Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего профессионального образования
«Севастопольский государственный университет»

С.В.Доценко

ДОПОЛНИТЕЛЬНЫЕ РАЗДЕЛЫ ТЕОРИИ ИНФОРМАЦИИ

Севастополь
2016

УДК 004.415.2

Конспект лекций составлен в соответствии с требованиями программы дисциплины «Дополнительные разделы теории информации» для студентов направления 09.04.02 и утверждены на заседании кафедры «Информационные системы»,
протокол №   от «   »________ 201 г.

Допущено учебно-методическим центром СевГУ

Рецензенты

Содержание

Введение

1. **Системы счисления**
2. **Задача хранения данных**
3. **Количество информации. Энтропия**
4. **Свойства энтропии**
5. **Энтропия совокупности случайных величин**
6. **Взаимная информация**
7. **Средняя взаимная информация**
8. **Энтропия непрерывных сообщений (непрерывных случайных величин)**
9. **Средняя взаимная информация непрерывных сообщений (непрерывных случайных величин)**
10. **Нормальное распределение**
11. **Двумерное нормальное распределение**
12. **Многомерное нормальное распределение**
13. **Энтропия и взаимная информация многомерных совокупностей гауссовских случайных величин**
14. **Преобразование формулы для взаимной информации**
15. **Примеры вычисления взаимной информации непрерывных случайных величин**
16. **Примеры вычисления взаимной информации дискретных случайных величин**

Заключение

Библиография

Введение

Теория информации — математическая наука. Этот факт подчеркивал и ее автор К. Шеннон, назвав свою первую, фундаментальную, работу, посвященную теории информации, «Математическая теория связи» [17]. Она позволила поставить ряд практически важных вопросов применения вычислительной техники для обработки и передачи больших объемов различного рода данных и дать количественные ответы на них.

В развитие теории информации внесен большой вклад отечественными и зарубежными математиками различных направлений. Однако в процессе выполнения исследований они стали применять понятия и методы теории информации к абстрактным математическим объектам всё более сложной структуры, нередко очень далеким от имеющих практическое значение. Это требовало от исследователя знания целого ряда современных разделов математики, не входящих в программы обучения специалистов по информационным технологиям и системам. Когда удельный вес математически сложных работ стал превалирующим, появился серьёзный отрыв математиков, работающих в области теории, от инженеров, занятых проектированием и применением информационных систем и систем связи. Это привело к противоестественной разобщенности теоретиков и инженеров и потере последними интересов к этой теории.

Цель настоящего пособия — найти достаточно простые подходы к понятиям теории информации и дать им физическое и/или информационное толкование, позволяющее осознанно применять эти понятия на практике.

1. Системы счисления

Одним из наиболее общих понятий науки, обозначающих некоторые сведения, совокупность каких-либо данных, знаний и т.п., является понятие *информации* (лат. *informatio* – разъяснение, изложение, осведомленность). Преобразование, переработка потоков информации, т.е. реализация информационного процесса, осуществляется *информационной системой* (ИС).

Существует особый вид информации, называемый «данными». *Данные* — это информация, представляемая обычно в цифровой форме или в каком-либо другом формализованном виде, предназначенная для обработки на компьютере (метеоданные, статистические данные, данные опросов и т.д.). Данные легко закодировать и передавать в форме дискретных электрических сигналов.

Любому дискретному сообщению или знаку сообщения можно приписать некоторый порядковый номер. Передача или хранение сообщений при этом сводится к передаче или хранению чисел, которыми описываются такие номера. Числа же можно выразить в какой-либо системе счисления. Таким образом, будет получен один из кодов, основанный на данной системе счисления.

Счисление — совокупность приемов представления *натуральных* чисел [2]. В любой *системе счисления* некоторые символы (слова или знаки) служат для обозначения определенных чисел и называются *узловыми.* Остальные числа (*алгоритмические*) получаются в результате каких-либо операций из узловых чисел. Системы счисления различаются выбором узловых чисел и способами образования из них алгоритмических чисел. С появлением письменных обозначений числовых символов системы счисления стали различаться характером числовых знаков и принципами их записи.

Например, у древних вавилонян узловыми являлись числа 1, 10, 60; у маори (коренных жителей Новой Зеландии) узловыми были числа 1, 11, 11^2, 11^3. В римской системе счисления узловыми являются числа 1, 5, 10, 50, 100, 500, 1000, обозначаемые соответственно знаками I, V, X, L, C, D, M.

Системы счисления, в которых алгоритмические числа образуются сложением узловых, называются *аддитивными*.

В некоторых системах счисления, называемых *алфавитными*, числа обозначались теми же знаками, что и буквы. Такие изображения чисел употребляли славяне и многие другие народы, например, римляне.

Система счисления называется *непозиционной,* если каждый числовой знак в записи любого числа в ней имеет одно и то же значение. Если же значение числового знака зависит от его расположения в записи числа, то система называется *позиционной.* Римская система счисления — непозиционная. Современная десятичная система счисления — позиционная

Общепризнанным в настоящее время является позиционный принцип образования системы счисления. Значение каждого символа (цифры) зависит от его положения — *позиции* в ряду символов, представляющих число (т.е. от его *разряда*).

Для представления чисел в позиционной системе применяют цифры или, точнее, цифровые ряды или (цифровые) слова, которые образуются путем упорядочения конечного числа основных знаков из конечного множества основных знаков (*алфавита* системы счисления). С этой целью выбирают некоторое *натуральное* число p, большее единицы, и используют его в качестве *базисного* числа (p-ичная система счисления). (Для p, равного единице, позиционной системы счисления не существует). Вводят p основных знаков, называемых *цифрами*. Эти знаки используют для образования цифровых последовательностей, которые служат для представления натуральных чисел. Будем цифры обозначать так: $b_0, b_1, \ldots, b_{p-1}$. Каждой цифровой последовательности, образованной только из одной цифры, однозначно сопоставим одно из p - 1 первых *натуральных* чисел или нуль. (Напомним, что наименьшее натуральное число равно 1, а каждое следующее получается из предыдущего добавлением единицы, т.е. последовательность натуральных чисел бесконечна и имеет вид 1, 2, 3, … . Ноль 0 не является натуральным числом и в наших рассуждениях должен быть введен дополнительно). Тогда всякое натуральное число a имеет точно одно представление в p - ичной системе счисления:

$$a = b_s p^s + b_{s-1} p^{s-1} + \ldots + b_0 p^0, \tag{1}$$

где s обозначает однозначно определенное натуральное число, $b_0, \ldots, b_s$— цифры, причем b_s— цифра, отличная от цифры, соответствующей нулю. Ряд знаков $b_s b_{s-1} \ldots b_0$ является цифровым представлением числа a. Число нуль представляется последовательностью, состоящей из одной цифры, соответствующей нулю. В позиционной системе представляемое число образуется аддитивно, причем каждая цифра b_j имеет *числовое значение* (число, которое соответствует цифре b_j) и *позиционное значение* (вес) p^j, если b_j стоит на j - м месте, считая справа (счет начинают с нуля, а не с единицы!) Аддитивные вклад этой цифры в значение числа a равен $b_j p^j$. Иными словами, для записи числа a применяется не формула (1), а выражение $a=b_s b_{s-1} \ldots b_0$.

Чем больше основание p системы счисления, тем меньше разрядов требуется для представления данного числа. Например, двузначное в десятичной системе число 29 в двоичной системе требует пять разрядов (11101). Действительно, здесь $p=2$, т.е. алфавит системы счисления состоит из двух чисел 0 и 1, и выражение (1) принимает вид 29 = $29 = 1 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$.

Однако с ростом основания существенно повышаются требования к линии связи и аппаратуре создания и распознавания элементарных сигналов, соответствующих различным символам. Логические элементы вычислительных устройств при этом должны иметь большее число устойчивых состояний.

Учитывая сказанное, целесообразно выбрать такую систему счисления, которая обеспечивает минимум произведения количества различных символов p основания системы счисления на количество разрядов s для выражения любого числа. Такой минимум имеется. Найдем его.

Выберем k - й разряд. Он в формуле (1) задается структурой

$$y=p^k. \tag{2}$$

Необходимо для этой структуры найти такое p, которое минимизирует величину

$$c=pk. \tag{3}$$

Из выражения (2) найдем $k=\ln y / \ln p$ и подставим в (3):

$$c = \ln\ y\ p / \ln\ p\ .$$

Минимум этого выражения достигается при $p = e = 2.718281828$ для любых k. Это число иррационально и не может служить основанием системы счисления. Полученный результат свидетельствует о том, что наиболее эффективной системой является троичная, основание которой является самым близким к e. Незначительно уступают ей двоичная и четверичная. Системы с основанием 10 и большим существенно менее эффективны. С точки зрения удобства физической реализации соответствующих логических элементов и простоты выполнения в них арифметических и логических операций предпочтение необходимо отдать двоичной системе.

Приведем примеры записи натуральных чисел в некоторых позиционных системах.

***Десятичная система*: $p = 10$, цифры 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Например,**

$$3 \cdot 10^5 + 2 \cdot 10^4 + 0 \cdot 10^3 + 0 \cdot 10^2 + 9 \cdot 10^1 + 1 \cdot 10^0 = 320091.$$

В двоичной системе оно имеет вид 1001110001001011011.

***Двоичная система* (*бинарная*, или *диадная* система): $p = 2$, цифры 0, 1. Например,**

$$1 \cdot 2^6 + 0 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 1010011.$$

Записанная в двоичной системе последовательность обозначает число, значение которого в десятичной системе есть $64 + 0 + 16 + 0 + 0 + 2 + 1 = 83$.

Чтобы иметь возможность представлять в позиционной системе также и некоторые рациональные числа, позиционные значения цифр в записи числа распространяют на степени p с отрицательными показателями.

В каждой позиционной системе можно точно представить только некоторое подмножество рациональных чисел, зависящее от выбора p. Например, рациональное число $1/3$ не может быть представлено в десятичной системе счисления в виде конечной последовательности цифр.

Сложение натуральных чисел вновь дает натуральное число. Вычитание натуральных чисел может привести к отрицательному, т.е. целому, но *не* натуральному, числу. Умножение натуральных чисел дает натуральное число, деление же их часто приводит к рациональному числу. Отсюда следует, что для того, чтобы при любых операциях над натуральными числами оставаться в пределах натуральных чисел, допустимо применять только операции сложения и умножения. Системы счисления, основанные на этом принципе, называются аддитивно-мультипликативными. Все известные позиционные системы счисления являются аддитивно-мультипликативными.

