

Федеральное агентство по образованию  
Федеральное Государственное бюджетное образовательное учреждение  
высшего профессионального образования  
Национальный исследовательский университет  
ресурсоэффективных технологий  
«ТПУ»

---

УТВЕРЖДАЮ  
Зав. каф. Промышленной  
и медицинской электроники  
проф., д-р техн. наук  
\_\_\_\_\_ Г.С. Евтушенко

## ЦИФРОВОЙ АНАЛИЗ СЛУЧАЙНЫХ ПРОЦЕССОВ

Методические указания к лабораторным работам  
по дисциплине "Методы обработки медико-биологических данных"

Томск - 2009

УДК 628.931

Цифровой анализ случайных процессов

Методические указания к лабораторным работам по дисциплине "Методы обработки биомедицинских сигналов и данных" для бакалавров по направлению 200300 «Биомедицинская инженерия»

Томск: Изд. ТПУ, 2009. – 34 с.

Составитель    ст. преп. Голованова И.С.

Рецензент                    к. т. н. доц. каф. ПМЭ А.И. Солдатов

Методические указания рассмотрены и рекомендованы к изданию методическим семинаром кафедры промышленной и медицинской электроники 14 декабря 2009 г.

Зав кафедрой

проф., д-р техн. наук \_\_\_\_\_ Г.С. Евтушенко

## ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

### ФИЗИЧЕСКИЕ ПРОЦЕССЫ

Физический процесс – это изменение физической величины во времени. В зависимости от характера изучаемой величины физический процесс может быть детерминированным или случайным.

Детерминированный процесс можно описать точным математическим выражением, позволяющим однозначно определить значение изучаемой величины в любой момент времени.

Случайный (стохастический) процесс – процесс, для которого характерно изменение изучаемой величины во времени случайным образом. Его нельзя описать аналитическим выражением, и, следовательно, нельзя точно предсказать его значение в определенный момент времени.

На практике заключение о том, является ли процесс случайным или детерминированным, делается на основании результатов его воспроизведения: если при проведении нескольких опытов результат наблюдений повторяется в пределах ошибки измерения, то процесс считается детерминированным, если нет – случайным.

Когда процесс в биосистеме регистрируется при помощи электронно-медицинской аппаратуры, то результатом наблюдения служит непрерывный сигнал (кардиограмма, реограмма и т.д.). В этом смысле биомедицинский процесс и биомедицинский сигнал взаимозаменяемы.

Когда процесс в биосистеме регистрируется иначе, например, при помощи ряда измерений какого-то показателя (артериального давления, температуры и т.д.) или по результатам опросов, то он представлен набором дискретных значений.

Непрерывный сигнал тоже можно преобразовать в набор дискретных значений, если фиксировать его в определенные моменты времени, например с некоторым постоянным интервалом.

Результат *однократного* наблюдения за процессом или однократный съем сигнала в течение определенного времени называется *реализацией процесса или временным рядом* и обозначается  $x(t)$ .

Провести наблюдения можно много раз, получив, таким образом, множество реализаций, причем каждая конкретная реализация является элементом этого множества.

### КЛАССИФИКАЦИЯ ФИЗИЧЕСКИХ ПРОЦЕССОВ

Классификация физических процессов необходима потому, что в зависимости от класса

выбирается математический аппарат обработки процесса. Детерминированные процессы делятся на классы, изображенные на Рис.1.



Рис. 1. Классификация детерминированных процессов

Гармонические процессы: реализация процесса описывается гармонической функцией времени:  $x(t) = A \cdot \sin(\omega t + \varphi)$ ,

где  $A$  – амплитуда,  $\omega$  – частота,  $\varphi$  – фаза,  $t$  – время.

Общие периодические процессы: реализация процесса описывается функцией времени, значения которой точно повторяются через промежуток  $T_0$ :  $x(t) = f(t + n T_0)$ .

Полигармонические процессы: реализация процесса описывается функцией времени, представляющей собой сумму синусоидальных волн с разными частотами, амплитудами и фазами:  $x(t) = \sum_{n=1}^{\infty} A_n \sin(\omega_n t + \varphi_n)$ .

Переходные процессы: реализация процесса описывается непериодической функцией времени:  $x(t) = A \cdot e^{-bt}$ ,  $x(t) = A \cdot e^{-at} \cos(bt)$ ,  $x(t) = \text{const}$ .

Классификация случайных процессов, изображенная на Рис. 2., намного сложнее, так как их поведение описывается большим числом характеристик.

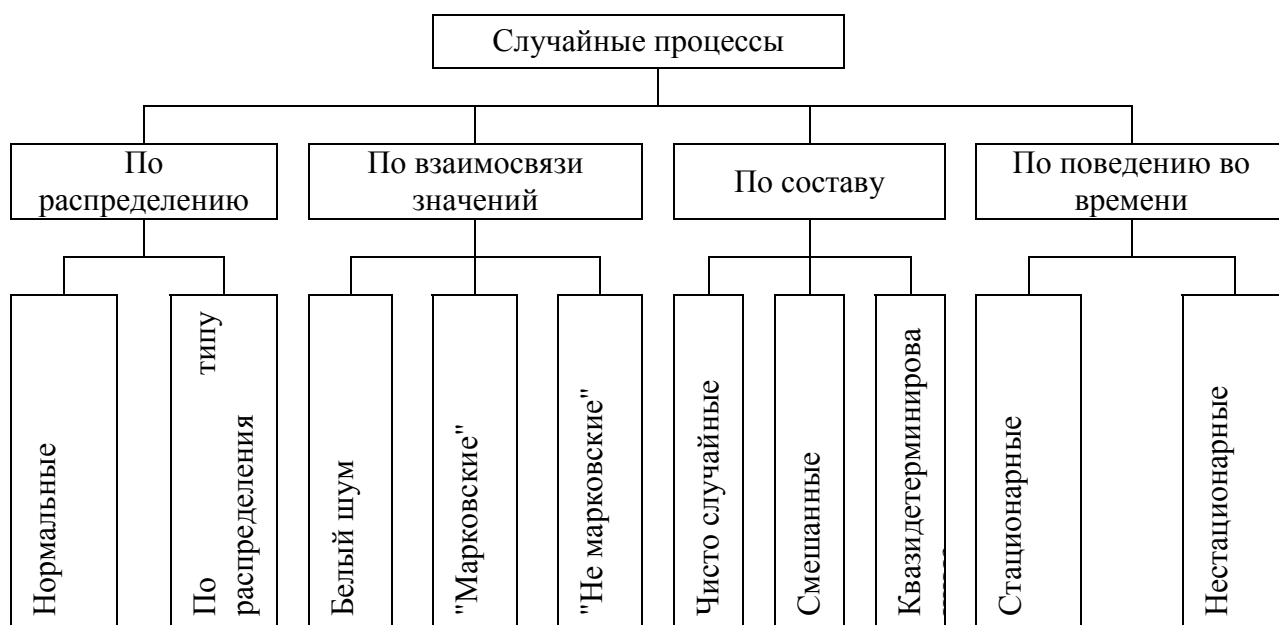


Рис. 2. Классификация случайных процессов

В зависимости от типа распределения случайные процессы делятся на:

- нормальные (гауссовы), распределение которых подчиняется нормальному закону,
- другие – по типу закона распределения (релеевские, равномерные, биномиальные, Пуассона и т.д.)

В зависимости от взаимосвязи значений процесса в разные моменты времени случайные процессы делятся на:

- совершенно случайные процессы ("белый шум"), значения которых не зависят друг от друга,
- "марковские", в которых значения процесса связаны только с непосредственно предшествующими и не связаны с другими (например, распад радиоактивного вещества, где вероятность распада частицы не зависит от предшествующего течения процесса),
- "не марковские", в которых значения процесса взаимосвязаны

В зависимости от наличия или отсутствия детерминированных составляющих случайные процессы делятся на:

- чисто случайный процесс, в котором отсутствуют детерминированные составляющие,
- смешанный процесс, представляющий собой сумму детерминированной и случайной составляющих (сигнал + шум),
- квазидетерминированный процесс, который описывается аналитической функцией времени, содержащей один или более случайных параметров:  $x(t) = A \sin(\omega \cdot t + \varphi_{случ.})$ , где  $A$  – амплитуда и  $\omega$  - частота постоянны, а фаза  $\varphi$  меняется случайным образом.

