МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

им. М.В.ЛОМОНОСОВА

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра физики атмосферы

Методические указания задачи спецпрактикума

"Выделение и исследование свойств шумов (в задачах геофизики)"

составил доц. В.И. Захаров

Москва, МГУ - 2022

Ниже приведены только основные понятия и методические указания, необходимые для выполнения задачи спецпрактикума для магистров кафедра физики атмосферы физического факультета МГУ

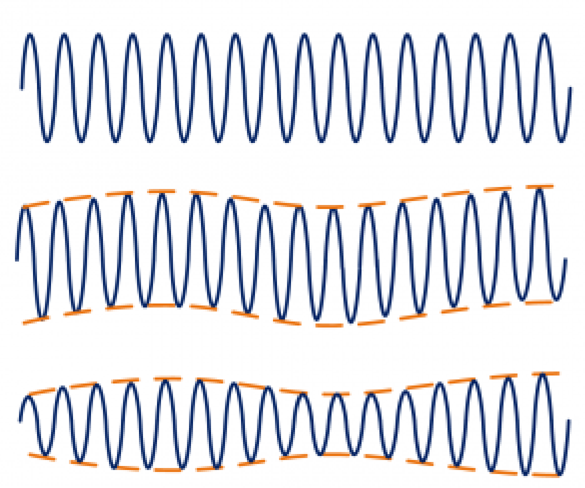
1. Шумы в измерениях

Модель измерений - смесь сигнала (полезного, информационного, периодического..) и шума (или мешающего сигнала, помехи) с определенными характеристиками.

Задачи 1) фильтрация, т.е. удаление, шума для приема информационной составляющей

2) выделение шума для определения его характеристик

Общие замечания.



Если в тракте передачи или преобразования информации функцию сигнала во времени t можно представить в виде

$$U(t) = U_S(t) + U_N(t),$$

где $U_S(t)$ - полезный сигнал, а $U_N(t)$ - помеха, то говорят о *аддитивной помехе*.

Если эту функцию сигнала можно представить в виде

$$U_S(t) * U_N(t) \quad \text{или} \quad U_S(t) / U_N(t),$$

то говорят о *мультипликативной* помехе.

На рисунке показана функция полезного сигнала во времени (сверху), в центре показан этот сигнал после воздействия аддитивной низкочастотной помехи, снизу - после воздействия этой же помехи, действующей мультипликативно.

В отличие от аддитивной, мультипликативная помеха является нелинейным эффектом и вызывает нелинейные искажения сигнала.

Аддитивная или мультипликативная помехи по-разному действуют в цепи измерения (по-разному влияют на абсолютную и относительную погрешности измерений).

Аддитивное или мультипликативное действие помехи - это обусловлено конкретными физическими явлениями, функциональной схемой данного канала передачи или преобразования сигнала, а также точкой приложения помехи в этой функциональной схеме.

2. Характеристики шумов

Как правило, шум естественного происхождения является случайным процессом, который характеризуется статистическими характеристиками, в общем случае, зависящими от времени. Для их описания широко используют такое понятие, как случайный (стохастический) процесс.

Пусть $x(t)$ - зависимость от времени некоторой величины, (например, температуры, давления, тока, напряжения в цепи и проч.) - так называемая реализация наблюдаемого процесса- смеси шума и регулярной составляющей. Смесь сигналов может принимать различные значения с заданным распределением вероятностей. Случайная $x(t)$ в общем случае может принимать действительные значения от $-\infty$ до $+\infty$. Из этого процесса может быть выделен шум.

Параметры шума как процесса

1) интегральная и дифференциальная функция распределения

2) моменты процесса (центральные или нецентральные).

Для описания случайного процесса часто вместо самих функций распределения случайного процесса рассматривают лишь отклонения характеризующих его физических величин от их средних значений, которые называют флюктуациями, а иногда пульсациями случайного процесса (например, флуктуации тока, напряжения или сопротивления).

Для гауссова случайного процесса описание с помощью моментов дает полную характеристику процесса.

Основные математические характеристики шума.

Наиболее важными для практических приложений вероятностными характеристиками случайного процесса $x(t)$ являются одновременная $w_1(x,t)$ и двухвременная $w_2(x_1,t_1;x_2,t_2)$ плотности вероятности. При этом $w_1(x,t)dx$ есть вероятность того, что в момент времени t случайный процесс принимает значение, лежащее в интервале dx вокруг значения случайной величины x, а $w_2(x_1,t_1;x_2,t_2)dx_1dx_2$ есть вероятность того, что в некоторый момент времени t_1 случайная величина принимает значение, лежащее в интервале dx_1 вокруг значения x_1, и в момент времени t_2 принимает значение, лежащее в интервале dx_2 вокруг значения x_2.

Шумы подразделяются на статистически стационарные и нестационарные. В случае, когда параметры рассматриваемой системы постоянны во времени и вероятностные характеристики шума, генерируемого в системе, не зависят от начала отсчета времени t, такие процессы называются стационарными. В этом случае

$$\int_{-\infty}^{+\infty} w_1(x,t)dx = 1$$

Для стационарных процессов одновременная плотность вероятности w_1 не зависит от времени, а двухвременная плотность вероятности w_2 зависит лишь от разности моментов времени $\tau = t_2 - t_1$.

Для нестационарных случайных процессов их вероятностные характеристики являются функциями времени. Многие из встречающихся в Природе случайных процессов являются статистически стационарными случайными процессами, которые в основном и рассматриваются в настоящей книге.