2. Задача хранения данных

Информационные процессы обычно связаны с необходимостью получения, хранения, передачи и обработки больших массивов данных. Однако ясно, что одинаковые по объему массивы данных не всегда содержат равные по объему сведения об интересующих нас процессах и явлениях.

Оценить информационные качества данных позволяет *теория информации*. Рассмотрим некоторые положения, лежащие в ее основе. Остановимся вначале на задаче *хранения данных*.

Данные о процессах, явлениях, событиях и т.д. могут принимать форму некоторых символов, графиков, цифр и т.д. Любое устройство хранения данных будем называть *накопителем*. Какой бы вид не имели данные, их представление должно быть таким, чтобы при необходимости можно было из хранимых данных восстановить оригинал процесса или сообщения или его некоторый эквивалент.

Ясно, что при хранении данных нет необходимости, а часто и невозможно использовать в накопителе тот же физический носитель информации, который применялся для их первичного представления. Например, текст не обязательно хранить в буквенном виде, а можно записать магнитным или электрическим способом, сведения об изменении температуры можно представить в виде кривой на ленте самописца и т.д. Звук же хранить в виде звука вообще невозможно. То же можно сказать о температуре, освещенности и т.д. Необходимым условием представления данных является обратимость этого представления, т.е. возможность получения исходных сведений по находящемся в накопителе данным.

Для хранения данных необходимо, чтобы сведения о каждом из возможных событий, содержащих первичную информацию, могли быть представлены в накопителе. Это означает, что пустой накопитель должен обладать лишь способностью переводиться в некоторое число различных состояний, природа же этих состояний совершенно несущественна для решения вопроса о том, какие данные в нем можно и нужно запомнить. Необходимо лишь, чтобы эти состояния были *отождествимы и различимы*. Поэтому накопители, предназначенные для хранения одной и той же информации, могут быть выполнены на различных физических принципах.

Способность пустого накопителя усваивать данные зависит от общего числа различных состояний, в которых он может находиться, и характеризуется *информационной емкостью.* Дадим количественное определение этой величины.

Пусть имеются два частных накопителя, каждый из которых имеет 5 состояний. В первом из них – состояния $X = \quad 1,2,3,4,5 \quad$, во втором – состояния $Y = \quad a,b,c,d,e \quad$. Общее число состояний, в которых они могут пребывать совместно, равно $5^2 = 25$. Действительно, эти состояния — элементы декартова произведения

$$X \times Y = \quad 1,a \; , \; 1,b \; ,\ldots, \; 1,e \; ,\ldots, \; 2,a \; ,\ldots, \; 1,e \; ,\ldots, \; 5,e \quad .$$

Обратим внимание, что общие состояния этой пары накопителей представляют собой кортежи — упорядоченные множества. Это означает, что первый накопитель обязательно находится слева от второго, т.е. позиции этих накопителей должны быть упорядочены. С точки зрения терминологии систем счисления расположение частных накопителей должно быть *позиционным*. Только при выполнении этого условия он может быть переведен в любое из 25 состояний.

Если каждый из двух частных накопителей имеет n возможных состояний, то общее число состояний, в которых они могут пребывать совместно, равно $N = n^2$. Однако естественно считать, что коль скоро здесь *два* одинаковых накопителя, то информационная емкость общего суммарного накопителя *вдвое больше* информационной емкости одного из накопителей. Эти соображения подсказывают *логарифмическую меру* информационной емкости накопителя. Эта мера была введена в 1928 г. американским инженером-связистом Р.В.Л.Хартли посредством соотношения

$$C = \log N, \tag{4}$$

где N — число различимых состояний накопителя. Она носит название *мера Хартли информационной емкости* накопителя. Выбор основания логарифма в этом выражении определяет *единицу информационной емкости.* В технических приложениях чаще всего пользуются логарифмами при основании *2*. Это хорошо согласуется с применяемой в электронных цифровых электрических машинах двоичной системой счисления. При этом информационная емкость C измеряется в двоичных единицах, которые имеют сокращенное название *бит* (*bit* — сокращение английского *binary digit* — «двоичная единица»). При математических операциях, связанных с дифференцированием, часто удобнее пользоваться натуральными логарифмами, основание которых равно числу $e = 2{,}718\ldots$ При этом информационная емкость измеряется в натуральных единица (*«натах»)*, а при основании логарифмов *10* — в десятичных логарифмах *(«дитах»)*. С принципиальной точки зрения выбор основания логарифмов не имеет значения, так как по известной формуле $\log_b k = \log_a k \cdot \log_b a$ легко перейти от одной системы логарифмов к другой.

Накопитель, как правило, может быть расчленен на совокупность одинаковых элементарных накапливающих ячеек (двоичных, десятичных и др.). Тогда число его возможных состояний *при условии упорядочения* этих ячеек в силу изложенного выше есть

$$N = p^{n}, \tag{5}$$

где p — число состояний одной элементарной ячейки, а n — число таких ячеек. При этом информационная емкость накопителя $C = n \log p$ есть произведение информационной емкости одной элементарной ячейки на число таких ячеек в накопителе. В частности, если основание логарифма *2*, то C есть эквивалентное число двоичных накапливающих ячеек, т.е. число двоичных разрядов числа N.

Обратимся к формуле (5). Найдем логарифм величины N по основанию p, т.е. по величине емкости одного элементарного накопителя:

$$\log_p N = n \log_p p = n\,.$$

Здесь учтено, что для любого p имеет место тождество $\log_p p = 1$.

Следовательно, если взять логарифм числа N по основанию, равному емкости элементарной ячейки, то получим число элементарных ячеек n, необходимых для хранения этого числа. Так, например, для хранения любого из 32 чисел от 0 до 31 в двоичном коде необходимо $n = \log_2 32 = 5$ ячеек.

Далеко не все числа представляют собой целые степени числа p. Например, $\log_2 35 = 5.129\ldots$ Так как число ячеек должно быть целым, то полученный результат следует округлить до целого. Если выбрать ближайшее меньшее целое 5, то некоторые числа из заданных 35 будут искажены. Поэтому следует выбирать большее число ячеек 6. При этом будет некоторая избыточность емкости накопителя, но все числа будут храниться без искажений. (*Замечание.* Функция $\log x$ предназначена для работы с числами десятичной позиционной системы счисления. Применению ее с числами иной системы счисления должно предшествовать преобразование этих чисел в десятичную систему. То же касается и других математических функций).

Обратим особое внимание на то, что элементарные накапливающие ячейки для обеспечения возможности хранения в общем накопителе максимально возможного количества данных должны быть упорядочены, и этот порядок должен сохраняться при всех операциях над этими данными. Такое свойство рассматриваемых накопителей хорошо согласуется с позиционными системами счисления, и потому в информационных технологиях применяются именно эти накопители. При этом каждый частный накопитель предназначен для хранения информационного символа своего разряда.

Такое соотношение между записью числа в позиционной системе и применяемых для его хранения ячеек накопителя дано на *Рис.1*.

$$\boldsymbol{a}=\boldsymbol{b}_S\boldsymbol{p}^S+\boldsymbol{b}_{S-1}\boldsymbol{p}^{S-1}+\ldots+\boldsymbol{b}_0\boldsymbol{p}^0$$

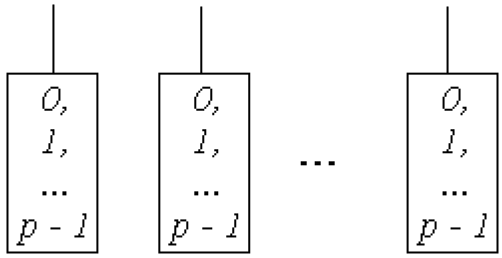


Рис. 1.- Запись числа в позиционной системе и соответствующее ему положение ячеек накопителя

Подчеркнем, что понятие «информационная емкость» характеризует физическую систему, предназначенную для хранения данных, а, следовательно, и максимальный объем данных, которые в ней могут находиться. Однако оно не дает возможность определить количество информации об исследуемом процессе или явлении, содержащемся в этих данных. Отличие количества информации, содержащегося в данных об исследуемом явлении, от их объема, совпадающего с информационной емкостью накопителя, предназначенного для их хранения, определяется статистической структурой этих данных и может быть найдено при математическом анализе последней. В частности, данные, содержащиеся в накопителе, могут вообще не содержать интересующей нас информации.

3 Количество информации. Энтропия

Информация имеет две стороны: количественную и качественную. Нас будет интересовать ее количественная сторона.

Первая работа, посвященная нахождению количественной меры информации, была опубликована Клодом Шенноном в 1948 г. [17]. В ней показано, что в качестве меры количества информации, возможности выбора из некоторого множества событий и неопределенности должна быть выбрана удовлетворяющая определенным условиям функция

$$H(X) = -\sum_{i=1}^{k} P(x_i)\log_2 P(x_i),$$

которую он назвал *энтропией*. А.Н. Колмогоров подчеркивал [8]: «Первые работы Шеннона были написаны на уровне физической строгости, противопоказанной познавательным возможностям среднего математика». Поэтому появились работы, ориентированные не на математиков, а на специалистов, работающих в области информационных технологий. Они лишены той математической строгости, которая присуща работам профессиональных математиков.

Последователи Шеннона использовали различные подходы к определению количественной меры информации. Так, в работах [4] и [7] она вводится через логарифм отношения вероятностей, в [13] при помощи введения величины исчезнувшей неопределенности, в [18] через принятую меру неопределенности, и т.д. Многие из них используют в качестве исходного загадочный термин «неопределенности», что не улучшает понимания сути дела.

Ясность в эту проблему внес академик А.Н. Колмогоров [8], применив аппарат комбинаторики. Рассмотрим этот подход более подробно.

Для хранения и передачи любого дискретного сообщения всегда используется некоторый (обычно конечный) алфавит $X = x_1, x_2, \dots, x_k$. Величина k есть число элементов данного алфавита. В бинарном алфавите $k = 2$, в десятичном $k = 10$, в кириллице $k = 33$, в латинице $k = 27$ (здесь включены пробелы, но не включены знаки препинания). Будем считать, что объем памяти, занимаемый каждым символом алфавита, одинаков. Сообщения в накопителе отличаются друг от друга порядком составляющих их элементов алфавита. Например, оба слова *сокол* и *колос* содержат по 2 буквы «*о*», по одной «*к*» и «*с»*, однако несут разные сообщения. Следовательно, содержание сообщения зависит от того, как в нем расставлены элементы алфавита (в нашем случае – буквы).

Если все буквы различны, то, как известно из комбинаторики, из n букв можно составить максимум $n!$ различных перестановок [1]. Например, из букв *а, б, р* можно составить 3! = 6 различных трехбуквенных слов (сообщений), а именно, *абр, бар, бра, рба, раб, арб*. (Мы здесь не обращаем внимания на то, имеют ли эти слова какой-либо смысл).