В зависимости от сдвига начала отсчета во времени случайные процессы делятся на:

- стационарные, у которых параметры распределения не зависят от сдвига начала отсчета по времени, (то есть если изменить начало отсчета с  $t_0$  на  $t_0 + \Delta t$ , то параметры распределения – математическое ожидание  $\mu$ , дисперсия  $\sigma^2$  и др. не изменятся),
- нестационарные, у которых параметры распределения являются функциями времени:  $\mu = \mu(t)$ ,  $\sigma^2 = \sigma^2(t)$ .

Стационарные случайные процессы обладают свойством *эргодичности*. Эргодическими называются такие случайные процессы, у которых параметры распределения, полученные усреднением по ансамблю реализаций равны параметрам распределения, полученным по одной, но достаточно длинной реализации (Рис. 3.).

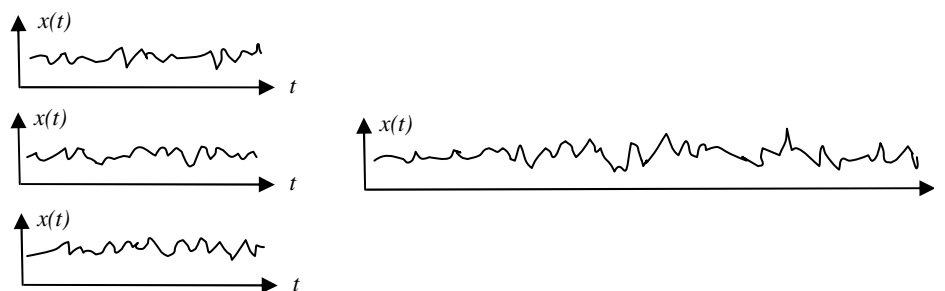


Рис. 3. Эргодические процессы

Нестационарные случайные процессы имеют различные виды нестационарности (Рис. 4.).



Рис. 4. Виды нестационарности случайных процессов.

Большой класс случайных процессов в том числе и биологических сигналов можно представить в виде суммы:  $x(t) = m(t) + a(t) \cdot y(t)$ , где  $x(t)$  - реализация (амплитуда) сигнала;  $m(t)$  - детерминированная составляющая;  $a(t)$  - детерминированный множитель;  $y(t)$  - реализация *стационарного* случайного процесса.

Когда  $a(t) = 1$ , а  $m(t)$  детерминированная функция времени, случайный процесс будет *нестационарным по среднему значению*. Например, если  $m(t)$  прямо-пропорционально  $t$ , процесс может иметь вид Рис. 5.

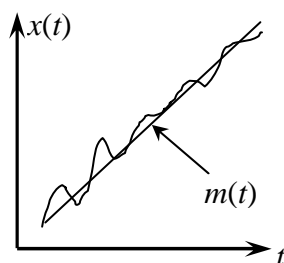


Рис. 5. Сигнал, нестационарный по среднему значению.

Если рассматривать  $m(t)$  как сигнал, а  $y(t)$  - как помеху, то такая помеха называется *аддитивной*.

Когда  $m(t) = \text{const}$ , а  $a(t)$  - детерминированная функция времени, случайный процесс будет *нестационарным по дисперсии*. Если рассматривать  $m(t)$  как сигнал, а  $y(t)$  - как помеху, такая помеха называется *мультипликативной*. (Рис. 6.)

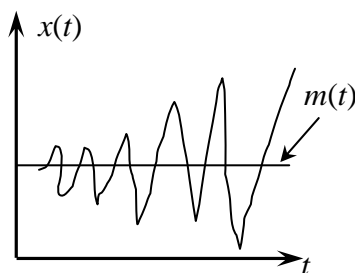


Рис. 6. Сигнал, нестационарный по дисперсии.

Интересно периодическое изменение дисперсии.

Примером может служить ЭКГ – систолические шумы с периодичностью сердечных сокращений (Рис. 7.).



Рис. 7. Электрокардиограмма.

Примером более сложных видов нестационарности сигнала может служить энцефалограмма человека, который следит за появлением случайной метки на экране и должен отреагировать на эту метку нажатием клавиши при прохождении теста на реакцию или координацию.

Для процессов разных классов применяются различные методы обработки. Так как медико-биологическая информация носит всегда до некоторой степени случайный характер, то к ней применяются методы обработки случайных процессов.

## ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ СИГНАЛОВ.

Процессы, протекающие в биосистемах, могут отображаться и фиксироваться электронно-медицинской аппаратурой в виде биомедицинских сигналов. Для их изучения необходимо знать вероятностно-статистические характеристики, играющие важную роль при анализе случайных сигналов.

К ним относятся: функция распределения, плотность распределения, матожидание, дисперсия, автокорреляционная функция и спектральная плотность.

Функция распределения вероятностей случайного сигнала  $F(x)$ , представляет собой вероятность того, что значения сигнала попадут в определённый интервал (амплитуд).

Для непрерывного случайного сигнала:  $F(x) = P(x < x(t) < x + \Delta x) = \lim_{T \rightarrow \infty} \frac{T_x}{T}$ ;

где  $T_x$  - сумма интервалов времени, в течение которых амплитуда сигнала находилась в диапазоне от  $x$  до  $x + \Delta x$ ,  $T_x = \sum_{i=1}^k \Delta t_i$ , здесь  $k$  - количество интервалов;  $\Delta t_i$  - длина  $i$ -того интервала (например, в секундах),  $T$  - время снятия сигнала (Рис. 8.). Для дискретного

временного ряда  $F(x) = \frac{M}{N}$ ; где  $M$  - число дискретных значений, попадающих в интервал  $[x; x + \Delta x]$ , а  $N$  - длина реализации или общее число наблюдений.

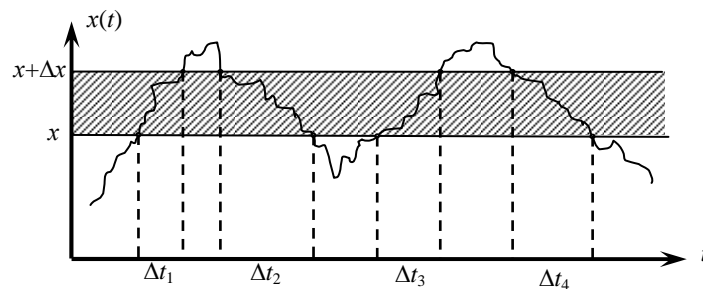


Рис. 8. Реализация непрерывного случайного сигнала

Плотность распределения случайного сигнала  $f(x)$  является вероятностью того, что амплитуда сигнала примет конкретное значение. Для непрерывного сигнала:

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x)}{\Delta x} = F'(x).$$

Для дискретного временного ряда  $f(x) = \frac{P}{N}$ , где  $P$  - количество наблюдений, равных заданному числу,  $N$  - длина реализации или общее число наблюдений.

Биомедицинские сигналы всегда несут элемент случайности и, следовательно, могут быть представлены в виде суммы статической (непосредственно сигнал) и динамической (помеха) составляющих.

Статическую составляющую или сам сигнал характеризует матожидание  $\mu$ , а динамическую составляющую - помеху - характеризует дисперсия  $\sigma^2$ .

Матожидание и дисперсия непрерывного случайного процесса вычисляется

соответственно по формулам:  $\mu = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt$ ;  $\sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x(t) - \mu]^2 dt$ .