Вероятностные характеристики дают наиболее полное описание случайного процесса. В теории случайных процессов для непрерывных случайных величин рассматривают начальные моменты распределения k-го порядка $m_k(x)$ (k = 1,2,…), которые вычисляются по формуле:

$$m_k(x) = \int_{-\infty}^{\infty} x^k w_1(x)dx$$

Заметим, что величина $x(t)$ является случайной функцией времени t. При этом среднее значение (или математическое ожидание) $m_1(x)$ случайной величины $x(t)$ определяется при значении k = 1:

$$m_1(x) = \overline{x(t)} = < x(t) > = \int_{-\infty}^{\infty} x w_1(x)dx .$$

Здесь черта сверху или скобки < > означают операцию усреднения по времени или по множеству реализаций случайного процесса (по ансамблю). Для статистически стационарных процессов справедлива эргодическая гипотеза, согласно которой усреднение по ансамблю и по большому интервалу времени эквивалентны, т.е. среднее по ансамблю равно среднему по времени. При этом нет необходимости изучать большую совокупность реализаций, а достаточно лишь одной реализации, например, при изучении шумов в идентичных по типу резисторах или в различных п/п приборах можно ограничиться одним образцом.

Начальный момент второго порядка (k = 2) является средним квадратом случайной величины x и определяется выражением:

$$\overline{x^2(t)} = < x^2(t) > = \int_{-\infty}^{\infty} x^2(t) w_1(x)dx .$$

Наряду с начальными моментами распределения рассматриваются также центральные моменты случайных величин различных порядков, которые вычисляются по формуле:

$$M_k(x) = \int_{-\infty}^{\infty} (x - \bar{x})^k w_1(x)dx .$$

Важнейшим центральным моментом является так называемая дисперсия - центральный момент второго порядка, который характеризует размах флуктуаций около среднего значения и интенсивность или мощность шума.

Дисперсия определяется формулой:

$$M_2(x) = \sigma^2 = \overline{[x(t) - \bar{x}]^2} = \int_{-\infty}^{\infty} (x - \bar{x})^2 w_1(x)dx .$$

При решении задач, связанных с электрическими флуктуациями, в большинстве случаев достаточно знания среднего значения случайного процесса и его дисперсии. Рассматриваемая в данном разделе корреляционная теория изучает свойства случайных процессов, которые определяются моментами первого и второго порядка.

Для стационарного случайного процесса среднее значение величины $x(t)$, дисперсия и другие статистические характеристики не зависят от времени. Классической моделью стационарного шума является белый шум (white noise), спектральные составляющие которого равномерно распределены по всему диапазону рассматриваемых частот. Его энергетический спектр (распределение интенсивности шума по частотам) представляет собой прямую, параллельную оси частот. Это эквивалентно тому, что мощность шума для постоянной полосы частот Δf, сосредоточенной на произвольно выбранной частоте f, не изменяется с частотой. "*Белый шум*" получил такое название, потому что он имеет одинаковое энергетическое распределение по всему диапазону частот подобно белому свету.

Если рассматривать флуктуирующие (шумовые) напряжения (или токи), то тогда средний квадрат напряжения $\overline{u^2(t)}$ соответствует средней мощности шума, выделяющейся на сопротивлении 1 Ом.

На практике обычно наблюдаемые случайные процессы в радиофизике или в других природных системах являются нормальными или гауссовыми, т.е. процессами с плотностью вероятности $w_1(x)$, описываемой гауссовым (нормальным) законом распределением. Плотность вероятности гауссова (нормального) распределения имеет вид:

$$w_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\bar{x})^2}{2\sigma^2}) \ ,$$

где (x - $\bar{x}$) - отклонение от среднего значения $\bar{x}$ флуктуирующей величины, $\sigma^2 = \overline{[x(t) - \bar{x}]^2}$ - дисперсия.

На рис. представлена функция плотности вероятности гауссова (нормального) распределения случайной переменной x

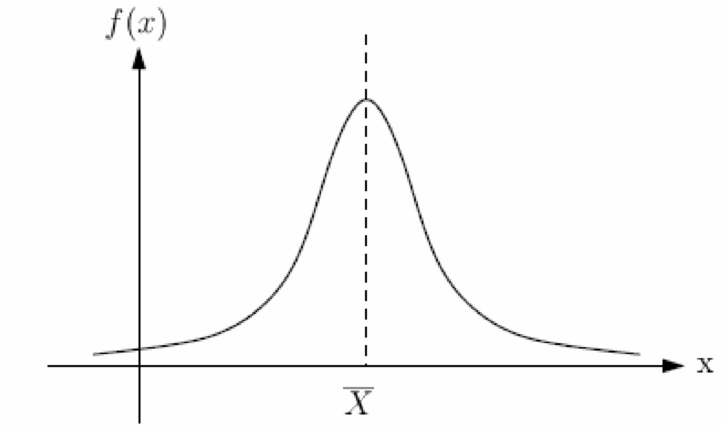


Рис. Функция плотности вероятности гауссова (нормального) распределения случайной переменной x.

