемножение чисел дает увеличение разрядности в 2 раза. При ограниченной разрядности числового представления с соответствующим округлением это может приводить к погрешностям суммирования.

В системах оперативной обработки данных часто возникает потребность вычислить свертку очень большого сигнала, поступающего на вход системы после-

довательными порциями (например, при получении данных от датчиков скважинных приборов). В таких случаях применяется так называемая *секционная свертка*. Суть ее состоит в том, что каждая из этих частей сворачивается с ядром отдельно, а затем полученные части объединяются. Для объединения достаточно размещать их друг за другом с наложением (перекрытием) в $N - 1$ точку (N – длина ядра свертки) и производить суммирование в местах перекрытия.

4.2 Порядок выполнения работы

- 1 Изучить методические рекомендации.
- 2 Получить исходные данные для выполнения работы у преподавателя.
- 3 Разработать математическую модель и алгоритм фильтрации с использованием дискретной свертки.
- 4 Разработать процедуру реализации алгоритма и провести расчетный эксперимент.
- 5 Оформить отчет.

Содержание отчета

- 1 Цель работы.
- 2 Постановка задачи и исходные данные.
- 3 Схема алгоритма и листинг программной процедуры.
- 4 Результаты моделирования математической модели и алгоритма фильтрации с использованием дискретной свертки.
- 5 Выводы по оценке результатов моделирования.

Контрольные вопросы

- 1 Что такое свертка?
- 2 Что такое дискретная свертка?
- 3 Для чего применяется дискретная свертка?



5 Изучение особенностей обработки экспериментальных данных различных типов

Цель работы: изучить особенности предварительной обработки экспериментальных данных (ЭД).

5.1 Основные теоретические положения

Пусть в полученной выборке значение x_1 параметра наблюдалось n_1 раз, значение x_2 – n_2 раз, значение x_k – n_k раз, $n_1 + n_2 + \dots + n_k = n$. Совокупность значений, записанных в порядке их возрастания, называют *вариационным рядом*, величины n_i – частотами, а их отношения к объему выборки $n_i = n_i/n$ – *относительными частотами* (частостями). Очевидно, что сумма относительных частот равна единице.

Под распределением понимают соответствие между наблюдаемыми вариантами и их частотами, или частостями. Пусть n_x – количество наблюдений, при которых случайные значения параметра X меньше x . Частость события $X < x$ равна n_x/n . Это отношение является функцией от x и от объема выборки: $F_n(x) = n_x/n$. Величина $F_n(x)$ обладает всеми свойствами функции распределения: $F_n(x)$ неубывающая функция, ее значения принадлежат отрезку $[0 - 1]$; если x_1 – наименьшее значение параметра, а x_k – наибольшее, то $F_n(x) = 0$, когда $x < x_1$, и $F_n(x_k) = 1$, когда $x \geq x_k$.

Функция $F_n(x)$ определяется по экспериментальным данным, поэтому ее называют *эмпирической функцией распределения*. В отличие от эмпирической функции $F_n(x)$ функцию распределения $F(x)$ генеральной совокупности называют теоретической функцией распределения, она характеризует не частость, а вероятность события $X < x$. Из теоремы Бернулли вытекает, что частость $F_n(x)$ стремится по вероятности к вероятности $F(x)$ при неограниченном увеличении n . Следовательно, при большом объеме наблюдений теоретическую функцию распределения $F(x)$ можно заменить эмпирической функцией $F_n(x)$.

График эмпирической функции $F_n(x)$ представляет собой ломаную линию. В промежутках между соседними членами вариационного ряда $F_n(x)$ сохраняет постоянное значение. При переходе через точки оси x , равные членам выборки, $F_n(x)$ претерпевает разрыв, скачком возрастая на величину $1/n$, а при совпадении l наблюдений – на l/n .

При большом объеме выборки (понятие «большой объем» зависит от целей и методов обработки, в данном случае будем считать n большим, если $n > 40$) в целях удобства обработки и хранения сведений прибегают к группированию ЭД в интервалы. Количество интервалов следует выбрать так, чтобы в необходимой мере отразилось разнообразие значений параметра в совокупности и в то же время закономерность распределения не искажалась случайными колебаниями частот по отдельным разрядам. Существуют нестрогие рекомендации по выбору количества u и размера h таких интервалов, в частности:

– в каждом интервале должно находиться не менее пяти-семи элементов.