Если же некоторые переставляемые символы одинаковы, то получается меньше возможных различных перестановок – некоторые перестановки будут совпадать друг с другом. Допустим, что имеется k различных элементов $x_1, x_2, ..., x_k$ алфавита X. Если в сообщении содержится n_1 элементов x_1, n_2 элементов $x_2, \ldots,$ n_k элементов x_k и полная длина сообщения есть $n = n_1 + n_2 + \ldots n_k$, то максимальное число различных сообщений (слов), которые можно создать из данных элементов, есть

$$\hat{N} = \frac{n!}{n_1!\, n_2! \ldots n_k!}.$$

В частности, из букв слова *сокол* можно составить $5!/(1!\,2!\,1!\,1!\,1!) = 60$ пятибуквенных слов, а не $5! = 120$, как из букв слова *спорт.*

Итак, последовательность n символов выбранного алфавита может нести не более, чем $\hat{N}$ различных сообщений. Следовательно, величина $\hat{N}$ характеризует *степень разнообразия* этих сообщений.

Логарифмическая мера этого числа сообщений

$$\hat{H} = \log \hat{N} = \log(n!) - \sum_{i=1}^{k} \log(n_i!) . \qquad (6)$$

Она представляет собой число элементарных ячеек памяти, объем каждой из которых равен основанию используемого логарифма, необходимое и достаточное для хранения числа $\hat{N}$. Если этих ячеек будет больше указанного, часть памяти окажется не занятой, если меньше — некоторые элементы сообщений будут искажены.

Преобразуем полученное выражение, считая, что сообщение имеет большую длину, т.е. $n >> 1$. Воспользуемся формулой Стирлинга для логарифма факториала большого n [8]

$$\log\, n! \approx n \log n. \qquad (7)$$

Эта формула тем точнее, чем больше число n. С ее учетом из (6) имеем

$$\hat{H} \approx n \log n - \sum_{i=1}^{k} n_i \log n_i . \qquad (8)$$

Допустим, что создается некоторое случайное сообщение длины n, состоящее из различных элементов x_i алфавита X. Примем во внимание, что в любом сообщении элементы алфавита появляются с некоторыми, вообще говоря, разными частотами (в частности, это относится к языковым текстам). Частота появления элемента x_i есть

$$q_i = n_i / n, \qquad (9)$$

и справедливо равенство (условие нормировки)

$$\sum_{i=1}^{k} q_i = \frac{1}{n} \sum_{i=1}^{k} n_i = 1. \qquad (10)$$

Выразив из (9) $n_i = n q_i$, подставив его в (8) и воспользовавшись формулой (10), получим, что логарифмическую меру всех возможных сообщений, состоящих из n символов данного типа можно записать как

$$\hat{H} \approx -n \sum_{i=1}^{k} q_i \log q_i .$$

При $n \to \infty$ частота q_i стремится к вероятности $P(x_i)$ появления в сообщении элемента x_i. В пределе имеем количество указанных выше элементарных ячеек, приходящихся на один символ сообщения:

$$H(X) = \lim_{n\to\infty}\left(\frac{\hat{H}}{n}\right) = -\sum_{i=1}^{k} P(x_i)\log P(x_i).$$

Полученный результат можно выразить так: *при употреблении отдельных букв (символов) с вероятностями* $P(x_i)$ *количество информации, приходящейся на одну букву (символ) сообщения, равно*

$$H(X) = -\sum_{i=1}^{k} P(x_i)\log P(x_i). \qquad (11)$$

Эта величина носит название *энтропии*. Будучи связанной с $\hat{N}$, она характеризует *степень разнообразия значений, приходящегося в среднем на один символ* при условии, что длина сообщения стремится к бесконечности.

Если в формуле (11) выбрать двоичный логарифм, то величина

$$H(X) = -\sum_{i=1}^{k} P(x_i)\log_2 P(x_i) \qquad (12)$$

характеризует собой *минимальное среднее количество двоичных единиц (бит), приходящихся на один символ, необходимое и достаточное для того, чтобы безошибочно хранить или передавать любое сообщение из данного ансамбля при условии, что* n *велико*. Если окажется, что в системе предусмотрено количество бит на символ меньшее, чем $H(X)$, некоторые сообщения будут переданы с ошибками; если большее, чем $H(X)$, – система будет избыточна для такого сообщения.

Измеряется энтропия в тех же единицах, что и информационная емкость накопителя (в битах, натах и т.д.). В обозначении энтропии символ X не является аргументом функции, а указывает множество, на котором определена дискретная случайная величина ξ.

Формула (12) содержит вероятности $P(x_i)$ символов сообщения x_i. На практике исследователь обычно имеет дело с частотами q_i этих символов. Эти частоты появляются при применении в формуле (6) приближенной формулы Стирлинга (7). Последняя тем ближе к точному равенству, чем большая длина n исследуемого сообщения. Выясним, как изменяется отношение правой и левой частей приближенного равенства при увеличении длины n. Результаты расчетов представлены в Таблице 1. В ней принято обозначение

$$\lambda_m(n) = \frac{n \log_m n}{\log_m(n!)} . \qquad (14)$$

Таблица 1

n	2	10	100	1000	10 000
$\lambda_m(n)$	2	1.524	1.266	1.168	1.122

Из нее следует, что это отношение не зависит от основания логарифма m, не превышает 2 и с ростом n медленно приближается к 1.

Полученные результаты позволяют объяснить понятие «неопределенность» и указать ее количественную меру [12].

Результаты наблюдения случайных явлений нельзя определить заранее (*a priori*), так как им присуща некоторая *неопределенность*. Эта неопределенность может быть большей или меньшей в зависимости от той начальной информации, которая нам о них известна. Так, если на пешеходном переходе выдает сигналы светофора «красный» и «зеленый» встречаются нами с вероятностями соответственно 0.9 и 0.1, то мы вправе с большой степенью уверенности ожидать появления именно сигнала «красный», причем начальная неопределенность рассматриваемого явления здесь сравнительно невелика. Если же сигналы «красный» и «зеленый» равновероятны, то мы не можем в своих предположениях относительно вида сигнала отдать предпочтение какому-либо из них, потому что здесь неопределенность опыта больше, чем в предыдущем случае.

Таким образом, априори можно только указать число возможных исходов опыта (или число возможных состояний некоторого объекта) и указать распределение вероятностей по этим исходам. Однако эти величины, взятые порознь, хотя и могут отражать степень неопределенности результатов наблюдений, но не дают четкой количественной меры неопределенности — единственного числа, ее характеризующего. Кроме того, при сложных событиях число возможных исходов и вероятностей их появления становится достаточно большим, и непосредственное использование этих величин теряет всякую наглядность и обозримость.

Естественно считать, что количественной мерой неопределенности результатов наблюдения случайных явлений является энтропия. Эта мера связана с видом (формой) распределения вероятностей исхода опыта, но не с множеством конкретных значений наблюдаемых явлений. Она не зависит от конкретных значений случайной величины и является в общем случае функционалом распределения вероятностей. Кроме того, как будет показано ниже, мера неопределенности имеет максимум, отвечающий наибольшей неопределенности при некотором распределении вероятностей, что соответствует интуитивному представлению неопределенности. Отсюда следует, что энтропию можно истолковать как *количественную меру неопределенности о состоянии* дискретной величины до ее измерения, т.е. как то количество информации, которое должно быть в среднем получено для опознавания любого из состояний множества $X=\{x_1,x_2,...,x_N\}$.

Рассмотрим конкретные примеры.

П**ример 1**. Допустим, что нам необходимо найти квартиру, в которой проживает наш знакомый. Нам известен номер его дома, известно, что дом восьмиэтажный и на каждом этаже расположены четыре квартиры. Неизвестно, на каком этаже и в какой квартире этого этажа проживает знакомый. Выясним степень неопределенности решаемой задачи. Для этого найдем ее энтропию.

Если ничего не известно о возможном этаже и номере квартиры знакомого, в соответствии с принципом равных возможностей следует считать все квартиры дома равновероятными, т.е. $P(x_i)=1/32$. Энтропия такого распределения, есть $H_{32}(X)=-\sum_{i=1}^{32}P(x_i)\log_2 P(x_i)=5\,бит$. Допустим теперь, что нам стал известен этаж, на котором проживает знакомый. Это означает, что его квартира — одна из четырех квартир этажа. Энтропия такого распределения есть $H_4(X) = 2$ *бит*. Как и следовало ожидать, неопределенность в последнем случае меньше, чем в первом. Здесь, как говорят, новое знание, позволило уменьшить (снять) начальную неопределенность.

Естественно, что если бы вероятности квартир были различными, то и неопределенность, т.е. энтропия была иная. □

Для получения дальнейших результатов обратимся к понятиям теории вероятностей.

Распределения вероятности $P(x_i)$ исчерпывающим образом характеризует дискретные случайные величины с вероятностной точки зрения. Эти законы распределения представляют собой некоторые функции от возможных значений исследуемых случайных величин.

Однако во многих вопросах практики нет необходимости характеризовать случайную величину полностью, исчерпывающим образом. Зачастую достаточно бывает указать только отдельные ее числовые параметры, например, такие как некоторое среднее значение, около которого группируются возможные значения случайной величины, или число, характеризующее степень разбросанности этих значений относительно среднего, и т.д. Эти *неслучайные* параметры, назначение которых — выразить в сжатой форме наиболее существенные особенности распределения, называются *числовыми характеристиками* случайной величины.

Числовые характеристики обычно являются математическим ожиданием (средним значением) некоторой заданной *детерминированной функции* $\varphi(\xi)$ от случайной величины ξ_i, что можно записать следующим образом

$$\mathrm{M}\left[\varphi(\xi)\right]=\sum_{i=1}^{k}\varphi(x_i)P(x_i),$$

где $\mathrm{M}\left[\xi\right]$ — обозначение математического ожидания случайной величины ξ.

Сравнивая этого выражение с формулой (11), видим, что последнюю можно записать как

$$H(X)=\sum_{i=1}^{k}I(x_i)P(x_i), \tag{13}$$

где обозначена

$$I(x_i)=-\log P(x_i). \tag{14}$$

величина, осредняемая в формуле (13).

Величину $I(x_i)$ естественно интерпретировать как *частную* информацию, которую несет конкретный символ x_i. Ее называют *собственной информацией* символа (или события) x_i. Она представляет собой количество информации, доставляемое *самим событием* x_i или любым другим, однозначно связанным с x_i. Это количество тем *больше*, чем *меньше* вероятность появления события x_i. Действительно, сенсационное известие, вероятность которого мала, несет в себе много больше информации, чем заурядное событие, вероятность которого велика. При этом выражение (1) дает математическое ожидание (среднее значение) такой информации по всем возможным символам.