## ЦИФРОВОЙ АНАЛИЗ МЕДИКО-БИОЛОГИЧЕСКИХ СИГНАЛОВ

В общем случае анализ медико-биологических данных включает в себя 4 этапа: сбор данных; регистрацию данных; подготовку данных и обработку данных.

*Сбор данных* – это, по сути, преобразование одного вида энергии в другой при помощи специального устройства – *преобразователя*. Такое преобразование позволяет количественно выразить изучаемый процесс, преобразуя его в аналоговый или цифровой сигнал.

Например, при съеме ЭКГ энергия мышечных сокращений сердца преобразуется в энергию биопотенциалов, которые усиливаются и регистрируются электрокардиографом в виде электрического напряжения.

В реальной практике преобразование энергии никогда не происходит без искажений, так как преобразователь всегда является источником помех.

*Регистрация* это запоминание данных на различных носителях. Регистрация не нужна, если обработка идет в реальном масштабе времени и необходима, если обработка идет в отложенном времени.

*Подготовка* данных включает в себя: сглаживание сигнала, его дискретизацию, квантование и форматирование.

*Сглаживание* сигнала – это исключение случайных выбросов и помех визуально или программно при помощи аналоговой или цифровой фильтрации.

*Дискретизация* сигнала – это представление непрерывного сигнала  $x(t)$  в виде набора дискретных отсчетов  $x_i = x(t_i)$ , снятых в моменты времени  $t_i$  через одинаковый интервал  $\Delta t$ , который называется *интервалом дискретности* или *шагом дискретизации*:  $t_i = t_0 + i \cdot \Delta t$ ;  $i = 0, 1, 2, \dots, N-1$ ;  $t_0$  – начало отсчета, которое выбирается произвольно,  $N$  – длина реализации (Рис. 9).

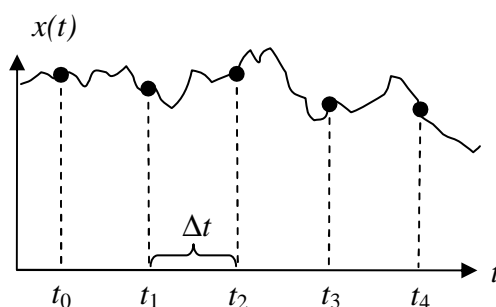


Рис. 9. Дискретизация сигнала

Необходимо правильно выбирать шаг дискретизации, так как если он слишком мал, то значения могут коррелировать между собой, а количество их будет избыточно велико,

если же он слишком велик, то возможно смешивание высоко- и низкочастотных составляющих – явление известное как "маскировка частот".

Поэтому шаг дискретизации принимают равным  $\Delta t = \frac{1}{2f_c}$ , где  $f_c$  – частота, называемая *частотой среза* (или частотой Найквиста) – максимально возможная частота в спектре сигнала. Например, для ЭКГ частота среза равна 250 Гц, тогда шаг дискретизации  $\Delta t = \frac{1}{500 \text{ Гц}} = 0,002 \text{ сек.} = 2 \text{ милсек.}$

На практике частоту среза либо выбирают из физических соображений, принимая ее такой, что существование в сигнале составляющих с более высокими частотами просто не возможно, либо отфильтровывают все составляющие с частотами выше некоторой определенной частоты, которую и принимают за частоту среза.

*Квантование* биомедицинских сигналов производится при помощи аналого-цифрового преобразователя (АЦП), который преобразует амплитуду аналогового сигнала в безразмерную цифровую форму. При этом нужно так выбирать АЦП, чтобы его шкала квантования (от 0 до 256) соответствовала размаху амплитуд сигнала.

*Форматирование* это до некоторой степени обратная процедура, когда безразмерные результаты вычислений переводятся в размерные единицы. Обычно форматирование осуществляется программно путем домножения на масштабный коэффициент.

*Обработка* биомедицинских сигналов и данных начинается с построения распределения вероятностей экспериментальной реализации.

Чтобы построить распределение, нужно найти размах наблюдаемых значений  $x_{\max} - x_{\min}$ , разбить его на  $n$  интервалов и подсчитать, сколько значений попадает в каждый интервал, то есть подсчитать *частоты распределения*.

Длина интервала определяется по формуле Стерджесса:  $k = \frac{x_{\max} - x_{\min}}{1 + 3,322 \ln N}$ , где  $N$  – количество наблюдений или длина реализации.

Количество интервалов равно:  $n = \frac{x_{\max} - x_{\min}}{k} = 1 + 3,322 \ln N$ .

Границы интервалов вычисляются по формуле:  $d_j = x_{\min} + j \cdot k$ , где  $j = 0, 1, 2, \dots, n$ .

Частоты распределения  $m_j$  (или количество попаданий в  $j$ -й интервал) рассчитывают из условия:  $d_{j-1} < x_i \leq d_j$ .

Тогда  $m_0$  – количество значений, удовлетворяющих неравенству:  $x_i \leq d_0$ ,

$m_1$  – количество значений, удовлетворяющих неравенству:  $d_0 < x_i \leq d_1$ ,  
и так далее ...  
 $m_n$  – количество значений, удовлетворяющих неравенству:  $d_{n-1} < x_i \leq d_n$ .

Сумма частот равна общему количеству наблюдений:  $\sum_{j=0}^n m_j = N$ .

График частот распределения  $m_j$ , построенный в виде столбчатой диаграммы, называют *гистограммой распределения*. Сами частоты относят к серединам интервалов, так что нулевая частота отнесена к середине интервала, выходящего за минимальное значение. (Рис. 10.). На этом же графике в виде ломаной линии изображены накопленные частоты  $M_j$ , где  $j = 0, 1, 2, \dots, n$ .

*Накопленной частотой  $M_j$*  называется сумма частот от начала распределения до  $j$ -того интервала, так что последняя накопленная частота равна  $N$ .

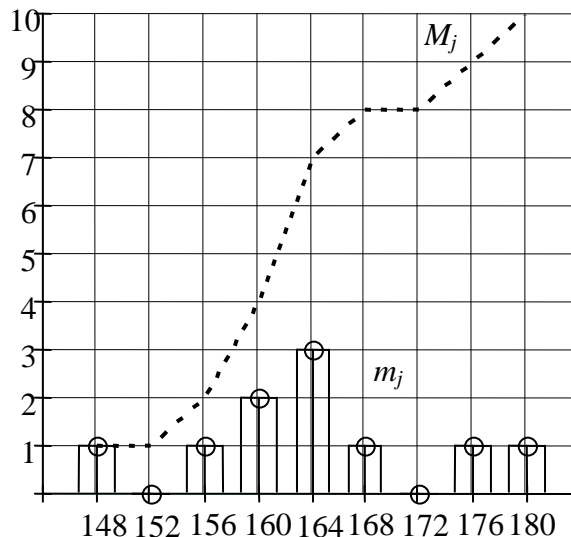


Рис. 10. Гистограмма распределения и накопленные частоты

Если нормировать частоты распределения и накопленные частоты, разделив их на сумму частот, равную объему выборки, то получим *плотность распределения* в виде частостей - дискретных вероятностей и функцию распределения в виде накопленных частостей. Качественный вид гистограммы при этом не изменится.

Построение распределения позволяет вычислить его статистики или выборочные параметры, такие как матожидание и дисперсия.

## ОЦЕНКА ОСНОВНЫХ СВОЙСТВ СЛУЧАЙНЫХ ПРОЦЕССОВ

К основным свойствам случайного процесса (сигнала) относятся:

- "нормальность" – согласованность с нормальным распределением,
- *стационарность* – постоянство параметров распределения во времени,

– "периодичность" – наличие или отсутствие в сигнале периодических составляющих.

В зависимости от этих свойств выбирается математический аппарат для анализа сигнала.

Схема этой оценки приведена на Рис. 11.