В крайних разрядах допустимо всего два элемента;

1) количество интервалов не должно быть очень большим или очень маленьким. Минимальное значение u должно быть не менее шести-семи. При объеме выборки, не превышающем несколько сотен элементов, величину u задают в пределах от 10 до 20. Для очень большого объема выборки ($n > 1000$) количество интервалов может превышать указанные значения. Некоторые исследователи рекомендуют пользоваться соотношением $u = 1,441 \ln(n) + 1$;

2) при относительно небольшой неравномерности длины интервалов удобно выбирать одинаковыми и равными величине

$$h = (x_{\max} - x_{\min})/u,$$

где x_{\max} и x_{\min} – максимальное и минимальное значения параметра.

При существенной неравномерности закона распределения длины интервалов можно задавать меньшего размера в области быстрого изменения плотности распределения;

– при значительной неравномерности лучше в каждый разряд назначать примерно одинаковое количество элементов выборки. Тогда длина конкретного интервала будет определяться крайними значениями элементов выборки, сгруппированными в этот интервал, т. е. будет различна для разных интервалов (в этом случае при построении гистограммы нормировка по длине интервала обязательна, в противном случае высота каждого элемента гистограммы будет одинакова).

Группирование результатов наблюдений по интервалам предусматривает определение размаха изменений параметра x , выбор количества интервалов и их величины, подсчет для каждого i -го интервала $[x_i - x_i + 1]$ частоты n_i или относительной частоты (частости n_i) попадания варианты в интервал. В результате формируется представление ЭД в виде *интервального или статистического ряда*.

Графически статистический ряд отображают в виде гистограммы, полигона и ступенчатой линии. Часто *гистограмму* представляют как фигуру, состоящую из прямоугольников, основаниями которых служат интервалы длиной h , а высоты равны соответствующей частоте. Однако такой подход неточен. Высоту i -го прямоугольника z_i следует выбрать равной $n_i/(nh)$. Такую гистограмму можно интерпретировать как графическое представление эмпирической функции плотности распределения $f_n(x)$, в ней суммарная площадь всех прямоугольников составит единицу. Гистограмма помогает подобрать вид теоретической функции распределения для аппроксимации ЭД.

Полигоном называют ломаную линию, отрезки которой соединяют точки с координатами по оси абсцисс, равными серединам интервалов, а по оси ординат – соответствующим частотам. Эмпирическая функция распределения отображается ступенчатой ломаной линией: над каждым интервалом проводится отрезок горизонтальной линии на высоте, пропорциональной накопленной частоте в текущем интервале. Накопленная частота равна сумме всех частот, начиная с первого и до данного интервала включительно.

Эффективность характеризует разброс случайных значений оценки около



истинного значения параметра. Среди всех оценок следует выбрать ту, значения которой теснее сконцентрированы около оцениваемого параметра. Для многих применяемых способов оценивания выборочные распределения параметров асимптотически нормальны, поэтому часто мерой эффективности служит дисперсия оценки. В таком понимании эффективная оценка – это оценка с минимальной дисперсией. При неограниченном увеличении n эффективная оценка является и состоятельной. В случае оценивания одного параметра дисперсия несмещенной оценки отвечает условию Рао-Крамера

$$D(\Theta^*) \geq \left[-nM \left(\frac{\partial^2 \ln f(x, \Theta)}{\partial \Theta^2} \right) \right]^{-1},$$

где $f(x, \Theta)$ – плотность распределения варианты;
 n – количество наблюдений.

Сравнительная эффективность оценки с дисперсией $D_n[\Theta^*]$ измеряется коэффициентом эффективности

$$\varepsilon = D[\Theta^*]/D_k[\Theta^*],$$

который не превышает единицы. Чем ближе коэффициент ε к единице, тем эффективнее оценка. Отмеченное ограничение применимо и к дискретным распределениям, если вместо плотности распределения подставить в него функцию вероятности.

Достаточность характеризует полноту использования информации, содержащейся в выборке. Другими словами, оценка Θ^* будет достаточной, если все другие независимые оценки на основе данной выборки не дают дополнительной информации об оцениваемом параметре. Эффективная оценка обязательно является и достаточной.

Для характеристики эмпирического распределения можно использовать оценки центральных и начальных моментов. Применение находят моменты до четвертого порядка включительно, т. к. точность выборочных моментов резко падает с увеличением их порядка, в частности, дисперсия начальных моментов порядка r зависит от моментов порядка $2r$. Она становится значительной для моментов высокого порядка даже при больших объемах выборки. Выборочные значения моментов определяют непосредственно по выборке или по сгруппированным данным.