Пример 2. Выясним, как выглядит собственная информация символа на примере дискретного *биномиального распределения*. Оно описывается следую

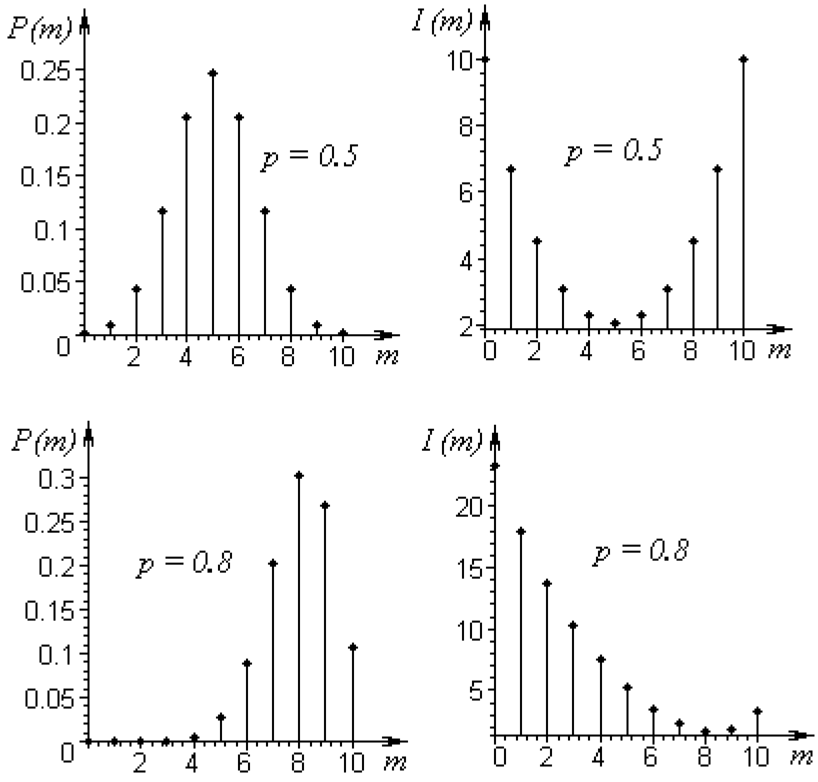


Рис.2.-. Биномиальное распределение и частная информация его составляющих

щей формулой для вероятности m успехов в n испытаниях при параметре p:

$$P(m) = \frac{n!}{m!\,(n-m)!}\,p^{m}\,(1-p)^{n-m}, \quad 0 \le m \le n\ .$$

Частная информация (14) в битах здесь принимает вид $I(m) = -\log_2 P(m)$.

Графики, соответствующие обеим последним формулам при $n = 10$, представлены на *Рис.2.* Они дают количественную величину введенных понятий. □

4 Свойства энтропии

Рассмотрим основные свойства энтропии [6].

1. Энтропия является неотрицательной вещественной величиной, так как для любого i $1 \le i \le k$ вероятность $P(x_i)$ изменяется в интервале от 0 до 1, $\log P(x_i)$ отрицателен и, следовательно, каждое ее слагаемое $-P(x_i)\log P(x_i)$ неотрицательно.

2. Минимальное значение энтропии равняется нулю. Оно достигается только в том случае, если вероятность одного из состояний равна единице (т.е. это состояние является достоверным, неслучайным событием). Тогда вероятности всех остальных состояний равны нулю. При этом слагаемое, соответствующее достоверному событию, есть $-1 \cdot \log(1) = 0$, а слагаемые, соответствующие остальным событиям, также равны нулю: $\lim\limits_{P(x_i) \to 0}\left(-P(x_i)\log P(x_i)\right) = 0$. Отсюда следует, что энтропия достоверного (неслучайного) события минимальна и равна нулю.

3. Энтропия максимальна, когда все элементы x_i алфавита равновероятны. Докажем это свойство с помощью метода неопределенных коэффициентов Лагранжа нахождения условного экстремума функции нескольких переменных, которые связаны между собой одним или несколькими условиями [15]. В нашем случае такой функцией будет

$$H\left(P_1,P_2,...,P_k\right)=-\sum_{i=1}^{k}P_i\ln P_i, \qquad (15)$$

где для краткости обозначено $P_i=P\left(x_i\right)$. Здесь выбран натуральный логарифм с целью упрощения решения поставленной задачи. Переменные P_i, от которых зависит энтропия, связаны известным условием

$$\sum_{i=1}^{k}P_i-1=0. \qquad (16)$$

Введем неопределенный коэффициент λ и рассмотрим следующую функцию $k+1$ переменных

$$\Phi\left(P_1,P_2,...,P_k,\lambda\right)=H\left(P_1,P_2,...,P_k\right)+\lambda\left(\sum_{i=1}^{k}P_i-1\right), \qquad (17)$$

состоящую из суммы выражения (15) и условия (16), умноженного на λ. Как известно, в методе Лагранжа нахождение условного экстремума функции (15) при условии (16) заменяется более простой задачей определения безусловного экстремума функции (17).

Необходимые условия экстремума функции $\Phi\left(P_1,P_2,...,P_k,\lambda\right)$ находятся из системы уравнений $\partial\Phi/\partial P_i=0,\ \ i=1,2,...,k$ и (19). Продифференцировав (17) по P_i с учетом (15), найдем, что экстремум обеспечивает P_i, удовлетворяющее равенству $\ln P_i=\lambda-1$. Обратим внимание, что правая часть этого равенства не зависит от номера символа i. Следовательно, все P_i одинаковы и, согласно (16), равны $1/k$ независимо от основания логарифма в формуле (15). Соответствующая энтропия максимальна и равна $H_{max}=\log k$.

Из всего изложенного в данном разделе следует, что энтропия дискретного сообщения лежит в пределах $0 \le H(X) \le H_{max} = \log k$. Конкретное значение этой величины зависит от вида распределения ее символов.

Пример 3. Допустим, некоторый символ может с одинаковой вероятностью может принимать 4 значения: $P = \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}$. Энтропия (неопределенность) такого распределения максимальна и равна $H_{max} = \log_2 4 = 2$ *бит.* Если же распределение вероятностей этого символа имеет вид $P = \left\{\frac{1}{8}, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}\right\}$,

то его энтропия, согласно формуле (12), есть

$$H(X) = -\left[\frac{1}{8}\log_2\left(\frac{1}{8}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{1}{8}\log_2\left(\frac{1}{8}\right)\right] =$$

$$= \frac{1}{8}\log_2 8 + \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 = \frac{3}{8} + \frac{1}{2} + \frac{2}{4} + \frac{3}{8} = \frac{7}{4},$$

что меньше, чем для предыдущего распределения. □

Пример 4. Пусть алфавит источника информации имеет два символа: $X = \{x_1, x_2\}$. Такой источник называется *бинарным* (*двоичным*). Обозначим через p вероятность символа x_1, т.е. $P(x_1) = p$. Тогда вероятность второго символа $P(x_2) = 1 - p$. Такое распределение носит название *распределения Бернулли*. Его энтропия

$$H_{дв}(X) = -p\log_2 p - (1-p)\log_2(1-p).$$

График этой зависимости представлен на *Рис.3.* Из него видно, что как и в общем случае $H_{дв}(X) = 0$ только тогда, когда достоверно известен либо символ x_1 (т.е. $p = 1$), либо символ x_2 (т.е. $1 - p = 1$, откуда $p = 0$).

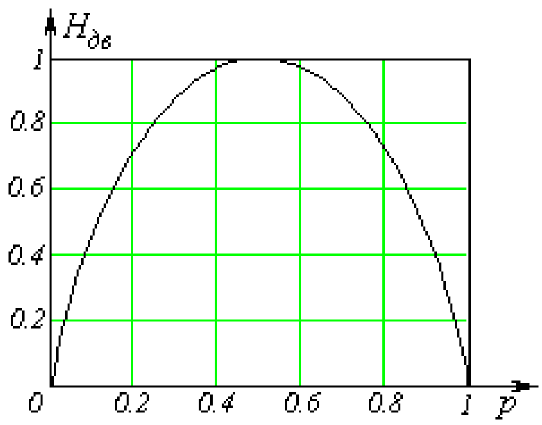


Рис.3.. - Энтропия двоичного источника

Величина $H_{дв}(X)$ максимальна и равна 1 *бит* при условии равной вероятности символов x_1 и x_2, т.е. при $p = 1/2$. Действительно, интуитивно ясно, что при таком распределении вероятностей неопределенность появления любого символа наибольшая. □

5 Энтропия совокупности случайных величин

При оценке неопределенности выбора часто бывает необходимо учитывать статистические связи, которые в большинстве случаев имеют место как между состояниями двух или нескольких источников, объединенных в одну систему, так и между состояниями, последовательно создаваемыми одним источником.

Определим энтропию объединения двух статистически связанных ансамблей $X = \{x_1, x_2, \ldots, x_m\}$ и $Y = \{y_1, y_2, \ldots, y_n\}$. *Объединение ансамблей* описывается в общем случае прямоугольной матрицей $\mathbf{P}\ X, Y$ вероятностей

$P\left(x_i, y_j\right)$ всех возможных комбинаций состояний x_i $1 \le i \le m$ ансамбля X и состояний y_j $1 \le j \le n$ ансамбля Y:

$$\mathbf{P}\ X,Y\ = \begin{pmatrix} P\ x_1,y_1 & \dots & P\ x_1,y_j & \dots & P\ x_1,y_n \\ \dots & \dots & \dots & \dots & \dots \\ P\ x_i,y_1 & \dots & P\ x_i,y_j & \dots & P\ x_i,y_n \\ \dots & \dots & \dots & \dots & \dots \\ P\ x_m,y_1 & \dots & P\ x_m,y_j & \dots & P\ x_m,y_n \end{pmatrix}. \quad (18)$$

Выше была найдена энтропия одной дискретной случайной величины (или события) ξ, равная

$$H\ X\ = -\sum_{i=1}^{m} P\left(x_i\right) \log P\left(x_i\right), \quad (19)$$

где m - число состояний ξ.