Это связано с тем, что согласно центральной предельной теореме любые физические явления, возникающие в результате воздействия суммы большого числа независимых случайных событий, подчиняются нормальному распределению плотности вероятности случайной величины, т.е. являются гауссовским случайным процессом. Центральная теорема имеет много практических приложений. Суперпозиция множества случайных возмущений лежит в основе механизмов возникновения шума в радиофизике и в подавляющем числе других природных систем. Поэтому наблюдаемые шумы, как правило, являются гауссовскими. Нормальный стационарный случайный процесс охватывает широкий класс явлений в радиотехнических устройствах и в физике вообще. Для нормального распределения все моменты случайной величины определяются первыми двумя моментами (средним и дисперсией).

Вместе с тем в наноразмерных проводниках и в ИС субмикронных размеров, имеющих малое число носителей заряда, а также в других современных электронных приборах может наблюдаться отклонение от нормального закона распределения. На практике центральная предельная теорема обычно выполняется лишь для относительно линейных физических систем и протекающих в них процессов, которые обусловливают возникновение изучаемого явления. Наличие нелинейных процессов в системе часто приводит к значительным отклонениям от нормального закона распределения.

Поэтому в общем случае предположение о нормальном законе распределения случайной величины подлежит проверке.

Спектральная плотность мощности стационарного случайного процесса

Важной характеристикой стационарных случайных процессов является спектральная плотность мощности, описывающая распределение мощности шума по частотному спектру. Рассмотрим стационарный случайный процесс, который может быть представлен беспорядочной последовательностью импульсов напряжения или тока, следующих друг за другом через случайные интервалы времени. Процесс со случайной последовательностью импульсов является непериодическим. Тем не менее, можно говорить о спектре такого процесса, понимая в данном случае под спектром распределение мощности по частотам.

Для описания шумов вводят понятие спектральной плотности мощности (СПМ) шума, называемой также в общем случае спектральной плотностью (СП) шума, которая определяется соотношением:

$$S(f) = \lim_{\Delta f \to 0} \frac{\Delta P(f)}{\Delta f},$$

где $\Delta P(f)$ - усредненная по времени мощность шума в полосе частот Δf на частоте измерения f.

Как следует из соотношения (2.10), СП шума имеет размерность Вт/Гц. В общем случае СП является функцией частоты. Зависимость СП шума от частоты называют энергетическим спектром, который несет информацию о динамических характеристиках системы.

Если случайный процесс эргодический, то можно находить энергетический спектр такого процесса по его единственной реализации, что широко используется на практике..

При рассмотрении спектральных характеристик стационарного случайного процесса часто оказывается необходимым пользоваться понятием ширины спектра шума. Площадь под кривой энергетического спектра случайного процесса, отнесенную к СП шума на некоторой характерной частоте f_0, называют эффективной шириной спектра, которая определяется по формуле:

$$\Delta f_{\text{эф}} = \frac{1}{S(f_0)} \int_0^{\infty} S(f)\,df,$$

Эту величину можно трактовать как ширину равномерного энергетического спектра случайного процесса в полосе $\Delta f_{\text{эф}}$, эквивалентного по средней мощности рассматриваемому процессу.

Мощность шума P, заключенная в полосе частот $f_1 \ldots f_2$, равна

$$P = \int_{f_1}^{f_2} S(f)\,df.$$

Если СП шума в полосе частот $f_1 \ldots f_2$ постоянна и равна S_0, тогда для мощности шума в данной полосе частот имеем: $P = S_0 \Delta f$, где $\Delta f = f_2 - f_1$ - полоса частот, пропускаемая схемой или измерительным прибором.

Важным случаем стационарного случайного процесса является белый шум, для которого спектральная плотность не зависит от частоты в широком диапазоне частот (теоретически - в бесконечном диапазоне частот). Энергетический спектр белого шума в диапазоне частот $-\infty < f < +\infty$ дается выражением:

$$S(f) = 2S_0 = \text{const}$$

Модель белого шума описывает случайный процесс без памяти (без последействия). Белый шум возникает в системах с большим числом простых однородных элементов и характеризуется распределением амплитуды флуктуаций по нормальному закону. Свойства белого шума определяются статистикой независимых одиночных событий (например, тепловым движением носителей заряда в проводнике или полупроводнике). Вместе с тем истинный белый шум с бесконечной полосой частот не существует, поскольку он имеет бесконечную мощность.

Итак, на практике при оценке величины шума какого-либо процесса измеряют среднее его значение и среднеквадратичную мощность (дисперсию) отклонений от среднего. При дискретных наблюдениях такие оценки проводят по *выборкам или реализациям случайного процесса, которые считаются независимыми.*

(Вопрос: Что изменится в оценках, если выборки зависимы?)

Спектральные оценки мощности шума проводят на основе теорема Винера-Хинчина.

Энергетический спектр $S(f)$ и автокорреляционная функция $K(\tau)$ стационарного случайного процесса $x(t)$ связаны друг с другом парой преобразования Фурье (теорема Винера - Хинчина). При этом энергетический спектр $S_x(f)$ стационарного случайного процесса $x(t)$ определяется как преобразование Фурье от автокорреляционной функции:

$$S_x(\omega) = 2 \int_{-\infty}^{\infty} K(\tau)\, e^{-i\omega\tau} d\tau = 4 \int_0^{\infty} K(\tau) \cos\omega\tau\, d\tau \,,$$

где $\omega = 2\pi f$ - угловая частота, i - мнимая единица.