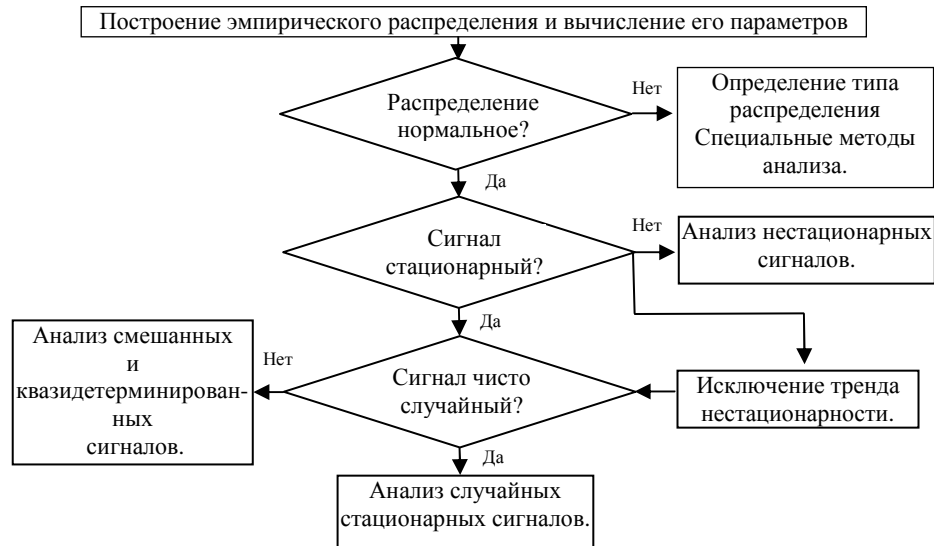


Рис. 11. Схема оценки основных свойств случайного процесса

### ОЦЕНКА «НОРМАЛЬНОСТИ»

Оценка основных свойств случайного процесса начинается с проверки согласованности эмпирического распределения с нормальным законом распределения вероятностей (закон Гаусса – Лапласа).

Плотность нормального закона распределения описывается формулой:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{или} \quad f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad \text{где} \quad t = \frac{x-\mu}{\sigma} \quad \text{называется}$$

центрированным и нормированным отклонением.

Функция распределения вычисляется как интеграл от плотности распределения и называется *интегралом вероятностей*:  $F(t) = \int_{-\infty}^t f(t)dt$ .

Графически функция и плотность распределения нормального закона приведены на (Рис. 12.).

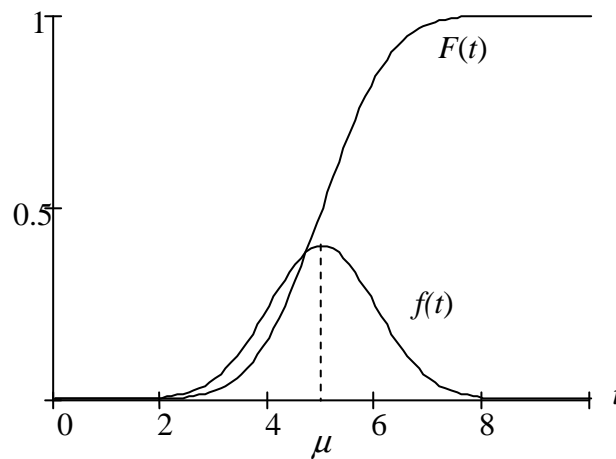


Рис. 12. Функция и плотность нормального закона распределения.

Очень многие природные явления распределены по нормальному закону. Это объясняет его физический смысл, который состоит в том, что под воздействием большого числа независимых случайных факторов (что и происходит в природе) случайная величина подчиняется нормальному закону распределения.

Согласие эмпирического распределения с нормальным законом распределения проверяется при помощи гипотез согласия.

Для их проверки используются критерии, которые называются *критериями согласия*. Критерии согласия бывают *параметрическими* и *непараметрическими*.

*Параметрические* критерии согласия зависят от параметров эмпирического распределения и требуют его обязательного построения.

Параметрические критерии дают надежные результаты, когда объем выборки не меньше 50, а эмпирические частоты не меньше 5.

Самым распространённым параметрическим критерием согласия является *критерий Пирсона*  $\chi^2$  ("хи-квадрат"). Его достоинство заключается в обеспечении высокой точности результатов при проверке гипотезы, однако этот критерий достаточно трудоёмок в вычислениях, и в этом – его главный недостаток.

Критерий Пирсона вычисляется по формуле:  $\chi^2 = \sum_{j=0}^n \frac{(m_j - m'_j)^2}{m_j}$ , где  $n$  – число интервалов распределения,  $m_j$  – эмпирические частоты,  $m'_j$  – теоретические или *выравнивающие* частоты:  $m'_j = k \sum_{j=0}^n m_j f(t_j)$ , где  $t_j = \frac{x_j - \bar{x}}{S}$ ,

$x_j$  – варианты –середины интервалов распределения,  $\bar{x}$  – выборочное среднее

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot m_i}{N}; \quad S = \frac{D \cdot N}{N-1}, \quad D - \text{выборочная дисперсия } D = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot m_i}{N};$$

$k$  – величина интервала распределения,  $N$  – объем выборки и  $f(t_j)$  – плотность нормального закона распределения, которая определяется из таблиц для каждого  $t_j$ .

Очень часто для удобства вычислений при цифровом анализе исходную реализацию представляют в виде центрированных и нормированных отклонений. Процесс центрирования – это приведение к нулевому среднему значению, так как очевидно, что среднее значение преобразованной реализации будет равно 0. Аналогично процесс нормировки – это приведение к единичной дисперсии, так как дисперсия преобразованной реализации равна 1.

Из таблиц распределения  $\chi^2$  в зависимости от уровня значимости  $\alpha$  и числа степеней свободы  $\nu = n-3$ , находят критическое значение критерия  $\chi_{\alpha,\nu}^2$  и принимают решение: если  $\chi^2 \leq \chi_{\alpha,\nu}^2$ , то гипотеза принимается и эмпирическое распределение согласуется с нормальным законом на уровне значимости  $\alpha$ .

Уровень значимости  $\alpha$  характеризует степень риска, то есть вероятность

принятия ошибочного решения при проверке гипотезы. Он задается произвольно в зависимости от практической ценности проверяемой гипотезы. Уровень значимости и доверительная вероятность (вероятность правильного решения при проверке гипотезы) связаны соотношением:  $\alpha = 1 - P_d$ .

Критерий Романовского вычисляется по формуле:  $R = \frac{\chi^2 - \nu}{2\sqrt{\nu}}$ ;

где  $\nu = N - 1$  число степеней свободы критерия.

Если  $R < 3$ , гипотеза принимается, а если  $R \geq 3$  гипотеза отвергается на заданном уровне значимости.

Критерий Колмогорова вычисляется по формуле:  $\lambda = \frac{|\Delta|}{\sqrt{N}}$ , где  $\Delta$  – максимальная

разность эмпирических  $M_j$  и выравнивающих  $M'_j$  накопленных частот. В зависимости от значений  $\lambda$  составлены таблицы вероятности, с которой принимается гипотеза согласия  $P(\lambda)$ . Для получения надежных результатов по критерию Колмогорова необходимо, чтобы объем выборки был не менее 100.

Непараметрические критерии согласия не требуют построения распределения, кроме этого они пригодны для относительно небольших объёмов выборки, что очень ценно для медицинских наблюдений.

Один из таких критериев - критерий Шапиро-Уилка.

Проверка гипотезы согласия при помощи критерия Шапиро-Уилка состоит из следующих этапов:

1) значения случайного процесса  $x_i$ , (где  $i = 1, 2, \dots, N$ ) располагаются по возрастанию, то есть строится *ранжированная выборка*;

2) вычисляется критерий Шапиро-Уилка  $V = \frac{b^2}{g}$ , где  $b = \sum_{i=1}^k a_i \cdot (x_{N-i+1} - x_i)$ ,

3)  $g = \sum_{i=1}^N x_i^2 - \frac{1}{N} \cdot \left( \sum_{i=1}^N x_i \right)^2$ ,  $k = \frac{N}{2}$  при четном  $N$  и  $k = \frac{(N-1)}{2}$  при нечетном  $N$ .