Допустим, что заданы две дискретные случайные величины ξ и η. Величина ξ может принимать любое из множества значений $X = \{x_1, x_2, \dots, x_m\}$ с соответствующими вероятностями $P\left(x_1\right), P\left(x_2\right), \dots, P\ x_m$, а величина η — любое из множества значений $Y = \{y_1, y_2, \dots, y_n\}$ с вероятностями $P\left(y_1\right), P\left(y_2\right), \dots, P\ y_n$, причем не обязательно $m = n$. Эти вероятности в силу условия согласованности могут быть получены из элементов матрицы $\mathbf{P}\ X,Y$:

$$P\left(x_i\right)=\sum_{j=1}^{n}P\left(x_i,y_j\right),\quad P\;\; y_j\;\; =\sum_{i=1}^{m}P\left(x_i,y_j\right).$$

Совокупность случайных величин ξ и η представляет собой кортеж (упорядоченную двойку) $\varsigma=\;\;\xi,\eta\;$. Он определен на множестве $Z=X\times Y$, которое является декартовым произведением указанных выше множеств. Элементы Z множества двумерны и описываются кортежами типа $z_{ij}=\left(x_i,y_j\right)$. Случайная величина ς двумерна, и ей соответствуют двумерные совместные вероятности $P\left(x_i,y_j\right)$.

Выше указывалось, что энтропия не зависит от того, что представляет собой физически случайный объект, неопределенность которого она оценивает, а определяется только совокупностью вероятностей его состояний. Поэтому в соответствии с формулой (19) можно определить *совместную энтропию* совокупности величин ξ_i и η_j

$$\begin{gathered}H\;\;Z\;\;=H\;\;X,Y\;\;=-\sum_{i=1}^{m}\sum_{j=1}^{n}P\left(z_{ij}\right)\log P\left(z_{ij}\right)=\\=-\sum_{i=1}^{m}\sum_{j=1}^{n}P\left(x_i,y_j\right)\log P\left(x_i,y_j\right).\end{gathered}\qquad(20)$$

Эту энтропию обычно обозначают $H\;\;X,Y\;$.

Преобразуем полученное выражение с целью выяснения его связи с энтропией $H\;\;X\;$ одной из составляющих пары рассматриваемых величин. В соответствии с теоремой умножения вероятностей имеем

$$P\left(x_i, y_j\right) = P\left(x_i\right) P\left(y_j / x_i\right), \qquad (21)$$

где $P\left(y_j / x_i\right)$ - условная вероятность исхода y_j при условии, что реализовался исход x_i. Подставим (21) в (20) и выполним простые преобразования, учитывая, что логарифм произведения чисел равен сумме их логарифмов:

$$\begin{aligned} H\ \ X,Y\ \ &= -\sum_{i=1}^{m}\sum_{j=1}^{n} P\left(x_i\right) P\left(y_j / x_i\right) \log\left[P\left(x_i\right) P\left(y_j / x_i\right)\right] = \\ &= -\sum_{i=1}^{m} P\left(x_i\right) \log P\left(x_i\right) \cdot \sum_{j=1}^{n} P\left(y_j / x_i\right) - \\ &- \sum_{i=1}^{m} P\left(x_i\right) \sum_{j=1}^{n} P\left(y_j / x_i\right) \log P\left(y_j / x_i\right). \end{aligned}$$

Учитывая, что для всех x_i справедливо условие нормировки

$$\sum_{j=1}^{n} P\left(y_j / x_i\right) = 1$$

(т.е. при любом x_i обязательно реализуется хотя бы одно y_j), последнюю формулу преобразуем следующим образом:

$$\begin{aligned} &H\ \ X,Y\ \ = \\ &= -\sum_{i=1}^{m} P\left(x_i\right) \log P\left(x_i\right) - \sum_{i=1}^{m} P\left(x_i\right) \sum_{j=1}^{n} P\left(y_j / x_i\right) \log P\ \ y_j / x_i\ \ = \qquad (22) \\ &= H\ \ X\ \ + \sum_{i=1}^{m} H\left(Y / x_i\right) P\left(x_i\right), \end{aligned}$$

где $H\ \ X$ - энтропия случайной величины ξ_i. В полученное выражение входит так называемая *частная условная энтропия*

$$H\left(Y / x_i\right) = -\sum_{j=1}^{n} P\left(y_j / x_i\right) \log P\left(y_j / x_i\right). \qquad (23)$$

Это — энтропия (т.е. неопределенность) случайной величины η при условии, что случайная величина ξ приняла конкретное значение x_i.

Второй член в последнем выражении (22) есть результат осреднения $H\left(Y / x_i\right)$ по всем значениям $X = \{x_1, x_2, \ldots, x_m\}$ величины ξ и называется *средней условной энтропией* величины η при условии реализации *любой* из величин ξ. Обозначив

$$H(Y / X) = \sum_{i=1}^{m} H\left(Y / x_i\right) P\left(x_i\right), \qquad (24)$$

окончательно получим

$$H(X, Y) = H(X) + H(Y / X). \qquad (25)$$

Основной смысл средней условной энтропии $H(Y / X)$ состоит в том, что она показывает, какую энтропию имеет случайная величина η, когда уже снята энтропия (неопределенность!) случайной величины ξ. Иными словами, она дает меру неопределенности события η, когда о событии ξ все известно.

Принимая во внимание теорему умножения вероятностей (21), формулу (24) с учетом (23) можно также преобразовать к виду

$$H(Y / X) = -\sum_{i=1}^{m} \sum_{j=1}^{n} P\left(x_i, y_j\right) \log P\left(y_j / x_i\right) \qquad (26)$$

Из очевидного равенства $H(X,Y)=H(Y,X)$ получим, кроме (25), еще одно выражение для совместной энтропии

$$H(X,Y)=H(Y)+H(X/Y), \tag{27}$$

где средняя условная энтропия

$$H(X/Y)=-\sum_{i=1}^{m}\sum_{j=1}^{n}P\left(x_i,y_j\right)\log P\left(x_i/y_j\right). \tag{28}$$

Таким образом, совместная энтропия двух случайных величин равна сумме безусловной энтропии одной из них и условной энтропии — второй величины.

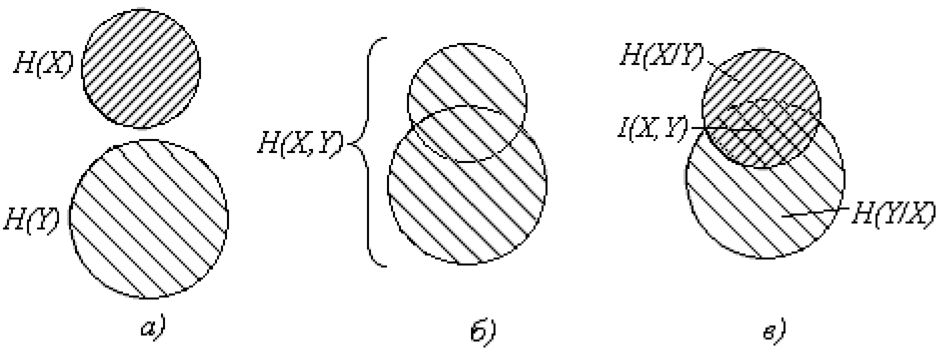


Рис. 4. - Графическое изображение соотношений между энтропиями

Уяснению соотношений между рассмотренными энтропиями дискретных источников информации способствует их графическое изображение в виде диаграмм, похожих на диаграммы Эйлера-Венна (*Рис. 4*). В позиции *а)* представлены два круга. Круг $H(X)$, характеризует энтропию случайной величины ξ, т.е. степень неопределенности знания о ее состоянии. Аналогично, круг $H(Y)$ характеризует энтропию случайной величины η. Из сравнения размеров этих кругов видно, что неопределенность величины η больше, чем величины ξ. Так как эти круги не пересекаются, то в сведениях о величине η отсутствуют сведения о величине ξ и наоборот. В позиции *б)* представлен тот случай, когда в совместной энтропии две неопределенности имеют общую часть. В позиции *в)* указано, что представляют собой отдельные части диаграммы совместной энтропии.

Отметим следующие основные свойства совместной энтропии двух случайных величин.

1. При статистической независимости величин ξ и η совместная энтропия равна сумме энтропий этих величин (*Рис.4а)*:

$$H(X,Y) = H(X) + H(Y). \qquad (29)$$

Действительно, при статистически независимых величинах условная вероятность равна безусловной вероятности: $P\left(y_j / x_i\right) = P(y_j)$, а совместная вероятность равна произведению вероятностей: $P\left(x_i, y_j\right) = P\left(x_i\right) P(y_j)$. Тогда условную энтропию получим из формулы (26):

$$H(Y/X) = -\sum_{i=1}^{m} \sum_{j=1}^{n} P\left(x_i\right) P(y_j) \log P(y_j).$$

Произведя суммирование по i с учетом условия нормировки $\sum_{i=1}^{m} P(x_i) = 1$, найдем

$$H(Y/X) = -\sum_{j=1}^{n} P(y_j) \log P(y_j) = H(Y).$$

Следовательно, при статистически независимых величинах условная энтропия равна безусловной, и формула (25) превращается в (29). Этот же вывод можно сделать, сравнивая диаграммы позиций *а)* и *в) Рис. 4*.

2. При детерминированной взаимнооднозначной зависимости величин ξ и η совместная энтропия равна безусловной энтропии одной из величин.

Очевидно, что при взаимнооднозначной зависимости условные вероятности $P\left(y_j / x_i\right)$ и $P\left(x_i / y_j\right)$ равны нулю или единице. При этом имеют место равенства $P\left(y_j / x_i\right)\log P\left(y_j / x_i\right) = 0$ и $P\left(x_i / y_j\right)\log P\left(x_i / y_j\right) = 0$. Поэтому в силу формул (26) и (28) условные энтропии $H\ Y / X\ = H\ X / Y\ = 0$, и $H\ X,Y\ = H\ X\ = H\ Y$. Применительно к *Рис.4* это означает, что такая ситуация возможна только в том случае, когда круги $H\ X$ и $H\ Y$ имеют одинаковую величину и совмещены друг с другом.

3 Условная энтропия может изменяться в пределах

$$0 \le H\ Y / X\ \le H\ Y\ .$$

Выше показано, что условная энтропия положительна, равна нулю при взаимнооднозначной зависимости событий, максимальна при полной статистической независимости событий и равна при этом безусловной энтропии. Отсюда непосредственно вытекает это свойство.

4. Для совместной энтропии всегда справедливо соотношение

$$H\ X,Y\ \le H\ X\ + H\ Y\ .$$

Это свойство совместной энтропии является непосредственным следствием предыдущего соотношения.