Автокорреляционная функция в свою очередь есть обратное преобразование Фурье от спектральной плотности мощности шума $S_x(f)$.

$$K(\tau) = \frac{1}{4\pi} \int_{-\infty}^{\infty} S_x(\omega)\, e^{i\omega\tau}\, d\omega$$

Поскольку автокорреляционная функция стационарного случайного процесса $K(\tau)$ является четной функцией временного сдвига τ, последнее выражение можно преобразовать к виду, удобному для расчетов:

$$K(\tau) = \frac{1}{2\pi} \int_0^{\infty} S_x(\omega) \cos\omega\tau\, d\omega = \int_0^{\infty} S_x(f) \cos 2\pi f\tau\, df$$

Отсюда при $\tau = 0$ (и при $\overline{x(t)} = 0$) получим выражение для дисперсии случайной величины $x(t)$:

$$\sigma_x^2 = \overline{x^2(t)} = K(0) = \int_0^{\infty} S_x(f)\, df \,,$$

которая есть не что иное, как среднеквадратичное значение случайной величины $x(t)$, а для случая электрических сигналов - мощность шума, выделяемая на сопротивлении 1 Ом.

Автокорреляционная функция $K(\tau)$ и энергетический спектр $S(f)$ стационарного случайного процесса, как пара преобразования Фурье, обладают всеми присущими этому преобразованию свойствами. В частности, чем шире эффективная ширина спектра, тем быстрее изменяется значение флуктуирующей переменной, и тем быстрее процесс забывает свое начальное состояние, т.е. с увеличением ширины полосы частот энергетического спектра $\Delta f_{эф}$ время корреляции τ_K уменьшается. Т.е., чем меньше ширина функции корреляции случайного процесса на оси временного сдвига τ, тем шире его спектр. Для всех процессов с одинаковой формой энергетического спектра и, следовательно, с корреляционной функцией одного вида произведение $\Delta f \tau_K$ является некоторой константой.

Широкополосные и узкополосные случайные процессы.

Стационарный случайный процесс с непрерывным энергетическим спектром называют узкополосным, когда спектр его сосредоточен в относительно узкой полосе частот Δf около некоторой фиксированной частоты $\omega_0 = 2\pi f_0$ (рис. 2а), так что условие узкополосности случайного процесса имеет вид $\Delta f \ll f_0$, или широкополосным, когда указанное условие не выполняется (рис. б).

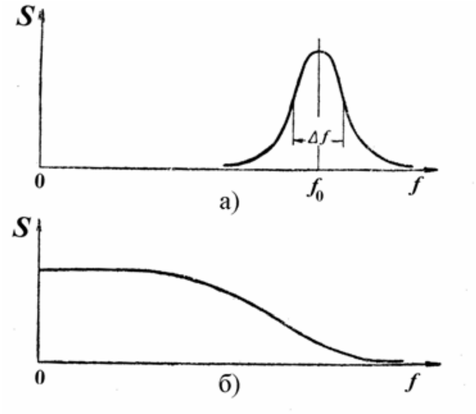


Рис. Энергетические спектры узкополосного (а) и широкополосного (б) случайных процессов.

Если узкополосный шум подать на осциллограф и развертку синхронизировать на частотах, кратных центральной частоте полосы f_0, то реализация узкополосного случайного процесса на экране осциллографа будет похожа на синусоиду фиксированной частоты f_0. Близость к синусоиде имеет место на протяжении довольно большого числа периодов $T = 1/f_0$.

Для энергетического спектра широкополосного случайного процесса $S(\omega) = 2S_0$ = const в широком диапазоне частот. Хорошим примером широкополосного случайного процесса является белый шум, который называют иногда дельта-коррелированным случайным процессом, в связи с тем, что функция корреляции для белого шума является дельта-функцией, $\delta(\tau)$:

$$K(\tau) = \frac{1}{4\pi} \int_{-\infty}^{\infty} S_x(\omega)\, e^{i\omega\tau}\, d\omega = \frac{S_0}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau}\, d\omega = S_0\,\delta(\tau),$$

где $\delta(\tau)$ - функция равная нулю при всех значениях $\tau \neq 0$ и $\delta(0) = \infty$.

Замена реального случайного процесса на δ-коррелированный в различных физических системах означает, что граничная частота спектра шумов не вводится в рассмотрение, поскольку она гораздо больше всех других частот, существенных для рассматриваемой системы.

Вместе с тем для практики наиболее важное значение имеет ограниченный по полосе частот белый шум, энергетический спектр которого аппроксимируется прямоугольником в полосе частот $|f| < \Delta f_0$, как показано на рис.

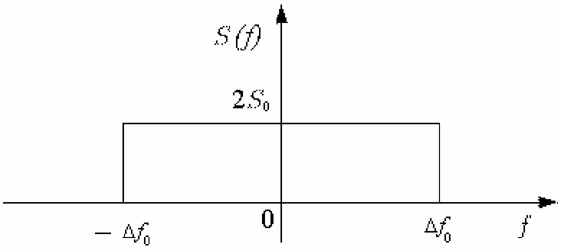
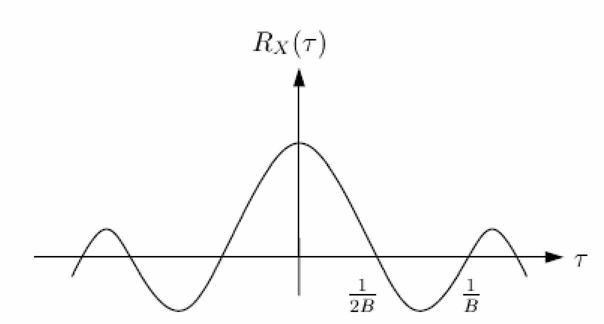


Рис. СП белого шума в ограниченной полосе частот $|f| < \Delta f_0$.