Коэффициенты  $a_i$  - фиксированные эмпирические константы, протабулированные в зависимости от  $N$  (Таблица 1.).

4) значение критерия  $V$  сравнивается с критическим значением  $V_{P_\delta, N}$ , которое определяется из графика, приведенного на Рис. 13, в зависимости от доверительной вероятности  $P_\delta$  и длины реализации  $N$ .

5) по результатам сравнения принимается решение: если  $V \leq V_{P_\delta, N}$ , то случайный процесс подчиняется нормальному закону распределения с доверительной вероятностью  $P_\delta$ .

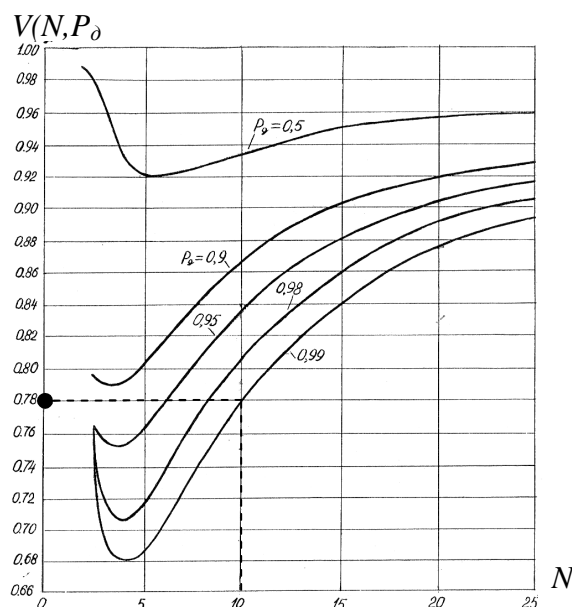


Рис. 13. Критические значения критерия Шапиро-Уилка

Таблица 1.

Таблица для коэффициентов критерия Шапиро – Уилка .

$a_i$	$N$ – длина реализации											
$i$	10	11	12	13	14	15	16	17	18	19	20	21
1	0,574	0,560	0,547	0,536	0,525	0,515	0,506	0,497	0,489	0,491	0,473	0,464
2	0,329	0,331	0,332	0,332	0,332	0,330	0,329	0,327	0,325	0,323	0,321	0,318
3	0,214	0,226	0,235	0,241	0,246	0,249	0,252	0,254	0,255	0,256	0,256	0,258
4	0,122	0,143	0,158	0,170	0,180	0,188	0,194	0,199	0,203	0,206	0,208	0,212
5	0,040	0,069	0,092	0,110	0,124	0,135	0,145	0,152	0,159	0,164	0,169	0,174
6			0,030	0,054	0,073	0,088	0,100	0,111	0,120	0,127	0,133	0,140
7					0,024	0,043	0,059	0,072	0,084	0,093	0,101	0,109
8							0,020	0,036	0,050	0,061	0,071	0,080
9									0,016	0,030	0,042	0,053
10											0,014	0,026

## ОЦЕНКА СТАЦИОНАРНОСТИ

В отдельных случаях стационарность можно оценить из физических соображений. В иных случаях для проверки гипотезы стационарности используются непараметрические критерии, такие как *критерий серий (Вальда-Вольфовица)* или *критерий инверсий*.

Критерий серий.

Определим для исходной реализации  $x_i = x(t_i)$  среднее значение  $\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^N x_i$

и сравним с ним каждое  $x_i$ .

Если  $x_i \geq \bar{x}$ , то отнесем  $x_i$  к одной категории, условно обозначенной "+1".

Если  $x_i < \bar{x}$ , то отнесем его к противоположной категории, обозначенной "-1".

Таким образом, каждое значение  $x_i$  будет отнесено к одной из противоположных категорий: +1 или -1. В результате вместо исходной реализации получим последовательность, состоящую из нескольких серий (Рис. 14):

*Серией называется группа подряд идущих значений, до и после которой расположены значения противоположной категории или таковые отсутствуют.* На Рис.14. количество серий равно четырем.

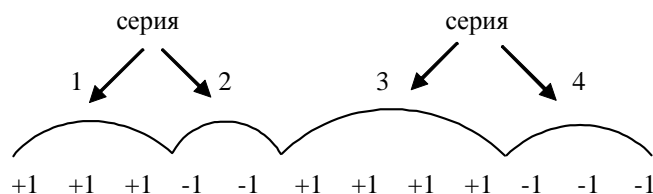


Рис. 14. Последовательность серий

Количество серий служит критерием проверки гипотезы стационарности. Если процесс стационарный, то естественно предположить, что количество отклонений от среднего со знаком "+" равно количеству отклонений со знаком "-" и равно  $N/2$ . Поэтому



подсчитывают количество серий  $r$  и сравнивают его с критическими значениями, которые определяется из таблиц для заданного уровня значимости  $\alpha$  и  $N/2$ :  $r_{1-\alpha, \frac{N}{2}}$  и  $r_{\alpha, \frac{N}{2}}$ .

Таблица 3

Критические значения распределения серий

N/2	$\alpha$					
	0,99	0,975	0,95	0,05	0,025	0,01
5	2	2	3	8	9	9
6	2	3	3	10	10	11
7	3	3	4	11	12	12
8	4	4	5	12	13	13
9	4	5	6	13	14	15
10	5	6	6	15	15	16
11	6	7	7	16	16	17
12	7	7	8	17	19	18
13	7	8	9	18	9	20
14	8	9	10	19	20	21
15	9	10	11	20	21	22
16	10	11	11	22	22	23
18	11	12	13	24	55	26
20	13	14	15	26	27	28
25	17	18	19	32	33	34
30	21	22	24	37	39	40
35	25	27	28	43	44	46
40	30	31	33	48	50	51
45	34	36	37	54	55	57
50	38	40	42	59	61	63

Использование критерия инверсий более эффективного, чем критерий серий, основано на понятии инверсии – появления в реализации за большим значением меньшего.

Критерий инверсий вычисляется по формуле:  $z = \frac{\gamma}{\sqrt{u}}$ , где  $\gamma = 1 - \frac{4I}{N(N-1)}$ ,  $I$  – количество инверсий,  $N$  – длина реализации,  $u$  вычисляется по эмпирической формуле:  $u = \frac{2(2N-5)}{9N(N-1)}$ . В зависимости от уровня значимости  $\alpha$  из таблиц интеграла вероятностей определяют критическое значение  $z_{\alpha}$  и проверяют гипотезу, сравнивая с ним значение  $z$ .  
Если  $|z| < z_{\alpha}$  то гипотеза стационарности принимается на уровне значимости  $\alpha$ .

## ОЦЕНКА ПЕРИОДИЧНОСТИ

В тех случаях, когда имеются какие либо предварительные сведения о физике процесса, отражаемого экспериментальным сигналом, присутствие периодической составляющей в сигнале можно выявить из физических соображений.

Кроме этого, если амплитуда периодической составляющей в сигнале больше, чем амплитуда случайной составляющей, то ее наличие очевидно, и ее выявление не требует никаких специальных приемов.

Если же предварительные сведения отсутствуют, и амплитуда периодической составляющей сравнима с амплитудой случайной составляющей, то такие приемы необходимы. В частности, можно выявить наличие периодической составляющей по виду эмпирического распределения (Рис. 15.).