6 Взаимная информация

Для того, чтобы дать определение количества информации, содержащейся в какой-либо случайной величине или в событии о другой случайной величине или событии, рассмотрим некоторый дискретный канал связи (*Рис.5*). На вход канала подается символ или дискретная случайная величина (или символ) ξ, принимающая значения из множества $X = \{x_1, x_2, ..., x_N\}$, а на его выходе образуется также дискретная случайная величина η, принимающая значения из множества $Y = \{y_1, y_2, ..., y_N\}$. Будем считать, что количества элементов в этих множествах одинаковы. В канале связи обычно присутствуют шумы. Поэтому при поступлении на его вход величины x_i (что можно также трактовать как осуществление некоторого события x_i), на его выходе не обязательно будет зарегистрирована величина y_i (т.е. событие y_i), соответствующая именно данному x_i. Отличие принятой величины от переданной характеризует ошибки канала связи.

Пусть величина x_i имеет априорную вероятность $P\left(x_i\right)$. Эта вероятность определяется только статистической структурой входного сообщения и не зависит ни от строения канала связи, ни от статистических свойств шумов в канале. Очевидно, что если бы канал связи не имел помех, то две последовательности $\{y_i\}$ и $\{x_i\}$ в точности соответствовали

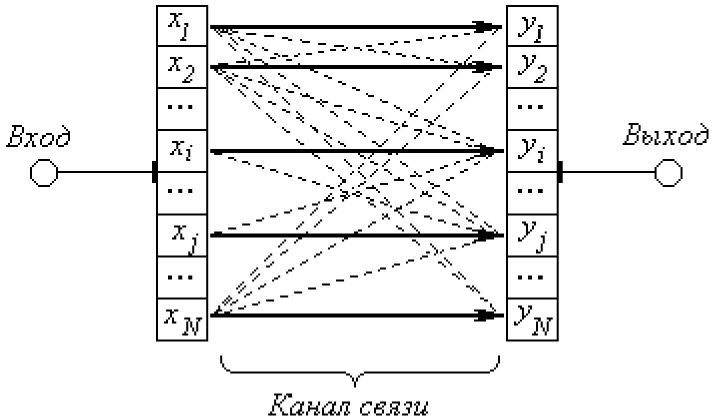


Рис.5. — Прохождение дискретной случайной величины через канал связи

бы друг другу. Это означает, что при передаче величины x_i на выходе канала была бы получена обязательно величина y_i. На *Рис.5* такая ситуация отмечена жирными линиями.

Однако наличие шумов ведет к тому, что при получении величины y_j на выходе канала остается некоторое незнание о том, какая величина x_i на самом деле была передана (пунктирные линии на рисунке). Следовательно, выходной сигнал должен характеризоваться не априорной вероятностью $P\;x_i$, а апостериорной вероятностью $P\left(x_i / y_j\right)$. Эта величина есть вероятность того, что при получении на выходе канала y_j на его вход была подана именно величина x_i. Иначе говоря, помеха приводит к тому, что наблюдение события y_j изменяет вероятность того, что на входе канала осуществилось именно событие x_i, с априорной $P(x_i)$ на апостериорную $P(x_i / y_j)$. Применяя теорему умножения вероятностей, эту вероятность можно записать в виде

$$P(x_i / y_j) = P(x_i, y_j) / P(y_j), \tag{30}$$

где $P(x_i, y_j)$ - совместная двумерная вероятность осуществления событий x_i на входе и y_j на выходе канала.

Введем математическое определение количества *взаимной информации*. Следуя К. Шеннону, определим количество *частной* информации $I(x_i; y_j)$ относительно события x_i, доставляемое событием y_j, как разницу между априорной информацией и апостериорной информацией, т.е. согласно формуле (14),

$$I(x_i; y_j) = -\log P(x_i) - \left(-\log P(x_i / y_j)\right).$$

Это можно представить как логарифм отношения апостериорной вероятности к априорной:

$$I(x_i; y_j) = \log \frac{P(x_i / y_j)}{P(x_i)}. \tag{31}$$

Выбор основания логарифма определяет единицу измерения информации и осуществляется точно так же, как при определении информационной емкости накопителя. Поэтому и название единицы количества взаимной информации такое же, как и информационной емкости. Обычно это биты или производные от них (байты, килобайты, мегабайты и т.д.).

Рассмотрим основные свойства частичной информации, следующие из ее определения, и убедимся, что они отвечают нашим интуитивным представлениям о количестве информации.

1. Информация, содержащаяся в событии y_j о событии x_i, равна информации, содержащейся в событии x_i о событии y_j. Для доказательства этого учтем в выражении (31) формулу (30):

$$I\left(x_i;y_j\right)=\log\frac{P\left(x_i,y_j\right)}{P\left(x_i\right)P\left(y_j\right)}. \tag{32}$$

Отсюда видно, что мера количества информации симметрична относительно x_i и y_j, т.е. $I\left(x_i;y_j\right)=I\left(y_j;x_i\right)$. Поэтому ее можно назвать *взаимной информацией* двух случайных событий относительно друг друга.

Примером могут здесь служить две различных книги рассказов разных авторов. Если в них обеих есть определенный рассказ одного и того же автора, то ясно, что сколько информации имеется в первой книге о содержании второй, то столько же ее имеется в содержании второй книги о первой.

2. Структура правой части формулы (32) характеризует меру статистической связи между событиями x_i и y_j,. Так, если события x_i и y_j статистически независимы, то $I\left(x_i;y_j\right)=0$. При независимых событиях результат наблюдения или измерения y_j не увеличивает знаний относительно x_i.

В качестве примера можно рассмотреть указанный выше канал связи. Очевидно, что если появление символов на его входе и на его выходе не связано друг с другом, то никакой информации на выходе о входных символах получено быть не может.

3. Из формулы (31) следует, что при фиксированной вероятности $P\left(x_i\right)$ взаимная информация $I\left(x_i;y_i\right)$ будет наибольшей, если $P\left(x_i/y_i\right)=1$, т.е. когда отсутствуют ошибки перехода x_i в y_i, и y_i достоверно и однозначно определяет x_i (жирные стрелки *Рис.5*). Это максимальное значение совпадает с (14), равно

$$I\left(x_i\right)=-\log P\left(x_i\right) \tag{33}$$

и называется *собственной информацией события* x_i. Оно представляет количество информации, доставляемое самим событием x_i или любым другим, однозначно связанным с x_i. Это количество тем больше, чем меньше вероятность появления события x_i.

Так как в общем случае $P\ x_i/y_j\ \leq 1$, то из формулы (31) видно, что само событие несет о себе большее (по крайней мере, не меньшее) количество информации, чем какое-либо другое, статистически с ним связанное. Следовательно, шумы в канале связи могут *только уменьшить информативность* выходного сигнала. Поэтому взаимная информация не превышает значения собственной информации каждого из событий, т.е.

$$I\left(x_i;y_j\right)\leq I\left(x_i\right),\ I\left(x_i;y_j\right)\leq I\left(y_j\right).$$

В силу условия $0\leq P\ x_i\ \leq 1$ собственная информация всегда неотрицательна $I\ x_i\ \geq 0$. Взаимная же информация может быть как положительной, так и отрицательной. Она положительна, если $P(x_i / y_j) > P(x_i)$, и отрицательна (дезинформация) в противоположном случае.

4. Предположим, что величина ξ может с равной вероятностью принимать N возможных состояний. Очевидно, вероятность появления любого из ее состояний x_i есть $P(x_i) = 1/N$. Тогда собственная информация события x_i в соответствии с формулой (33) равна $I(x_i) = \log N$. Но правая часть этой формулы равна правой части выражения (4). Следовательно, собственная информация величины с равномерно распределенными состояниями совпадает с информационной емкостью накопителя, необходимой для ее регистрации и хранения: $I(x_i) = C$. Если состояния не равновероятны, эти величины не равны.

7 Средняя взаимная информация

Найдем среднее количество информации о случайной величине η относительно случайной величины ξ: $I(X,Y) = \mathrm{M}\left[I(x_i,y_j)\right]$. Осредняя входящее сюда выражение в виде (31) и (32), получим для него две эквивалентные формы:

$$I(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i,y_j) \log \frac{P(x_i / y_j)}{P(x_i)} \tag{34}$$

и

$$I(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i,y_j) \log \frac{P(x_i,y_j)}{P(x_i)P(y_j)}. \tag{35}$$

На практике наибольший интерес представляет именно среднее количество взаимной информации $I(X,Y)$, а не количество информации $I(x_i,y_j)$ в некотором определенном состоянии y_j величины η относительно определенного состояния x_i величины ξ.

Формулу (34) можно привести к виду

$$I(X,Y)=\sum_{i=1}^{m}\sum_{j=1}^{n}P\left(x_i,y_j\right)\log P\left(x_i/y_j\right)-\sum_{i=1}^{m}\sum_{j=1}^{n}P\left(x_i,y_j\right)\log P\left(x_i\right).$$

Согласно формуле (26), первое слагаемое ее правой части представляет собой условную энтропию, взятую с обратным знаком, т.е. $-H(X/Y)$. Второе слагаемое преобразуем с учетом равенства $\sum_{j=1}^{n}P(x_i,y_j)=P(x_i)$. Получим

$$\sum_{i=1}^{m}\sum_{j=1}^{n}P\left(x_i,y_j\right)\log P\left(x_i\right)=\sum_{i=1}^{m}P\left(x_i\right)\log P\left(x_i\right)=-H(X).$$

Поэтому имеем

$$I(X,Y)=H(X)-H(X/Y). \tag{36}$$

Таким образом, применительно к рассматривавшейся ранее схеме канала связи среднее количество информации, получаемое при неполной достоверности сообщений, равно разности безусловной энтропии $H(X)$, характеризующей начальную (априорную) неопределенность сообщений на входе канала, и условной энтропии $H(X/Y)$, характеризующей остаточную (апостериорную) неопределенность сообщений на выходе канала. Графически средняя взаимная информация $I(X,Y)$ на *Рис.4* изображена областью в позиции *в)*, заштрихованной в обе стороны.

Из равенства (27) имеем $H(X/Y) = H(X,Y) - H(Y)$. Поэтому выражение (36) можно привести к виду

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{37}$$

Следовательно, количество передаваемой информации может быть выражено через сумму энтропий передаваемого ξ и принимаемого η сообщений за вычетом их совместной энтропии $H(X,Y)$.

Рассмотрим некоторые свойства взаимной информации.

1. Так как $H(X,Y) = H(Y,X)$, то из последнего равенства следует

$$I(Y;X) = I(X;Y),$$

т.е. количество информации, которое содержится в сообщении Y относительно сообщения X, равно количеству информации, содержащемуся в X относительно Y. Поэтому $I(Y,X)$ и $I(X,Y)$ называют также *полной взаимной информацией*.