Для ограниченного по полосе частот белого шума автокорреляционная функция дается отношением:

$$K(\tau) = 2S_0 \Delta f_0 \frac{\sin 2\pi \Delta f_0 \tau}{\pi \Delta f_0 \tau},$$



B= Δf_0

Рис. Автокорреляционная функция ограниченного по полосе частот белого шума.

Задача выделения шума из измерений.

Эффективное выделение шума возможно, если пространственно-временные его характеристики отличаются от свойств полезного (информационного) сигнала. Простой пример - спектры сигналов разделены или один из сигналов узкополосный.

Суть фильтрации - сглаживание наблюдаемого сигнала по нескольким отсчетам

Скользящее среднее (англ. *moving average*, *MA*) - общее название для семейства функций, значения которых в каждой точке определения равны среднему значению исходной функции за предыдущий период. Скользящие средние обычно используются с данными временных рядов для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов. Математически скользящее среднее является одним из видов свёртки.

Метод базируется на предположении о независимости шума измерений, когда при определении средних значений случайные (шумовые) отклонения погашаются. При сглаживании этим методом фактические значения ряда динамики заменяются средними значениями, которые характеризуют срединную точку периода.

Простое сглаживание основывается на составлении нового ряда из простых средних арифметических, исчисленных для ряда x(t), длиной n отсчетов и промежутков времени длиной q:

$$\overline{\mathrm{x}}(\mathrm{k}) = \frac{1}{q}\sum_{\mathrm{i=k}}^{\mathrm{q+k}} \mathrm{x(t)}, \quad k = 1,2,....,n-q+1$$

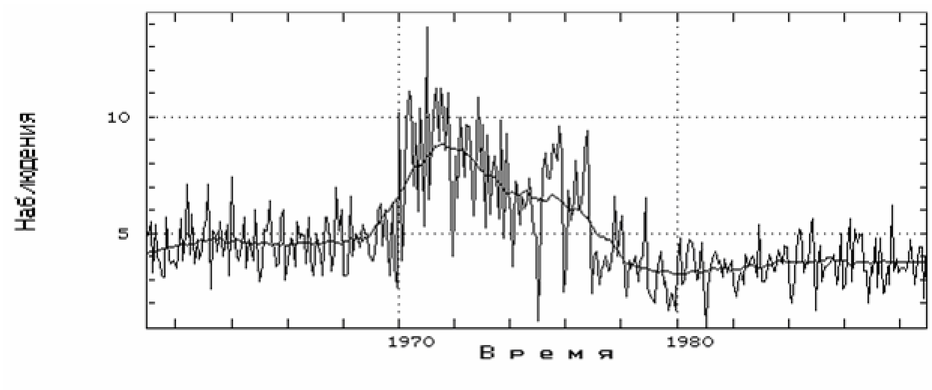


Рис. Построение оценки $\overline{\mathrm{x}}(\mathrm{k})$ ряда методом простого скользящего среднего

Первое значения отсчета сглаженного ряда минимально можно приписывать точке q/2, а последнее соответственно максимально точке n-q/2+1.

После вычитания гладкого тренда $\overline{\mathrm{x}}(\mathrm{k})$ из исходного ряда x(t) получаем новый ряд остатков y(t) , показанный на рис.

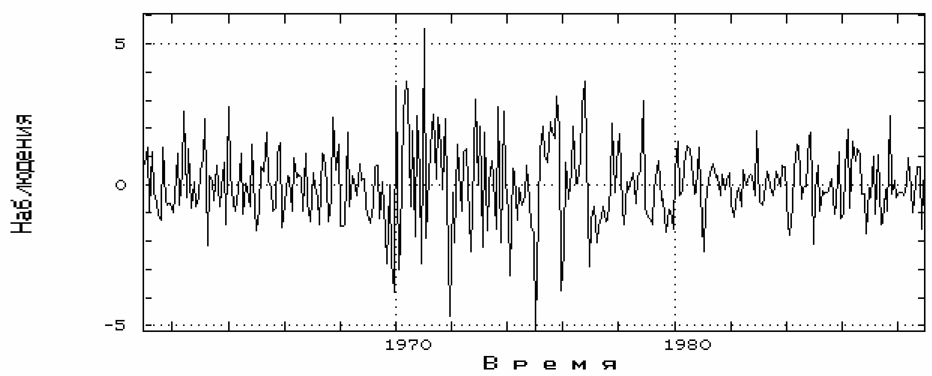


Рис. Выделение шума y(t) методом простого скользящего среднего

Ряд остатков имеет признаки случайного процесса, в приведенном примере использованная процедура сглаживания компенсирует до 40 % вариабельности ряда).

В качестве иных процедур сглаживания можно использовать взвешенное скользящее среднее по трем точкам с весами 0.25, 0.5 и 0.25, соответствующее формуле:

$$\overline{\mathrm{x}}(\mathrm{k}) = 0.25\ x(t-1) + 0.5\ x(t) + 0.25\ x(t+1)$$

(концевые точки копируются : = $x(0)$, = $x(n)$). Это так называемый сглаживающий фильтр Хэмминга.

Возможны и другие оценки трендов во времени, например, построение полинома некоторой степени, меньшей длины окна сглаживания.