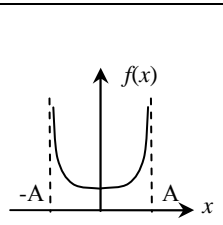
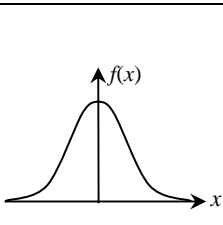
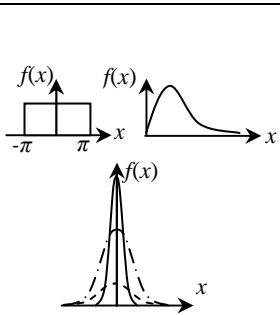
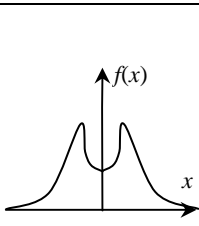
Гармоническое колебание $x(t) = A \sin(\omega \cdot t)$	Идеальный "белый шум"	Реальный чисто случайный сигнал	Смешанный сигнал
			
$f(x) = \begin{cases} \frac{1}{\pi\sqrt{A^2 - x^2}} &  x  < A \\ 0 &  x  \geq A \end{cases}$	Нормальный закон распределения $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Различный вид плотности распределения.	Вид плотности распределения.

Рис. 15. Плотность распределения различных сигналов

Однако наличие периодической составляющей будет заметно на распределении, если ее амплитуда все-таки сравнительно велика.

Более эффективным способом проверки гипотезы периодичности случайного сигнала является подсчет его *автокорреляционной функции (А.К.Ф.)*

Автокорреляционная функция описывает поведение случайного процесса во временной области.

Она показывает зависимость значений случайного сигнала в момент времени  $t$  от его значения в момент времени  $t+\tau$ , где  $\tau$  – сдвиг по времени.

АКФ вычисляется по формуле:  $R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T - \tau} \int_0^{T-\tau} x(t) \cdot x(t + \tau) dt$  и представляет

собой для непрерывных действительных случайных сигналов четную функцию, имеющую максимум при  $\tau = 0$ :

$$R(\tau) = R(-\tau), R(0) > |R(\tau)| \text{ при любом } \tau \neq 0.$$

При оценке периодичности чаще всего применяют *нормированную автокорреляционную функцию*,

$$\rho(\tau) = \frac{R(\tau)}{R(0)}, \text{ которая также является четной и имеет максимум, равный 1, при } \tau=0.$$

Для получения достоверных результатов желательно, чтобы соблюдалось соотношение:  $\tau \approx \frac{T}{10}$ . График АКФ. называют *автокоррелограммой*.

Выявить наличие периодичности можно по виду автокоррелограммы (Рис. 16.).

Для чисто периодического сигнала гармонического колебания вида  $x(t) = A \sin(\omega \cdot t)$  А.К.Ф. имеет косинусоидальный вид.

Для идеального белого шума она имеет вид  $\delta$ -функции Дирака.

Для реальных чисто случайных процессов, значения А.К.Ф. довольно быстро спадают к нулю уже при малых  $\tau$

Для процессов, в которых присутствует периодические составляющие, значения А.К.Ф. не прекращают осциллировать при сколь угодно больших  $\tau$ .

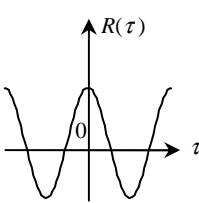
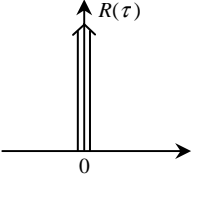
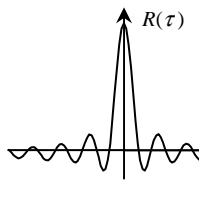
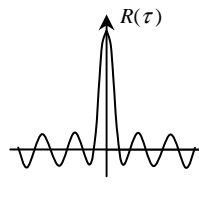
Гармоническое колебание $x(t) = A \sin(\omega \cdot t)$	Идеальный "белый шум"	Реальный чисто случайный сигнал	Смешанный сигнал
			
$R(\tau) = \frac{A^2}{2} \cos(\omega \cdot \tau)$	$R(\tau) = \begin{cases} x^2 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}$	Спадает от $\overline{x^2}$ при $\tau = 0$ до $(\bar{x})^2$ при $\tau \neq 0$ Центрированный сигнал спадает от $\overline{x^2}$ при $\tau = 0$ до 0 при $\tau \neq 0$ .	Не прекращает осциллировать при достаточно больших $\tau$

Рис. 15. Автокоррелограммы различных сигналов

Автокорреляционная функция дискретизированной реализации  $x_i = x(t_i)$  вычисляется

по формуле:  $R(r) = \frac{1}{N-r} \sum_{i=1}^{N-r} x_i \cdot x_{i+r}$ , где  $r = 0, 1, 2, \dots, m$ ;

$r \cdot h = \tau$  – сдвиг по времени,  $m \cdot h$  – максимальный сдвиг по времени,  $h$  – шаг дискретизации.

Для получения достоверных результатов желательно, чтобы соблюдалось соотношение:

$m \approx \frac{N}{10}$ , но в любом случае  $m$  не должно превышать  $N/2$ .

Самый эффективный способ выявления периодических составляющих – это вычисление спектральной плотности. Спектральная плотность, характеризует поведение случайного процесса в частотной области или его частотную структуру. Спектральная плотность  $G(\omega)$  – всегда действительная и неотрицательная функция. Она показывает распределение энергии мощности между гармоническими составляющими разных частот и вычисляется по формуле:

$$G(\omega) = \lim_{\Delta\omega \rightarrow 0} \frac{\Psi^2(\omega, \Delta\omega)}{\Delta\omega}, \text{ где } \Psi^2(\omega, \Delta\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x^2(\omega, \omega + \Delta\omega, t) dt.$$

$\Psi^2(\omega, \Delta\omega)$  среднее значение квадратов амплитуд гармонических составляющих с частотами в диапазоне от  $\omega$  до  $\omega + \Delta\omega$ .

Спектральная плотность и АКФ связаны друг с другом взаимным косинус–преобразованием Фурье, то есть для действительных непрерывных случайных процессов их можно вычислить по формулам Хинчина – Винера:

$$G(\omega) = 2 \int_0^{\infty} R(\tau) \cos(\omega\tau) d\tau \quad R(\tau) = \frac{1}{\pi} \int_0^{\infty} G(\omega) \cos(\omega\tau) d\omega$$

Для дискретизированной реализации  $x_i = x(t_i)$  значения которой сняты в моменты времени  $t_i = i \cdot \Delta t = i \cdot h$ , где  $i = 0, 1, 2, \dots, N-1$ , а  $h = \Delta t$  – шаг дискретизации, рекомендуется рассчитывать спектральную

плотность для дискретных частот  $f_k = \frac{k \cdot f_c}{m}$ , где  $k$  называется номером или порядком гармоники  $k = 0, 1, \dots, m$ .

Максимальный порядок гармоники  $m = \frac{N}{2}$ . Частота гармоники максимального порядка равна частоте среза  $f_m = f_c$ .

Спектральная плотность вычисляется по формуле:

$$G(f_k) = G_k = 2 \cdot h \sum_{r=0}^m R(r) \cdot \cos(2\pi \cdot f_k \cdot r \cdot h)$$

При  $r=0 \cos(0) = 1$ , а при  $r=m$

$$\cos(2\pi f_k m h) = \cos\left(\frac{2\pi \cdot k \cdot f_c m h}{m}\right) = \cos(2\pi \cdot k \cdot f_c h) = \cos\left(\frac{2\pi \cdot k h}{2h}\right) = \cos(k\pi) = (-1)^k.$$

С учетом того, что  $\cos(0) = 1$ , а  $\cos(k\pi) = (-1)^k$  получим:

$$G(f_k) = G_k = 2 \cdot h \cdot \left[ R(0) + \sum_{r=1}^{m-1} R(r) \cdot \cos\left(\frac{\pi \cdot r \cdot k}{m}\right) + (-1)^k \cdot R(m) \right]$$

Величину  $G_k$  - называют  $k$  - той гармоникой (гармоникой  $k$  - того порядка).

Гипотеза периодичности проверяется по спектру сигнала.