2. При статистической независимости X и Y взаимная информация равна нулю: $I(X,Y) = 0$. Она не может быть отрицательной: $I(X,Y) \geq 0$ Действительно, при статистической независимости справедливо равенство (29): $H(X,Y) = H(X) + H(Y)$, что в соответствии с формулой (37) дает указанный результат.

3. При полной статистической зависимости (детерминированной связи) X и Y взаимная информация

$$I(X;Y) = H(X) = H(Y). \tag{38}$$

При указанных условиях выполняется соотношение $H(X,Y) = H(X) = H(Y)$, подстановка которого в формулу (37) дает результат (38).

Пример 5. Взаимная информация между парой соседних букв русскоязычного текста составляет 0,83 *бит*, англоязычного текста 0,71 *бит*. □

8 Энтропия непрерывных сообщений (непрерывных случайных величин)

Выше была рассмотрена мера неопределенности выбора для дискретных случайных величин. На практике очень часто встречаются источники информации, множество возможных состояний которых составляет континуум. Такие источники называются *непрерывными* источниками информации.

Во многих случаях непрерывные источники преобразуются в дискретные посредством применения устройств дискретизации и квантования. Вместе с тем существует немало и таких систем, в которых информация передается и преобразуется непосредственно в форме непрерывных сигналов. Примерами могут служить системы проводной телефонной и радиосвязи.

Оценка неопределенности выбора для непрерывного источника информации имеет определенную специфику. *Во-первых*, значения, реализуемые источником, математически отображаются *непрерывной* случайной величиной. *Во-вторых*, вероятности значений этой случайной величины не могут использоваться для оценки неопределенности, так как в данном случае вероятность любого конкретного значения равна нулю.

Поэтому естественно связывать неопределенность выбора значений непрерывной случайной величины с плотностью распределения вероятностей этих значений. Учитывая, что для совокупности значений, относящихся к сколь угодно малому интервалу непрерывной случайной величины, вероятность конечна, найдем формулу для энтропии непрерывного источника информации, используя операции квантования и последующего предельного перехода при уменьшении кванта до нуля.

Для этого разобьем диапазон изменения непрерывной случайной величины ξ, описываемой плотностью вероятностей $p(x)$, на конечное число n интервалов шириной Δx (*Р.ис.6*). При реализации любого значения *непрерывной* случайной величины ξ, принадлежащей интервалу $\left(x_i - \Delta x/2, x_i + \Delta x/2\right)$, будем считать, что реализовалось значение x_i *дискретной* случайной величины $\tilde{\xi}$. Так как Δx мало, вероятность этого есть

$$P\left(x_i\right) = P\left(x_i - \Delta x/2 \le \xi < x_i + \Delta x/2\right) = \int_{x_i-\Delta x/2}^{x_i+\Delta x/2} p(x)\,dx \approx p\left(x_i\right)\Delta x. \qquad (39)$$

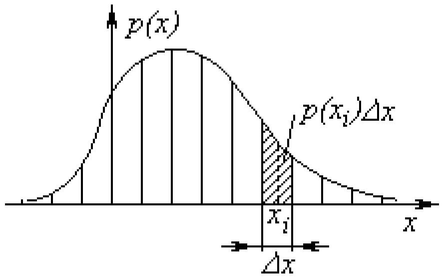


Рис.6. К определению энтропии непрерывных сообщений

Тогда энтропия дискретной случайной величины $\tilde{\xi}$ может быть записана в виде

$$H\left(\tilde{X}\right) = -\sum_{i=1}^{n} P\left(x_i\right)\log P\left(x_i\right) = -\sum_{i=1}^{n} p\left(x_i\right)\Delta x \log\left[p\left(x_i\right)\Delta x\right]$$

или

$$H(\tilde{X}) = -\sum_{i=1}^{n} p\left(x_i\right)\Delta x \log p\left(x_i\right) - \sum_{i=1}^{n} p\left(x_i\right)\Delta x \log(\Delta x).$$

Так как $\sum_{i=1}^{n} p(x_i)\Delta x = 1$, то отсюда имеем

$$H\left(\tilde{X}\right) = -\sum_{i=1}^{n} p\left(x_i\right)\Delta x \log p\left(x_i\right) - \log \Delta x .$$

По мере уменьшения Δx приближенные равенства (39) становятся все точнее, и свойства дискретной случайной величины $\tilde{\xi}$ все больше приближаются к свойствам непрерывной случайной величины ξ. Переходя к пределу при $\Delta x \to 0$ $n \to \infty$, получаем следующее выражение для энтропии $H(X)$ непрерывной случайной величины:

$$H(X) = \lim_{\Delta x \to 0} H\left(\tilde{X}\right) = \lim_{\Delta x \to 0}\left\{-\sum_{i=1}^{n}\left[p\left(x_i\right)\log p\left(x_i\right)\right]\Delta x\right\} - \lim_{\Delta x \to 0}\log \Delta x$$

или

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx - \lim_{\Delta x \to 0}\log \Delta x . \qquad (40)$$

Эта величина при $\Delta x \to 0$ стремится к бесконечности, что полностью соответствует интуитивному представлению о том, что неопределенность выбора из бесконечно большого числа возможных состояний (значений) бесконечно велика.

Первый член в правой части соотношения (40) имеет конечное значение, которое зависит только от закона распределения непрерывной случайной величины ξ и не зависит от шага квантования Δx. Он имеет такую же структуру, как энтропия дискретного источника.

Второй член того же соотношения, наоборот, зависит лишь от шага квантования случайной величины ξ. Именно в нем кроется причина того, что величина $H\ X$ обращается в бесконечность. Однако ясно, что непрерывный сигнал всегда воспринимается приближенно, с ограниченной точностью. Поэтому в последнем выражении величине Δx вполне логично придать смысл некоторого интервала неопределенности, которым, например, характеризуется система, воспринимающая сигнал, или же шага квантования возможных значений сигнала, т.е. так или иначе можно считать величину Δx конечной.

Следует отметить, что при использовании энтропии в теории информации непрерывных сообщений одно и то же бесконечное слагаемое обязательно фигурирует во всяком выражении дважды – один раз со знаком (+) и второй – со знаком (-); таким образом, в итоге это слагаемое всегда уничтожается. Поэтому здесь можно определить энтропию просто как

$$H\ X = -\int\limits_{-\infty}^{\infty} p\ x\ \log p\ x\ dx. \tag{41}$$

Поскольку для определения этой величины используется только плотность распределения вероятностей, т.е. дифференциальный закон распределения, она получила название *дифференциальной энтропии* непрерывного источника (непрерывного распределения случайной величины ξ).

Итак, формула (41) получена из (40) устранением слагаемого $-\lim\limits_{\Delta x \to 0} \log\ \Delta x$. Такая процедура ведет к тому, что, например, в формуле (41) аргументом логарифма является величина $p(x)$, имеющая размерность, обратную размерности ξ, что недопустимо с математической точки зрения. Устранить этот недостаток можно, введя в аргумент логарифма сомножитель с размерностью ξ [11]. Обозначим этот сомножитель d_0. Тогда формула (41) приобретает вид

$$H(X) = -\int\limits_{-\infty}^{\infty} p(x)\log[d_0 \cdot p(x)]\,dx.$$

При $d_0 = 1$ с размерностью, в которой измеряется величина ξ, получаем формулу (41).

Аналогично изложенному, используя операции квантования и предельного перехода, найдем выражение для *условной энтропии* непрерывного источника информации:

$$H\left(Y/X\right) = -\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p\left(x,y\right)\log\frac{p\left(x,y\right)}{p_X\left(x\right)}dx\,dy - \lim_{\Delta x\to 0}\log\left(\Delta x\right). \qquad (42)$$

Отметим, что второй член в правой части этого выражения идентичен соответствующему члену в соотношении (40) для энтропии непрерывной случайной величины. Обычно под условной энтропией непрерывного источника информации понимают первый член правой части выражения (42), т.е.

$$H\left(Y/X\right) = -\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p\left(x,y\right)\log\frac{p\left(x,y\right)}{p_X\left(x\right)}dx\,dy.$$

Эта величина конечна и называется *дифференциальной условной энтропией* непрерывного источника. Она характеризует неопределенность выбора непрерывной случайной величины η при условии, что известны результаты реализации значений другой статистически связанной с ней непрерывной случайной величины ξ.

Отметим некоторые свойства дифференциальной энтропии.

1. Дифференциальная энтропия в отличие от энтропии дискретного источника является *относительной* мерой неопределенности. Ее значение зависит от масштаба случайной величины ξ, а, следовательно, и от выбора единицы изме-

рения. Изменим масштаб случайной величины ξ в k раз. Если $u=kx$, то $p(u)=p(x)/k$. Действительно, это следует из соотношения

$$\int_{-\infty}^{\infty} p(x)\,dx=\int_{-\infty}^{\infty}\frac{p(x)}{k}d(kx)=\int_{-\infty}^{\infty}p(u)\,du=1.$$

Тогда

$$H(U)=-\int_{-\infty}^{\infty}p(u)\log p(u)\,du=$$
$$=-\int_{-\infty}^{\infty}\frac{p(u)}{k}\log\frac{p(u)}{k}k\,dx=-\int_{-\infty}^{\infty}p(x)\log p(x)\,dx+\log k.$$

Из относительности дифференциальной энтропии следует, что эта энтропия может принимать положительные, отрицательные и нулевые значения.

2. Дифференциальная энтропия не зависит от изменения всех значений случайной величины ξ на постоянное число. Действительно, масштаб ξ при этом не меняется и справедливо равенство

$$H(X+x_0)=-\int_{-\infty}^{\infty}p(x+x_0)\log p(x+x_0)\,d(x+x_0)=-\int_{-\infty}^{\infty}p(x)\log p(x)\,dx=H(X).$$

3. Соотношения для дифференциальной энтропии объединения статистически зависимых непрерывных источников аналогичны соответствующим формулам для дискретных источников:

$$H(X,Y)=H(X)+H(Y/X)=H(Y)+H(X/Y),$$

где

$$H(X,Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x,y)\,dx\,dy.$$

Справедливость этого соотношения легко доказывается применением указанных выше формул.

Так как

$$H(X/Y)\le H(X)\ \text{ и }\ H(Y/X)\le H(Y),$$

то $H(X,Y) \le H(X) + H(Y)$, причем равенство имеет место только в случае отсутствия связи между ξ и η.