Существуют и весьма развитые методы спектральной фильтрации сигналов для выделения шума.

Метод скользящих средних имеет ряд преимуществ перед другими методами:

- скользящая средняя дает функцию тренда, в наибольшей мере приближенную к значениям исследуемого ряда, поскольку для отдельных частей ряда выбирается наилучшая тенденция;
- к исследуемому ряду могут быть прибавлены новые значения;
- нахождение тренда не связано с большими вычислительными трудностями.

Недостатком метода скользящей средней является то обстоятельство, что при увеличении периода скольжения теряется информация о крайних периодах ряда, что недопустимо при некоторых приемах анализа временных рядов (например, при спектральном анализе). Кроме того, этот метод (и другие, подобные ему) может вызывать автокорреляцию остатков, даже если она отсутствовала в исходном ряду.

Кроме того, результат фильтрации и статистические характеристики выделенного шума зависят от длины окна фильтра q. Оптимальный выбор величины окна требует численных экспериментов.

4. Исследование характеристик шумов

Под характеристиками шума понимаются

1) *дифференциальная функция распределения*

Под такой функцией для эксперимента обычно понимают гистограмму.

Построение гистограммы связано с группировкой данных и потому требует тестирования получаемых результатов.

Общий подход заключается в следующем.

1. Определяется максимальное y_{max} и минимальное y_{min} значение в выборке.
2. Выбирают шаг гистограммы Δ (или количество ячеек N_G). Они связаны соотношением

$$\Delta=(y_{max}-y_{min})/N_G.$$

3. Строят распределение отсчетов ряда y, попадающих в последовательные ячейки i

$$y_{min}+\Delta*i < y \le y_{min}+\Delta*(i+1)$$

4. В результате для каждой ячейки получают число точек ряда y, попавших в эту ячейку, т.е. функцию $G(i)$, , $0 \le i \le N_G$

Замечание. При построении гистограммы важно выбрать шаг. Он должен быть достаточным для получения особенностей функции распределения, и, одновременно, позволять получить в каждой ячейке основного тела гистограммы не менее 5 отсчетов. Обычно под основным телом гистограммы понимают область вокруг максимума, содержащее не менее 70% всех отсчетов наблюдаемого процесса.

2) *моменты до 4 включительно*

Числовые характеристики эмпирического распределения называются выборочными характеристиками. Рассмотрим некоторые из них (n - объем исследуемой серии):

- *выборочное среднее*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \;;$$

- *выборочная дисперсия* (несмещённая) и *среднее квадратическое отклонение*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \;; \qquad S = \sqrt{S^2}\;;$$

- *выборочный коэффициент асимметрии*

$$Sk = \frac{\mu_3}{S^3}; \qquad \mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^3 \;;$$

- *выборочный коэффициент эксцесса*

$$Ex = \frac{\mu_4}{S^4} - 3; \qquad \mu_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^4 \;;$$

3) *спектр наблюдаемой выборки*

4) *функции авто и взаимной корреляции.*

Построение всех оценок связано с конкретными величинами окна фильтра q.

5. Проверка статистических гипотез о свойствах шума.

Теперь ясно, что таких гипотез может быть две

1. Гипотеза о том, что исследуемая выборка имеет определенный закон распределения. Проверка таких гипотез строится по довольно простой схеме. Сначала определяют параметры распределений, например, среднее значение и выборочная дисперсия. Затем строят теоретическое распределение и проверяют его соответствие наблюдаемым экспериментальным.

2. Гипотеза о том, что экспериментальные выборки являются однородными и принадлежат к одной генеральной совокупности.

Подтверждение обоих гипотез имеет важный смысл. Первая позволяет решить вопрос о функции распределения изучаемого процесса. Вторая, не решая вопрос о собственно функции распределения, позволяет дать заключение о не менее важном вопросе о степени однородности данных.

Примеры критериев

Критерий согласия χ^2-Пирсона позволяет осуществлять проверку эмпирического и теоретического (либо другого эмпирического) распределений одного признака. Данный критерий применяется, в основном, в двух случаях:

- Для сопоставления эмпирического распределения признака с теоретическим распределением (нормальным, показательным, равномерным либо каким-то иным законом);

- Для сопоставления двух эмпирических распределений одного и того же признака.

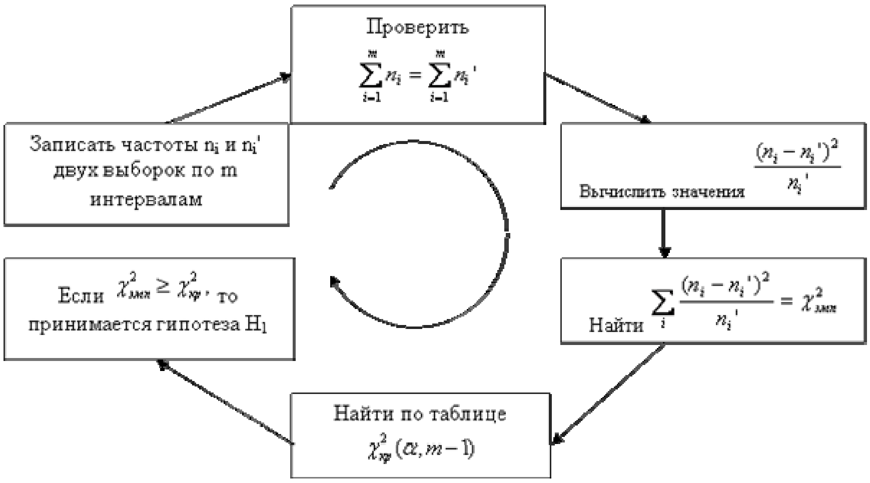
Идея метода - определение степени расхождения соответствующих частот n_i и n'_i; чем больше это расхождение, тем больше значение

$$\chi^2 = \sum_{\mathrm{i}} \frac{(n_i - n_i^{'})^2}{n_i^{'}}$$

Объемы выборок должны быть не меньше 50 и необходимо равенство сумм частот n_i и n'_i.