Для гармонического колебания - чисто периодического сигнала спектральная плотность имеет вид  $\delta$ -функции Дирака. Для идеального белого шума она постоянна, это значит, что все гармонические составляющие вносят одинаковый энергетический вклад в частотную структуру сигнала. Реальные чисто случайные процессы имеют размытый спектр. Это значит, что спектральные плотности частотных составляющих близки между собой. Спектр смешанных сигналов, имеет выраженные пики на частотах периодических составляющих (Рис. 16.).

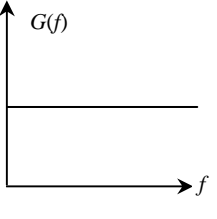
Гармоническое колебание $x(t) = A \sin(\omega \cdot t) = A \sin(2\pi \cdot f \cdot t)$	Идеальный "белый шум"	Реальный чисто случайный сигнал	Смешанный сигнал
			
$G(f) = \begin{cases} \frac{A^2}{2} & \text{при } f = f' \\ 0 & \text{при } f \neq f' \end{cases}$	Спектральная плотность постоянна $G(f) = 2 \cdot x^2$	Размытый спектр	Спектр имеет выраженные пики на частотах периодических составляющих.

Рис. 16. Спектры различных сигналов

## ЦЕЛЬ РАБОТЫ

Изучение применения методов математической статистики для цифрового анализа случайных процессов.

## ПОРЯДОК РАБОТЫ

1. Скопировать средствами WINDOWS в личную папку файл с исходными данными по заданию преподавателя.
2. Лабораторная работа выполняется средствами специализированного программного комплекса MathCAD. Запустить MathCAD с рабочего стола или через меню "Пуск Программы".

Для оформления документа в виде отчета каждый фрагмент документа нужно сопровождать текстовым комментарием. Для ввода текста нужно щелчком мыши задать место вставки и из меню Insert (Вставка) командой Text Region (Область текста) вызвать появление шаблона текста – прямоугольника

с курсором, указывающим позицию ввода.



Шаблон текста также можно получить, напечатав знак кавычки в заданном месте, или просто начав вводить текст.

Рис. 17. Шаблон текста

После нажатия пробела вводимый фрагмент станет областью текста. По мере ввода область текста увеличивается в размерах. По окончании набора нужно просто щелчком вывести курсор за пределы области текста.

Для форматирования текста используется панель форматирования Format Bar:



Рис. 18. Панель форматирования

3. Ввести текст "Отчет по лабораторной работе" и указать фамилию и номер группы.
4. Задать нумерацию элементов массивов с единицы командой  $\text{ORIGIN}:=1$
5. Считать экспериментальные данные из файла, скопированного в личную папку, и присвоить считанные значения массиву  $x$  командой  $x:=\text{READPRN}(\text{"имя файла.txt"})$ .
6. Определить длину реализации  $N$  по формуле:  $N:=\text{length}(x)$ .
7. Задать диапазон значений дискретного индекса для нумерации элементов массива по формуле:  $i:=1 \dots N$ . Диапазон значений дискретной величины задается клавишей «точка с запятой» или кнопкой на панели «Матрицы».

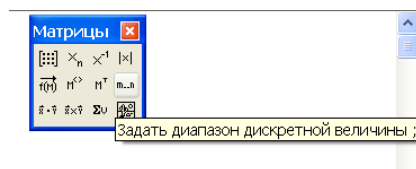


Рис. 19. Задание диапазонов значений дискретной величины

8. Построить график исходной реализации с помощью панели «Графики».

По горизонтальной оси задается индекс  $i$ , по вертикальной оси задается  $x_i$ .

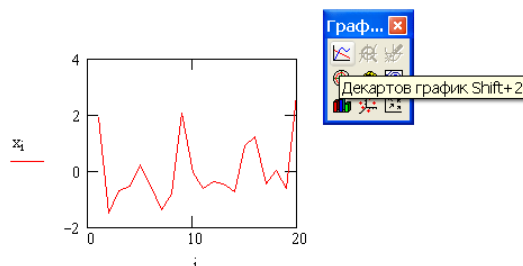


Рис. 20. График исходной реализации

### ПОСТРОИТЬ РАСПРЕДЕЛЕНИЕ ЭКСПЕРИМЕНТАЛЬНОЙ РЕАЛИЗАЦИИ.

9. Для построения распределения необходимо:

найти минимальное значение  $x_{\min} := \min(x)$

найти максимальное значение  $x_{\max} := \max(x)$

задать количество интервалов распределения  $n := 5$ .

найти величину интервала распределения  $k := \frac{x_{\max} - x_{\min}}{n}$

задать индекс интервала распределения  $j := 1 \dots n+1$

рассчитать массив границ интервалов распределения  $d_j := x_{\min} + (j-1) \cdot k$

рассчитать массив частот распределения  $h$  с помощью встроенной функции hist (гистограмма) по формуле:  $h := \text{hist}(d, x)$

10. построить гистограмму распределения, как показано на рисунке Рис. 21.

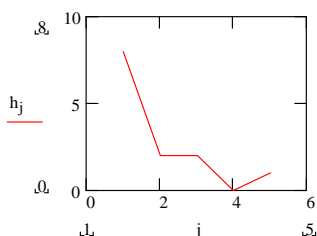


Рис. 21. Гистограмма распределения

11. Двойным щелчком левой кнопки мыши в области графика вызвать диалоговое окно форматирования Formatting Currently Selected X-Y Plot.

Formatting Currently Selected X-Y Plot

Оси X-Y   След   Метки   Умножение

Имя	Символ	Линия	Цвет	Тип	Толщина
trace 1			red	solidbar	1
trace 2	none	dot	blu	lines	1
trace 3	none	dash	gm	lines	1
trace 4	none	dashdot	mag	lines	1
trace 5	none	solid	cya	lines	1
trace 6	none	solid	bm	lines	1

trace 1   none   solid   red   solidbar   1

☐ Скрывать аргументы   ☒ С

lines  
points  
error bar  
step  
draw  
stem  
solidbar

OK   Отмена   Применить   Справка



- ## ОЦЕНИТЬ СТАЦИОНАРНОСТЬ ЭКСПЕРИМЕНТАЛЬНОЙ РЕАЛИЗАЦИИ ПО КРИТЕРИЮ СЕРИЙ












- Бүлево 
-     
-     
- Больше либо равно Ctrl+0

Рис. 23. Панель «Булево»

15. Построить график массива серий  $y$



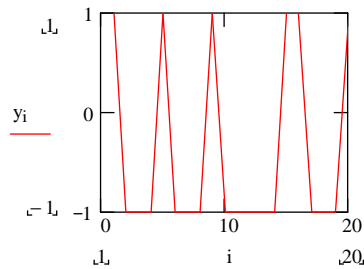


Рис. 24. График массива серий

Расширить границы от -2 до 2.

Через окно форматирования (см. п.8) задать в закладке «След» тип траектории «step» - шаг.

Задать в закладке «Оси X-Y» Вид оси – Пересечение.

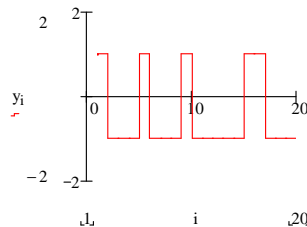


Рис. 25. Отформатированный график массива серий

16. Рассчитать количество серий.

Из последнего графика (Рис. 25) видно, что количество серий на 1 больше, чем количество смен знака в массиве  $y$ .

Количество смен знака в массиве  $y$  рассчитывается при помощи условной функции  $\text{if}()$ .

Задаем вспомогательный индекс  $j:=1 \dots N-1$  и рассчитываем количества смен знака по формуле:

$$n := \sum_j \text{if}(y_j \cdot y_{j+1} < 0, 1, 0)$$

Количество серий рассчитывается по формуле:  $r:=n+1$

17. По критерию серий проверить стационарность экспериментальной реализации на уровне значимости, заданном преподавателем, и записать вывод в виде комментария.