Выборочные значения центральных моментов случайной величины X вычисляются по выборке с применением с формул:

$$\begin{aligned} \mu_1 &= \frac{1}{n} \sum_{i=1}^n x_i; \\ \tilde{\mu}_k &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu_1)^k; \quad k = 2, 3, 4. \end{aligned} \quad (5.1)$$



Эти величины являются оценками соответствующих теоретических моментов $m_1 \dots m_4$ и должны рассматриваться как случайные. Вычисления по формулам (5.1) дают состоятельные, но смещенные оценки моментов старше первого. Смещение удастся устранить введением поправочных коэффициентов, зависящих от объема выборки. Несмещенными и состоятельными будут оценки

$$\begin{aligned}\mu_2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_1)^2; \\ \mu_3 &= \frac{n^2}{(n-1)(n-2)} \mu_3; \\ \mu_4 &= \frac{n(n^2 - 2n + 3)\mu_4 - 3n(2n-3)\mu_2^2}{(n-1)(n-2)(n-3)}.\end{aligned}\quad (5.2)$$

Оценки моментов по сгруппированным ЭД:

$$\begin{aligned}\mu_{1,g} &= \frac{1}{n} \sum_{j=1}^y n_j X_{u,j}; \\ \tilde{\mu}_{k,g} &= \frac{1}{n} \sum_{j=1}^y n_j (X_{u,j} - \mu_{1,g})^k, k = 2, 3, 4, \dots,\end{aligned}\quad (5.3)$$

где $X_{u,j}$ – центр j -го интервала;

y – количество интервалов.

Группирование и приписывание соответствующей частоты значения варианты в середине интервала группирования вносят некоторые искажения. Если распределение непрерывно и имеет достаточно высокий порядок соприкосновения с осью абсцисс (значения функции плотности распределения быстро убывают при удалении от центра распределения), то для снижения ошибок группирования используют поправки Шеппарда. Уточненные значения выборочных моментов для случая равной длины всех интервалов определяются через оценки моментов, вычисленные по сгруппированным данным:

$$\begin{aligned}\mu_1 &= \mu_{1,g}; \mu_2 = \mu_{2,g} - h^2 / 12; \\ \mu_3 &= \mu_{3,g}; \mu_4 = \mu_{4,g} - \mu_{2,g} h^2 / 2 + 7h^4 / 240,\end{aligned}\quad (5.4)$$

где h – длина интервала группирования. Указанные поправки ведут к уточнению только при соблюдении указанного условия, в противном случае они могут привести к еще большей ошибке.

Начальный эмпирический момент порядка r по несгруппированным данным определяется соотношением



$$\eta_r = \frac{1}{n} \sum_{i=1}^n x_i^r; r = 1, 2, 3, \dots \quad (5.5)$$

Центральные и начальные оценки моментов связаны между собой следующими зависимостями:

$$\begin{aligned} \mu_1 &= \eta_1; \tilde{\mu}_2 = \eta_2 - \eta_1^2; \\ \mu_3 &= \eta_3 - 3\eta_1\eta_2 + 2\eta_1^3; \mu_4 = \eta_4 - 4\eta_1\eta_3 + 6\eta_1^2\eta_2 - 3\eta_1^4. \end{aligned} \quad (5.6)$$

В процессе обработки ЭД проще сначала определить оценки начальных моментов, потом перейти к смещенным оценкам центральных моментов и затем вычислить несмещенные оценки.

Квантиль, отвечающий уровню вероятности g , – такое значение варианты x_g , при котором функция распределения случайной величины принимает значение g , т. е. квантиль – это значение аргумента x_g функции распределения, при котором $F(x_g) = g$. Эмпирический квантиль находят по заданному значению вероятности g , используя вариационный ряд или ступенчатую ломаную линию.

Наряду с указанными параметрами для описания распределений применяются и другие характеристики:

- среднеквадратическое отклонение $\sigma = \sqrt{\mu_2}$;
- коэффициенты асимметрии $\beta_1 = \mu_3 / \mu^3$ и эксцесса $\beta_2 = \mu_4 / \mu^2$;
- стандартизованные переменные $u = (x - m_1) / s$.

Коэффициент асимметрии характеризует «скошенность» распределения относительно симметричного нормального распределения (у любого симметричного распределения $b_1 = 0$) (рисунок 5.1, а). Этот показатель в основном зависит от крайних значений выборки.