Проверяются **нулевая гипотеза** H_0={два распределения практически не различаются между собой}; или **альтернативная гипотеза** - H_1={расхождение между распределениями существенно}.

Общий план проверки гипотезы схематически можно представить след. диаграммой при заданном уровне значимости α и объема выборки m



Критерий Колмогорова.

На практике, кроме критерия χ^2, часто используется критерий Колмогорова, в котором в качестве меры расхождения между теоретическим и эмпирическим распределениями рассматривают максимальное значение абсолютной величины разности между эмпирической функцией распределения $F_n(x)$ и соответствующей теоретической функцией распределения

$$D = \max_{-\infty \le x \le \infty} \left| F_n(x) - F(x) \right|,$$

называемой *статистикой критерия Колмогорова.*

Доказано, что какова бы ни была функция распределения $F(x)$ непрерывной случайной величины X, при неограниченном увеличении числа наблюдений $(n \to \infty)$ вероятность неравенства $P(D\sqrt{n} \ge \lambda)$ стремится к пределу

$$P(\lambda) = 1 - \sum_{k=-\infty}^{+\infty} (1)^k e^{-2k^2\lambda^2} .$$

Задавая уровень значимости α, из табулированного или предельного соотношения при объемах выборок более 100

$$P(\lambda) = \alpha$$

можно найти соответствующее критическое значение λ_α.

Схема применения критерия Колмогорова.

1. Строятся эмпирическая функция распределения $F_n(x)$ и предполагаемая теоретическая функция распределения $F(x)$.

2. Определяется мера расхождения между теоретическим и эмпирическим распределением d по формуле (1) и вычисляется величина

$$\lambda = D\sqrt{n} \ .$$

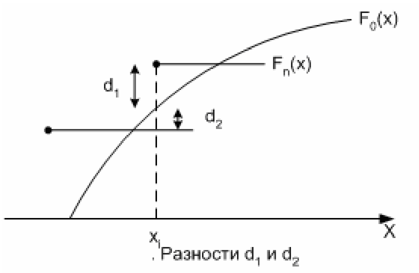


Рис. Определение отклонений функций

3. Если вычисленное значение λ окажется больше критического λ_α, определенного на уровне значимости α, то нулевая гипотеза H_0 о том, что случайная величина X имеет заданный закон распределения, отвергается (односторонний критерий). Если $\lambda \le \lambda_\alpha$, то считают, что гипотеза H_0 не противоречит опытным данным.

Замечание *Отметим, что решение подобных задач можно найти с помощью критерия χ^2. Потенциальное преимущества критерия Колмогорова в том, что он не требует группирования данных, но дает возможность рассматривать индивидуальные наблюдаемые значения. Этот критерий можно успешно применять для малых выборок. Считается, что его мощность, вообще говоря, выше, чем у критерия χ^2.*

Проверка гипотез об однородности выборок

Гипотезы об однородности выборок - это гипотезы о том, что рассматриваемые выборки извлечены из одной и той же *генеральной совокупности*.

Пусть имеются две независимые выборки, произведенные из генеральных совокупностей с неизвестными теоретическими функциями распределения $F_1(x)$ и $F_2(x)$.

Проверяемая нулевая гипотеза имеет вид H_0: $F_1(x) = F_2(x)$ против конкурирующей $H_1: F_1(x) \ne F_2(x)$. Будем предполагать, что функции $F_1(x)$ и $F_2(x)$

непрерывны.

Критерий Колмогорова-Смирнова использует ту же самую идею, что и критерий Колмогорова, но только в критерии Колмогорова сравнивается эмпирическая функция распределения с теоретической, а в критерии Колмогорова-Смирнова сравниваются две эмпирические функции распределения.

Статистика критерия Колмогорова-Смирнова имеет вид:

$$\lambda' = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \max\left|F_{n_1}(x) - F_{n_2}(x)\right|,$$

где $F_{n_1}(x)$ и F_{n_2} - эмпирические функции распределения, построенные по двум выборкам с объемами n_1 и n_2.

Гипотеза H_0 отвергается, если фактически наблюдаемое значение статистики λ' больше критического λ'_{kp}, т.е. $\lambda' > \lambda'_{\text{kp}}$, и принимается в противном случае.

При малых объемах выборок $(n_1, n_2 \le 20)$ критические значения λ'_{kp} для заданных уровней значимости критерия можно найти в специальных таблицах. При $n_1, n_2 \to \infty$ (а практически при $n_1, n_2 \ge 50$) распределение статистики λ' сводится к распределению Колмогорова для статистики λ. Поэтому гипотеза H_0 отвергается на уровне значимости α, если фактически наблюдаемое значение λ' больше критического λ_α, т.е. $\lambda' > \lambda_\alpha$, и принимается в противном случае.