ОЦЕНИТЬ СОГЛАСИЕ РАСПРЕДЕЛЕНИЯ ЭКСПЕРИМЕНТАЛЬНОЙ РЕАЛИЗАЦИИ С НОРМАЛЬНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ ПО КРИТЕРИЮ ШАПИРО-УИЛКА

18. Для оценки нормальности необходимо из исходного массива  $x$  создать массив  $z$ , в котором значения будут расположены по возрастанию. Массив  $z$  создается по формуле:

$$z := \text{sort}(x)$$

19. Задать вектор коэффициентов  $a$ . Для этого необходимо написать  $a:=$  и затем на панели «Матрицы» нажать кнопку «Создать матрицу или вектор» или нажать клавиши Ctrl+M.

В диалоговом окне «Создание матрицы» задать 1 столбец и число строк, равное  $N/2$  и в местозаполнители ввести значения коэффициентов из Таблицы 1. в зависимости от длины реализации  $N$ .



Рис. 26. Создание матрицы или вектора

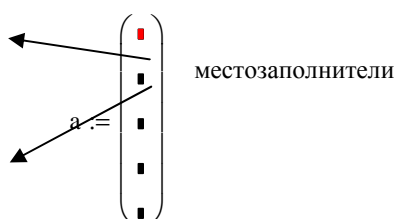


Рис.27. Задание вектора коэффициентов

20. Далее вычислить  $k := \frac{N}{2}$   $b := \sum_{i=1}^k a_i \cdot (z_{N-i+1} - z_i)$  ;  $g := \sum_{i=1}^N z_i^2 - \frac{1}{N} \cdot \left( \sum_{i=1}^N z_i \right)^2$  и сам критерий Шапиро-Уилка  $V := \frac{b^2}{g}$

Следует помнить, что значения критерия Шапиро-Уилка не должны превышать 1. Если  $V > 1$ , то вычисления ошибочны и нуждаются в исправлении.

21. По критерию Шапиро-Уилка проверить «нормальность» экспериментальной реализации на уровне значимости, заданном преподавателем, и записать вывод в виде комментария.

ОЦЕНИТЬ НАЛИЧИЕ ИЛИ ОТСУТСТВИЕ ПЕРИОДИЧЕСКИХ СОСТАВЛЯЮЩИХ В ЭКСПЕРИМЕНТАЛЬНОЙ РЕАЛИЗАЦИИ ПО АВТОКОРРЕЛОГРАММЕ И СПЕКТРУ СИГНАЛА.

22. Задать диапазон изменения сдвига по времени:  $t := 0..k$ ,

где  $k := \frac{N}{2}$  было задано ранее.

23. Вычислить Автокорреляционную функцию:

$$R(t) := \frac{1}{N-t} \cdot \sum_{i=1}^{N-t} x_i \cdot x_{i+t}$$

24. Построить график автокорреляционной функции и отформатировать его через окно форматирования (см. п.8), задав в закладке «Оси X-Y» Вид оси – Пересечение.

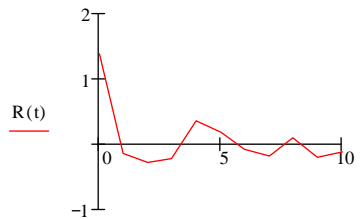


Рис. 28. График автокорреляционной функции

25. Задать диапазон изменения номера гармоники:  $j := 0..k$ ,

где  $k := \frac{N}{2}$  было задано ранее.

26. Считая шаг дискретизации  $h = 1$  сек., вычислить частоты гармоник по

формуле:  $f(j) = \frac{j}{2}$ , так как  $f(j) = j \cdot f_c = \frac{j}{2h} = \frac{j}{2}$ , при  $h = 1$ .

27. Вычислить спектральную плотность

$$G(j) := 2 \cdot \left[ R(0) + 2 \cdot \sum_{t=1}^{k-1} \left( R(t) \cdot \cos\left(\frac{\pi \cdot t \cdot j}{K}\right) \right) + (-1)^j \cdot R(k) \right]$$

28. Построить график спектральной плотности (амплитудный спектр) и отформатировать его через окно форматирования (см. п.8), задав в закладке «След» тип траектории «stem».

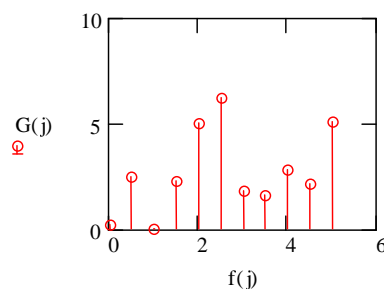


Рис. 29. График спектральной плотности. Амплитудный спектр.

29. По виду автокоррелограммы и амплитудного спектра сделать вывод о наличии или отсутствии периодических составляющих.

Если в экспериментальной реализации присутствуют периодические составляющие, нужно определить их частоты. Для этого нужно параллельно вывести частоты и амплитуды гармоник, как показано на рисунке Рис. 30., и визуально определить те частоты, которые соответствуют пикам спектральной плотности.

$f(j) =$	$G(j) =$
0	0.263
0.5	2.498
1	0.026
1.5	2.275
2	5.028
2.5	6.225
3	1.827
3.5	1.649
4	2.834
4.5	2.16
5	5.114

Рис. 30. Таблицы значений частот и соответствующих амплитуд гармоник.

30. Записать в виде комментария вывод об отсутствии или наличии периодических составляющих с указанием их частот.
31. Письменно в виде комментария ответить на контрольные вопросы по заданию преподавателя.

## СОДЕРЖАНИЕ ОТЧЕТА

Отчетом является файл документа с комментариями и ответами на контрольные вопросы.

### КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Чем отличаются детерминированные и случайны процессы?
2. Каков физический смысл автокорреляционной функции и спектральной плотности?
3. Как на практике классифицируется процесс, то есть определяется детерминированный он или случайный?
4. Является ли критерий Шапиро – Уилка параметрическим критерием согласия или непараметрическим и почему?
5. Что называется реализацией процесса?
6. Какие процессы называются общепериодическими?
7. По какому принципу исходная реализация разбивается на серии?
8. Какие процессы называются гармоническими?
9. Что подразумевается под понятием "уровень значимости"?
10. Какие процессы называются переходными?
11. Как взаимосвязаны между собой произвольные значения «белого шума»?
12. Что такое «квазидетерминированный процесс»?
13. Что означает «накопленная частота распределения»?
14. Что означает понятие «эргодический процесс»?
15. Что такое «дискретизация» сигнала?
16. Что означает «частота распределения»?
17. Для чего необходима оценка, основных свойств случайных процессов?
18. Какая частота называется частотой среза?
19. Чем отличаются стационарный и нестационарный случайные процессы?
20. Какая помеха называется «аддитивной»?
21. Каков физический смысл математического ожидания и дисперсии?
22. Какая помеха называется «мультипликативной»?
23. Что такое «сглаживание» сигнала?
24. Каков физический смысл функции и плотности распределения?

В методических указаниях даны понятия о физических процессах, их классификации, об основных свойствах случайных процессов и статистических методах оценки этих свойств, о параметрических и непараметрических критериях оценки и их применении. Приведены алгоритмы и примеры использования интегрированной системы технических вычислений MathCad, представляющей собой мировой стандарт для научно-технических расчетов.

## ЦИФРОВОЙ АНАЛИЗ СЛУЧАЙНЫХ ПРОЦЕССОВ

Методические указания к лабораторным работам

Составитель                      Голованова И.С.

Подписано к печати

Формат 60х84/16. Бумага писчая № 2.

Плоская печать. Усл. печ. л. 2,3. Уч. – изд. л. 2,30.

Тираж 50 экз. Заказ №                      . Бесплатно.

ИПФ ТПУ. Лицензия ЛТ №1 от 18.07.94.

Ротапринт ТПУ, 634034, г. Томск, пр. Ленина, 30