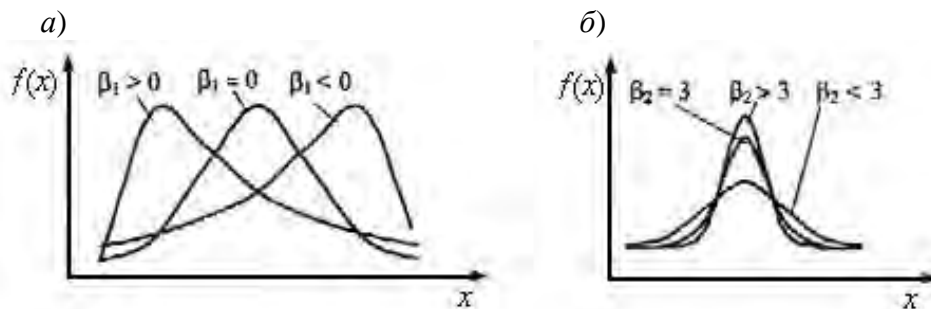


Рисунок 5.1 – Асимметрия распределения (а) и эксцесс распределения (б)

Коэффициент эксцесса характеризует островершинность распределения относительно нормального распределения (этот коэффициент у нормального распределения равен трем) (рисунок 5.1, б). Термин «эксцесс» (превышение) целесообразно применять не к величине b_2 , а к сравнению этой величины изучаемого распределения с величиной данного коэффициента нормального рас-

пределения, т. е. с величиной, равной трем. Исходя из этого, часто вместо b_2 используют выражение $b_2 - 3$.

Стандартизация переменной позволяет упростить расчеты, кроме того, в литературе многие справочные статистические таблицы приводятся именно для стандартизованных переменных. Нетрудно показать, что математическое ожидание стандартизованной переменной равно нулю, а дисперсия равна единице, т. е. после такого преобразования ЭД справедливы следующие соотношения:

$$M(u) = \frac{1}{n} \sum_{i=1}^n u_i = 0; D(u) = \frac{1}{n} \sum_{i=1}^n (u_i - 0)^2 = 1.$$

Величина u называется *центрированной и нормированной*. Переход от центрированной и нормированной величины к исходной осуществляется простым преобразованием $x = u_s + m_1$. Потери информации при стандартизации и обратном преобразовании не происходит.

5.2 Порядок выполнения работы

- 1 Изучить методические рекомендации.
- 2 Разработать математическую модель и алгоритм обработки экспериментальных данных по заданию преподавателя.
- 3 Разработать процедуру реализации алгоритма и провести расчетный эксперимент.
- 4 Оформить отчет.

Содержание отчета

- 1 Цель работы.
- 2 Постановка задачи и исходные данные.
- 3 Схема алгоритма и листинг программной процедуры.
- 4 Результаты моделирования математической модели и алгоритм фильтрации с использованием дискретной свертки.
- 5 Выводы по оценке результатов моделирования.

Контрольные вопросы

- 1 Назовите методы оценки экспериментальных данных.
- 2 Как оценить данные по гистограмме?



6 Изучение алгоритма быстрого преобразования Фурье и его возможностей

Цель работы: изучить быстрое преобразование Фурье (БПФ) и его возможности, разработать алгоритм анализа заданного сигнала с использованием БПФ.

6.1 Основные теоретические положения

Быстрое преобразование Фурье (БПФ, fast Fourier transform – FFT) базируется на том, что при вычислениях среди множителей (синусов и косинусов) есть много периодически повторяющихся значений (в силу периодичности функций). Алгоритм БПФ группирует слагаемые с одинаковыми множителями в пирамидальный алгоритм, значительно сокращая число умножений за счет исключения повторных вычислений. В результате быстрое действие БПФ в зависимости от N может в сотни раз превосходить быстрое действие стандартного алгоритма. При этом следует подчеркнуть, что алгоритм БПФ даже точнее стандартного, т. к. сокращая число операций, он приводит к меньшим ошибкам округления.

Допустим, что массив чисел s_k содержит $N = 2^r$ отсчетов (r – целое). Разделим исходный массив на два первых промежуточных массива с четными и нечетными отсчетами:

$$s_k' = s_{2k}, s_k'' = s_{2k+1}; 0 \leq k \leq N/2 - 1.$$

Дискретное преобразование Фурье (ДПФ) может быть получено непосредственно из интегрального преобразования дискретизаций аргументов ($t_k = k \cdot \Delta t, f_n = n \cdot \Delta f$):

$$S(f) = \int_{-\infty}^{+\infty} s(t) e^{-2j\pi ft} dt, S(f_n) = \Delta t \cdot \sum_{k=-\infty}^{+\infty} s(t_k) e^{-2j\pi f_n k \Delta t}; \quad (6.1)$$

$$s(t) = \int_{-\infty}^{+\infty} S(f) e^{2j\pi ft} df; s(t_k) = \Delta f \cdot \sum_{n=-\infty}^{+\infty} S(f_n) e^{2j\pi n \Delta f t_k}. \quad (6.2)$$