Рекомендованный порядок выполнения работы.

Все данные для анализа рекомендуется брать из реальных наблюдений о различных геофизических полях (например, температуры, давления, влажности, сигналов GPS (псевдодальности и фазы), солнечных пятен, потока солнечного радиоизлучения и т.п., т.е. использовать свои реальные данные для подготовки квалификационных работ.

Допускается использование любого программного обеспечения. Это могут быть модули, написанные вами на любом языке программирования, специализированные математические и статистические пакеты обработки.

Каждое упражнение необходимо иллюстрировать доступными вам средствами

Задания.

1. Для выполнения работы в качестве данных возьмите 3-5 выборки из изучаемого вами в научной работе процесса. Объемы выборок должны быть достаточными для получения статистически значимых оценок с ошибкой не более 1%. Оцените необходимые вам длины рядов данных.

2. Выделите шумовую составляющую при различных значениях величины окна фильтра. В качестве метода фильтрации используйте два любых, первый обязательно - метод простого скользящего среднего.

2a) Исследуйте устойчивость получаемых вами оценок моментов и функций распределения. Для из этого каждого ряда выделите шум при различных величинах окна фильтра q. Исследуйте изменения функций дифференциального распределения в зависимости от параметра q. Занесите полученные результаты в таблицу отчета по работе.

Определите величину q, при которой изменения в оценки выборочных моментов менее статистических ошибок эксперимента.

Выборка	Параметр q1		Параметр q2		Параметр q3		Параметр q4	
	$\overline{x}$	σ	$\overline{x}$	σ	$\overline{x}$	σ	$\overline{x}$	σ
N1								

N2								
N3								

Исследуйте устойчивость функций распределения для данного объема выборки. Посмотрите, как меняются оценки среднего и дисперсии при разных q, посмотрите, соответствуют ли эти изменения величинам выборочных ошибок.

2б) Повторите п. 2а) для другого выбранного вами метода фильтрации шума.

3. Постройте оценки спектров выборок N1, N2, N3... при одной определенной в упражнении 2 величине окна q. Шаг спектра возьмите одинаковым. Сравните спектры шумов. Для сравнения 1) определите частоту максимума спектра (или спектральный диапазон частот), в котором частота спадает до уровня 0.8 по амплитуде 2) определите амплитуду спектральной оценки в максимуме и на границах выбранного уровня.

4. Постройте оценки функции взаимной корреляции для всех парных выборок. Результаты представьте в таблице.

5. Проведите проверку статистических гипотез о том, что

1) Выделенные вами в п.2 шумы имеют нормальное распределение с определенными выборочными средним и дисперсией; проверку проведите для всех выборок.

2) Исследуемые вами выборки папарно принадлежат к одной генеральной совокупности.

Постарайтесь объяснить полученные результаты. Учтите, что используя данные за целый день, вы должны быть уверены, что свойства сигналов остаются неизменными.

Все задания иллюстрируйте средствами вашего ПО.

Литература.

1. Баскаков С.И. Радиотехнические цепи и сигналы. Учебник для вузов. - М., Высшая школа, 1988.
2. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. - М.: Мир, 1989. - 540 с.
3. Васильев Д.В. Радиотехнические цепи и сигналы: Учебное пособие для вузов. - М.: Радио и связь, 1982. - 528 с.
4. Гольденберг Л.М. и др. Цифровая обработка сигналов: Учебное пособие для вузов. - М.: Радио и связь, 1990.- 256 с.
5. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. - М.: Мир, 1988. - 488 с.
6. Зиновьев А.Л., Филиппов Л.И. Введение в теорию сигналов и цепей: Учебное пособие для вузов. - М.: Высшая школа, 1975. - 264 с.
7. Канасевич Э.Р. Анализ временных последовательностей в геофизике. - М.: Недра, 1985.- 300 с.
8. Купер Дж., Макгиллем А. Вероятностные методы анализа сигналов и систем. - М.: Мир, 1989.
9. Макс Ж. Методы и техника обработки сигналов при физических измерениях: В 2-х томах. - М.: Мир, 1983.
10. Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов. - М.: Связь, 1979. - 416 с.
11. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. - М.: Мир, 1982. - 428 с.
12. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. - М.: Мир, 1978.
13. Хемминг Р.В. Цифровые фильтры. - М.: Недра, 1987. - 221 с.
14. Сергиенко А.Б. Цифровая обработка сигналов. - СПб.: Питер, 2003. - 608 с.
15. Хуанг Т.С. (ред.)_Быстрые алгоритмы в цифровой обработке изображений. М., Радио и Связь, 1984. - 222 с.
16. Корн Г., Корн Е. Справочник по математике для научных работников и инженеров. - М.: Наука, 1984.
17. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения М.: Высшая школа, 2000. - 480с.
18. Пытьев Ю.П., Шишмарев И.А. Курс теории вероятностей и математической статистики для физиков. М.: Изд-во Моск. ун-та, 1983. - 256с.
19. Пугачев В.С. Теория вероятностей и математическая статистика. М.: Наука, 1979. - 496с.
20. Чистяков В.П. Курс теории вероятностей. М.: Наука, 1982. - 256с.
21. Боровков А.А. Теория вероятностей. М.: Наука, 1976. - 352с.
22. Кремер Н.Ш. Теория вероятностей и математическая статистика. М.: ЮНИТИ, 2000. - 543с.