Дискретизация функции по времени приводит к периодизации ее спектра, а дискретизация спектра по частоте – к периодизации функции. Значения (6.1) числового ряда $S(f_n)$ являются дискретизацией непрерывной функции $S'(f)$ спектра дискретной функции $s(t_k)$, равно как и значения (6.2) числового ряда $s(t_k)$ являются дискретизацией непрерывной функции $s'(t)$, и при восстановлении этих непрерывных функций $S'(f)$ и $s'(t)$ по их дискретным отсчетам соответствие $S'(f) = S(f)$ и $s'(t) = s(t)$ гарантировано только при выполнении теоремы Котельникова-Шеннона.

Выполним ДПФ каждого массива с учетом того, что шаг функций равен 2 (при $\Delta t = 1$), а период промежуточных спектров будет соответственно ра-



вен $N/2$: $s_k' \Rightarrow S_n'$, $s_k'' \Rightarrow S_n''$, $0 \leq n \leq N/2-1$.

Для получения одной половины искомого спектра S_n сложим полученные спектры с учетом теоремы запаздывания, т. к. отсчеты функции s_k'' сдвинуты относительно s_k' на один шаг дискретизации:

$$S_n = S_n' + S_n'' \cdot e^{-j2\pi n/N}.$$

Вторая половина спектра, комплексно сопряженная с первой, с учетом периода повторения $N/2$ промежуточных спектров определяется выражением

$$S_{n+N/2} = S_n' + S_n'' \cdot e^{-j2\pi(n+N/2)/N} = S_n' - S_n'' \cdot e^{-j2\pi n/N}. \quad (6.3)$$

Нетрудно видеть, что для вычисления полного спектра в данном случае потребуется $N^2/4$ операций для вычисления промежуточных спектров плюс еще N операций комплексного умножения и сложения, что создает ощутимый эффект по сравнению с ДПФ.

Но деление массивов на две части может быть применено и к первым промежуточным массивам, и ко вторым, и т. д. до тех пор, пока в массивах не останется по одному отсчету, фурье-преобразование которых равно самому отсчету. Тем самым алгоритм преобразования превращается в пирамидальный алгоритм перестановок со сложением/вычитанием и с единичным умножением на значение $e^{-j2\pi n/N}$ соответствующего уровня пирамиды. Первый алгоритм БПФ на данном принципе (из множества модификаций, существующих в настоящее время) был разработан Кули-Тьюки в 1965 г. и позволил повысить скорость вычислений в N/r раз по сравнению с ДПФ. Чем больше N , тем больше эффект БПФ. Так, при $N = 1024$ имеем $r = 10$ и, соответственно, $N/r \approx 100$. Условия по количеству точек $N = 2^r$ рассматривается в варианте $N_k J 2^r$, где r – минимальное целое. Массивы с $N_k < 2^r$ дополняются до 2^r нулями, что не изменяет форму спектра. Изменяется только шаг $\Delta\omega$ по представлению спектра ($\Delta\omega = 2\pi/2^r < 2\pi/N$), который несколько избыточен по адекватному представлению сигнала в частотной области. В настоящее время существуют и алгоритмы БПФ с другими основаниями и их комбинациями, при которых не требуется дополнения сигналов нулями до 2^r .

Заметим, что в соответствии с (6.3) отсчеты, сопряженные с правой половиной главного частотного диапазона $(0, \pi)$, относятся не к диапазону $(-\pi, 0)$, а к диапазону $(\pi, 2\pi)$, что, учитывая периодичность спектра дискретных данных, значения не имеет. То есть выходной частотный диапазон БПФ равен $(0, 2\pi)$. Общее количество отсчетов комплексного спектра в этом условно главном диапазоне равно количеству точек исходного сигнала (с учетом нулевых точек при дополнении сигнала до $N = 2^r$). Алгоритм быстрого обратного преобразования (ОБПФ) тождественен алгоритму прямого БПФ [1].

Алгоритмы прямого и обратного БПФ широко используются в современном программном обеспечении для анализа и обработки цифровых данных.

6.2 Порядок выполнения работы

- 1 Изучить методические рекомендации.
- 2 Получить исходные данные для выполнения работы у преподавателя.
- 3 Разработать математическую модель и алгоритм быстрого преобразования Фурье.
- 4 Разработать процедуру реализации алгоритма и провести расчетный эксперимент.
- 5 Оформить отчет.

Содержание отчета

- 1 Цель работы.
- 2 Постановка задачи и исходные данные.
- 3 Схема алгоритма и листинг программной процедуры.
- 4 Результаты моделирования и алгоритм обработки данных с использованием БПФ.
- 5 Выводы по оценке результатов моделирования

Контрольные вопросы

- 1 Что такое преобразование Фурье?
- 2 Что такое быстрое преобразование Фурье?
- 3 Для каких целей можно использовать БПФ?

7 Алгоритмы статистической обработки экспериментальных данных различных типов

Цель работы: изучить алгоритмы статистической обработки различных типов данных.

7.1 Основные теоретические положения

Алгоритм статистической обработки данных на примере обработки биомедицинских данных.

7.1.1 Представление данных.

- 1 Определить тип признака, который необходимо описать. Если признак количественный, перейти к п. 2, если качественный – к п. 5.
- 2 Оценить нормальность распределения признака с помощью критерия Колмогорова-Смирнова (при числе исследуемых $n > 50$) или критерия Шапиро-Уилка (при $n < 50$). Если распределение нормальное, перейти к п. 3, если отличается от нормального – к п. 4.
- 3 При описании нормально распределенного количественного признака



указать среднее значение M , стандартное отклонение σ или стандартную ошибку m , 95-процентный доверительный интервал (ДИ).

4 При описании количественного признака, распределение которого отличается от нормального, указать медиану, значения нижнего и верхнего квартилей (или 25 и 75 % перцентилей).

5 При описании качественного признака для каждого его значения указать абсолютную величину, а также процентную долю в структуре всей совокупности.

7.1.2 Сравнение двух групп пациентов по количественному признаку.

1 Оценить нормальность распределения признака с помощью критерия Колмогорова-Смирнова (при числе исследуемых $n > 50$) или критерия Шапиро-Уилка (при $n < 50$). Если распределение нормальное, следует использовать алгоритм, описанный в пп. 2 и 3, если отличается от нормального – алгоритм, изложенный в пп. 4 и 5.

2 Для сравнения использовать t -критерий Стьюдента для несвязанных совокупностей. Оценить статистическую значимость различий показателей, сравнив рассчитанное значение t -критерия с критическим или определив уровень значимости p с помощью статистической программы.

3 Описать результаты сравнения исследуемых групп, указав вначале средние значения показателя для каждой группы, стандартное отклонение или стандартную ошибку, при необходимости – границы 95-процентного доверительного интервала. Отметить, в какой из сравниваемых групп среднее значение анализируемого показателя было выше. В выводе указать уровень значимости различий p . В качестве графического изображения для наглядного представления результатов применяют столбиковую диаграмму.

4 Для сравнения необходимо использовать непараметрический U -критерий Манна-Уитни. Оценить статистическую значимость различий показателей, сравнив рассчитанное значение U -критерия с критическим значением или определив уровень значимости p с помощью статистической программы.

5 Описать результаты сравнения исследуемых групп, указав вначале медиану и значения нижнего и верхнего квартилей показателя для каждой группы. Отметить, в какой из сравниваемых групп медиана анализируемого показателя выше. В выводе указать уровень значимости различий p . Для графического представления результатов сравнения следует использовать ящичную диаграмму («ящик с усами», box-plot).

7.1.3 Сравнение количественных параметров в динамике при двух-этапном измерении.

1 Оценить нормальность распределения признака с помощью критерия Колмогорова-Смирнова (при числе исследуемых $n > 50$) или критерия Шапиро-Уилка (при $n < 50$). Если распределение нормальное, необходимо использовать алгоритм, описанный в пп. 2 и 3, если отличается от нормального – алгоритм, изложенный в пп. 4 и 5.

2 Для сравнения нужно использовать t -критерий Стьюдента для связанных совокупностей (парный t -критерий Стьюдента). Оценить статистическую зна-

чимось различий показателей, сравнив рассчитанное значение t -критерия с критическим или определив уровень значимости p с помощью статистической программы.

3 Описать результаты сравнения исследуемых групп, указав вначале средние значения и стандартное отклонение или стандартную ошибку показателя для каждого этапа наблюдения, при необходимости отметить границы 95-процентного доверительного интервала. Определить, в каком направлении изменился показатель: снизился или увеличился. В выводе указать уровень значимости различий p , определенного с помощью парного t -критерия Стьюдента. В качестве графического изображения для наглядного представления результатов нужно использовать столбиковую диаграмму.

4 Для сравнения использовать ранговый T -критерий Уилкоксона. Оценить статистическую значимость различий показателей, сравнив рассчитанное значение критерия Уилкоксона с критическим или определив уровень значимости p с помощью статистической программы.

5 Описать результаты сравнения исследуемых групп, указав вначале медиану и значения нижнего и верхнего квартилей показателя для каждого этапа наблюдения. Отметить, в какой из сравниваемых групп медиана анализируемого показателя выше. В выводе указать уровень значимости различий p . Для графического представления результатов сравнения нужно использовать ящичную диаграмму («ящик с усами», box-plot).

7.1.4 Сравнение трех и более групп по количественному признаку.

1 Оценить нормальность распределения показателей в каждой из сравниваемых групп с помощью критерия Колмогорова-Смирнова (при числе исследуемых $n > 50$) или критерия Шапиро-Уилка (при $n < 50$). Если во всех группах распределение нормальное, необходимо следовать пп. 2–4 алгоритма, если хотя бы в одной группе отличается от нормального – пп. 5–7.

2 Для сравнения нужно использовать однофакторный дисперсионный анализ, в результате которого находится значение F -критерия Фишера. Оценить статистическую значимость различий показателей, сравнив рассчитанное значение F -критерия с критическим или определив уровень значимости p с помощью статистической программы.

3 Описать результаты сравнения исследуемых групп, указав вначале средние значения и стандартное отклонение или стандартную ошибку показателя для каждой группы исследуемых, при необходимости указать границы 95-процентного доверительного интервала. Сравнить между собой средние показатели, отметив, в какой из групп было получено наибольшее, а в какой наименьшее значение. В выводе нужно указать уровень значимости различий p , определенного при помощи F -критерия Фишера. Для графического представления результатов следует использовать столбиковую диаграмму, где каждый столбик соответствует значению показателя в определенной группе.

4 В случае выявления статистически значимых различий при сравнении всех групп для более точного описания наблюдаемых тенденций необходимо использовать апостериорные критерии, позволяющие оценить различия показа-

телей при сравнении групп попарно. Если размеры сравниваемых групп равны или имеют относительно небольшую разницу, парные сравнения проводятся с помощью критерия Тьюки. Если размеры сравниваемых групп существенно различаются, нужно использовать критерий Шеффе.

5 Для сравнения показателей, распределение которых отличается от нормального, следует использовать непараметрический критерий Краскела-Уоллиса. Оценить статистическую значимость различий показателей, сравнив рассчитанное значение критерия Краскела-Уоллиса с критическим или определив уровень значимости p с помощью статистической программы.

6 Описать результаты сравнения исследуемых групп, указав вначале медианы и значения нижнего и верхнего квартилей показателей для каждой из сравниваемых групп. Отметить, в какой группе медиана показателя имеет самое высокое значение, а в какой самое низкое. В выводе указать уровень значимости различий p . Для графического представления результатов использовать ящичную диаграмму («ящик с усами», box-plot), при этом каждой группе соответствует определенный «ящик».

7.1.5 Сравнение относительных показателей между двумя группами.

1 Построить и заполнить четырехпольную таблицу, строки в которой представляют значения фактора, а столбцы – значения исхода.

2 Рассчитать значения четырех ожидаемых явлений, представляющих собой результаты деления произведений сумм столбцов и ячеек на общую сумму всех ячеек:

$$(a + b) \cdot (a + c) / (a + b + c + d);$$

$$(a + b) \cdot (b + d) / (a + b + c + d);$$

$$(a + c) \cdot (c + d) / (a + b + c + d);$$

$$(b + d) \cdot (c + d) / (a + b + c + d).$$

3 Определить наименьшее значение ожидаемого явления (ОЯ) из четырех, выбрать метод анализа:

– если наименьшее значение ОЯ менее 5, использовать для сравнения точный критерий Фишера;

– если наименьшее значение ОЯ находится в интервале от 5 до 10, использовать для сравнения критерий Пирсона с поправкой на непрерывность Йейтса;

– если наименьшее значение ОЯ более 10, использовать для сравнения критерий Пирсона.

4 Для количественной оценки зависимости вероятности исхода от наличия фактора рассчитать показатель отношения шансов с 95-процентным доверительным интервалом.

5 Описать результаты сравнения исследуемых групп, указав вначале абсолютные и относительные значения частоты явления для каждой группы. Отметить, в какой из сравниваемых групп частота явления в процентах выше.



Указать уровень значимости различий p , определенный с помощью одного из критериев, указанных в п. 3. Отметить, во сколько раз отличается вероятность наступления исхода в одной группе по сравнению с другой с помощью отношения шансов (п. 4). В качестве графического изображения для наглядного представления результатов следует использовать столбиковую диаграмму.

7.1.6 Сравнение относительных показателей в динамике при двух-этапном измерении.

1 Построить и заполнить четырехпольную таблицу, строки в которой представляют этапы наблюдения (до – после), а столбцы – значения анализируемого явления.

2 Для сравнения нужно использовать критерий МакНемара (тест МакНемара). Оценить статистическую значимость различий показателей, сравнив рассчитанное значение критерия с критическим или определив уровень значимости p с помощью статистической программы.

3 Описать результаты сравнения исследуемых групп, указав вначале абсолютные и относительные значения частоты явления для каждого этапа наблюдения. Отметить, увеличилась или уменьшилась частота явления в результате эксперимента. Указать уровень значимости различий p , определенный с помощью критерия МакНемара. В качестве графического изображения для наглядного представления результатов следует использовать столбиковую диаграмму.

7.1.7 Выявление взаимосвязи между двумя количественными показателями.

1 Оценить нормальность распределения обоих показателей с помощью критерия Колмогорова-Смирнова (при числе исследуемых $n > 50$) или критерия Шапиро-Уилка (при $n < 50$). Если распределение нормальное, использовать алгоритм, описанный в пп. 2–4, если отличается от нормального – алгоритм, изложенный в п. 2–7.

2 Для оценки связи между двумя показателями необходимо использовать коэффициент корреляции r_{xy} Пирсона. Исходя из полученного значения r_{xy} определить тесноту связи по шкале Чеддока и ее направление (прямая или обратная).

3 Рассчитав величину t -критерия и сравнив его значение с критическим значением по таблице или с применением специальной статистической программы, найти уровень значимости p корреляционной связи.

4 Описать корреляционную связь:

– если связь является статистически значимой, то указывается значение r_{xy} Пирсона, теснота связи по шкале Чеддока (весьма высокая – высокая – заметная – умеренная – слабая), направление связи (прямая или обратная). Указывается уровень значимости p ;

– если связь статистически не значима, помимо констатации данного факта, указывается только уровень значимости.

5 Для оценки связи между двумя показателями, распределение которых отличается от нормального, следует использовать коэффициент ранговой кор-



реляции r_{xy} Спирмена. Исходя из полученного значения r_{xy} определить тесноту связи по шкале Чеддока и ее направление (прямая или обратная).

6 Рассчитав величину t -критерия и сравнив его значение с критическим по таблице или с применением специальной статистической программы, найти уровень значимости p корреляционной связи.

7 Описать корреляционную связь:

– если связь является статистически значимой, то указывается значение r_{xy} Пирсона, теснота связи по шкале Чеддока (весьма высокая – высокая – заметная – умеренная – слабая), направление связи (прямая или обратная). Указывается уровень значимости p ;

– если связь статистически не значима, помимо констатации данного факта, указывается только уровень значимости.

7.2 Порядок выполнения работы

- 1 Изучить методические рекомендации.
- 2 Получить исходные данные для выполнения работы у преподавателя.
- 3 На основе примера разработать свой алгоритм анализа.
- 4 Разработать алгоритм и провести расчетный эксперимент.
- 5 Оформить отчет.

Содержание отчета

- 1 Цель работы.
- 2 Постановка задачи и исходные данные.
- 3 Схема алгоритма и листинг программной процедуры.
- 4 Результаты выполнения программы анализа данных.
- 5 Выводы по оценке результатов моделирования.

Контрольные вопросы

- 1 Последовательность анализа данных.
- 2 Как определить тип признака?
- 3 Чем отличаются количественный и качественный признаки?
- 4 Как выявляется взаимосвязь между двумя количественными показателями?



Список литературы

1 **Горохов, В. А.** Основы экспериментальных исследований и методика их проведения : учебное пособие / В. А. Горохов. – Минск ; Москва : Новое знание ; ИНФРА-М, 2016. – 655 с. : ил.

2 **Карманов, Ф. И.** Статистические методы обработки экспериментальных данных. Лабораторный практикум с использованием пакета MathCad : учебное пособие / Ф. И. Карманов, В. А. Острейковский. – Москва : Высшая школа ; Абрис, 2012. – 208 с. : ил.

3 Алгоритмы статистического анализа данных [Электронный ресурс]. – Режим доступа: <http://medstatistic.ru/algorithm.html>. – Дата доступа: 26.06.2018.