9 Средняя взаимная информация непрерывных сообщений (непрерывных случайных величин)

Ранее было показано (формула (35)), что средняя взаимная информация двух *дискретных* случайных величин $\tilde{\xi}$ и $\tilde{\eta}$ дается формулой

$$I(\tilde{X},\tilde{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i,y_j)\log\frac{P(x_i,y_j)}{P(x_i)P(y_j)}. \qquad (43)$$

Теперь допустим, что случайные величины ξ и η *непрерывны*. Разобьем диапазон изменения величины ξ на m одинаковых малых интервалов шириной Δx, а величины η - на n одинаковых малых интервалов шириной Δy. При реализации любого значения ξ из интервала $(x_i - \Delta x/2,\ x_i + \Delta x/2)$ и любого значения η, принадлежащего интервалу $(y_j - \Delta y/2, y_j + \Delta y/2)$, будем считать, что реализовались значение x_i *дискретной* случайной величины $\tilde{\xi}$ и значение y_j *дискретной* случайной величины $\tilde{\eta}$. При этом

$$P(x_i) \approx p_X(x_i)\Delta x,\ P(y_j) \approx p_Y(y_j)\Delta y, P(x_i,y_j) \approx p(x_i,y_j)\Delta x\Delta y,$$

где $p_X(x_i)$ и $p_Y(y_j)$ - одномерные плотности вероятности величин ξ и η, а $p(x_i,y_j)$ - их двумерная плотность. Поэтому средняя взаимная информация (43) дискретных величин $\tilde{\xi}$ и $\tilde{\eta}$ может быть записана как

$$I(\tilde{X};\tilde{Y}) \approx \sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log\frac{p(x_i,y_j)}{p_X(x_i)p_Y(y_j)}\Delta x\Delta y.$$

Переходя к пределу при $\Delta x \to 0, \Delta y \to 0$ $m \to \infty, n \to \infty$, получаем следующее выражение для *средней взаимной информации непрерывных случайных величин* ξ и η:

$$I\ X;Y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p\ x,y\ \log\frac{p\ x,y}{p_X\ x\ p_Y\ y}\,dx\,dy. \quad (44)$$

Оно является *точным* и *не имеет слагаемого*, зависящего от малых Δx и Δy, характерного для энтропии непрерывного источника.

Обратим внимание на следующее обстоятельство. Для нахождения величины средней взаимной информации дискретных случайных величин необходимо знать двумерное распределение их вероятностей $P\ x_i, y_j$, а непрерывных случайных величин — двумерных плотностей вероятностей $p\ x,y$. Теоретические модели многомерных дискретных распределений в литературных источниках отсутствуют, а многомерные непрерывные модели представлены гауссовским (нормальным) распределением. Поэтому представляется естественным теоретическое исследование взаимной информации непрерывных случайных величин на примерах нормального распределения, а дискретных случайных величин — на экспериментальных данных, в частности, на цифровых фотоизображениях. Поэтому вначале рассмотрим достаточно подробно многомерное нормальное распределение. Начнем с одномерного.

10 Нормальное распределение

В настоящее время в теоретических исследованиях и практической деятельности используется несколько десятков одномерных распределений непрерывных случайных величин, описываемых аналитически [9]. Это распределения: равномерное, треугольное (Симпсона), экспоненциальное, гиперэкспоненциальное, Коши, Лапласа, гамма- , бета - распределения, и др.

Более подробно остановимся на нормальном (гауссовом) распределении, аналитическая запись плотности вероятности которого имеет вид (*Рис.7*)

$$p\left(x\right)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{\left(x-M_1\right)^2}{2\sigma^2}\right],\quad -\infty<x<\infty,\ -\infty<M_1<\infty,\ \sigma>0\ , \qquad (51)$$

где M_1 и σ - соответственно параметры, характеризующие центр распределения (математическое ожидание случайной величины ξ) и его масштаб (среднее квадратическое отклонение этой величины). Дисперсия нормального распределения равна σ^2. Это распределение симметрично относительно математического ожидания.

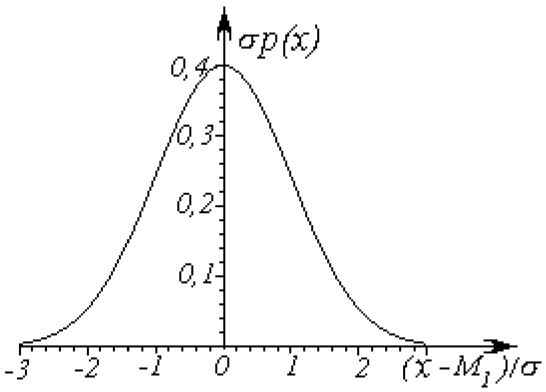


Рис.7— Плотность вероятности нормального распределения

Нормальное распределение на практике и в теоретических расчетах используется наиболее часто. Долгое время его важность переоценивалась вследствие ошибочного представления о том, будто оно является основным распределением в природе и что, согласно теории ошибок, нормальному закону подчиняются все измерения. В начале XX в. с помощью статистического анализа испытаний было показано, что это допущение справедливо не во всех случаях.

Теоретическим обоснованием роли нормального распределения является центральная предельная теорема — один из наиболее важных результатов математической статистики. Согласно этой теореме, распределение среднего n независимых случайных величин, распределенных по любому закону или даже имеющих до n различных распределений с конечными математическим ожиданием и дисперсией, при увеличении числа наблюдений в выборке (т. е. когда n стремится к бесконечности) приближается к нормальному. Важно, что этот результат справедлив независимо от того, по какому закону распределена каждая из n случайных величин, среднее которых рассматривается.

Хотя центральная предельная теорема связана с большими выборками, распределение выборочного среднего стремится к нормальному даже при относительно небольших значениях n, если значение дисперсии какого-либо элемента или небольшой группы элементов не является преобладающим и распределение элементов не слишком отклоняется от нормального.

Когда случайная величина представляет собой общий результат большого числа независимых «небольших» воздействий, то, согласно центральной предельной теореме, можно ожидать, что эта случайная величина будет распределена по нормальному закону. Кроме того, эмпирические результаты свидетельствуют о том, что нормальный закон удовлетворительно описывает многие реальные случайные величины. Примерами могут служить измерения, проводимые на живых организмах, компоненты скорости движения молекул в газе, число баллов при проверке умственных способностей, средняя температура воздуха в данном районе, уровень случайных шумов в электрической цепи. Инструментальные ошибки также часто имеют нормальное распределение относительно истинного значения либо относительно некоторой средней систематической ошибки. Во многих задачах нормальное распределение имеет то преимущество, что его легко задать математически и результаты решения этих задач можно представить в виде аналитических выражений. Ряд методов статистического вывода разработан при допущении, что рассматриваемые данные имеют нормальное распределение. Эти его свойства оказываются полезными и для рассматриваемых здесь задач информационной оптимизации.

Вследствие широкой распространенности нормального распределения и, возможно, благодаря его названию иногда полагают, что случайная величина имеет нормальное распределение, если не доказано обратное утверждение. Следует ясно представлять себе, что многие случайные величины *нельзя* рассматривать как результат воздействия большого числа мелких факторов и, следовательно, нет теоретических оснований ожидать появления нормального распределения. Возможно такое положение, когда будет преобладать какой-либо один эффект, не подчиняющийся нормальному закону.

Теоретически область изменения нормально распределенной случайной величины лежит в интервале от $-\infty$ до $+\infty$. Однако большинство реальных величин имеют верхний или нижний предел, а часто и оба. Тем не менее, это обстоятельство не мешает использовать нормальное распределение для описания такой случайной величины, как рост взрослого человека, где среднее отстоит от нуля на большое число среднеквадратических отклонений и ошибка принятия в качестве пределов $\pm\infty$ пренебрежимо мала. В то же время существуют другие случайные величины, для которых обычным является группирование значений вблизи некоторого физического предела, например, выпуск различных видов продукции в процентах на некотором предприятии или время безотказной работы системы. В таких случаях нормальный закон или какое-либо иное симметричное распределение не подходит. Наконец, следует указать, что для некоторых случайных величин нормальное распределение дает приемлемую аппроксимацию в центре, но не подходит для значений, лежащих в одной или в обеих областях больших отклонений.

Погрешность вследствие ошибочно принятого допущения о нормальности распределения будет различной в каждом конкретном случае. Многие статистические методы, разработанные при этом допущении, остаются справедливыми в случае умеренных отклонений; говорят, что эти методы являются устойчивыми. В то же время, если допущение о нормальности распределения не верно, то могут возникать серьезные ошибки.

На нормальном распределении основано значительное количество иных статистических распределений, В частности, это распределения логарифмически нормальное (логнормальное), Стьюдента (*t* – распределение), Рэлея, Максвелла, хи-квадрат и ряд других.

В дальнейшем изложении, если нормальность важна принципиально для получения некоторого результата, это будет оговариваться особо.

11 Двумерное нормальное распределение

Обобщение одномерного нормального распределения на двумерный случай называется двумерным нормальным распределением и часто используется для описания совместного поведения двух случайных величин. Рассмотрим такое распределение подробнее.

Плотность двумерного распределения центрированных случайных величин ξ_1 и ξ_2 дается выражением

$$p_2\ x_1, x_2\ = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}L^2\right), \quad |\rho|<1\ . \tag{46}$$

В его аргумент входит безразмерная величина, соответствующая равной плотности распределения

$$L^2 = \frac{\left(\frac{x_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1 x_2}{\sigma_1\sigma_2}\right) + \left(\frac{x_2}{\sigma_2}\right)^2}{1-\rho^2}, \tag{47}$$

где ρ - коэффициент корреляции между случайными величинами ξ_1 и ξ_2, а σ_1 и σ_2 — среднеквадратические значения этих величин.

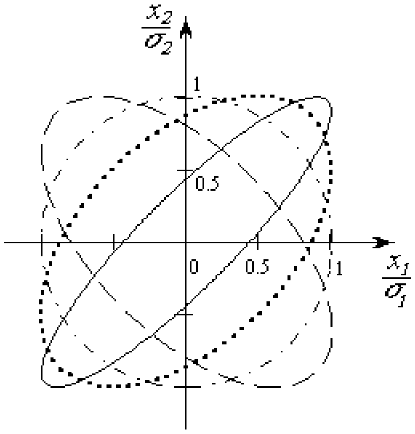


Рис.8 — Семейство эллипсов равной плотности вероятности для $l=1$ и различных ρ:

$- - - \rho=0$, $\cdots\cdots \rho=0.5$, $— - \rho=-0.6$, $—— \rho=0.9$

Кривые (47) для заданных значений L — эллипсы. Семейство таких эллипсов для $L=1$ и разных ρ представлено на *Рис.8* Из него видно, что величина ρ влияет на соотношение осей эллипса и их ориентацию.

При $\rho=0$ формулу (46) можно записать как

$$p_2(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left(\left(\frac{x_1}{\sigma_1}\right)^2+\left(\frac{x_2}{\sigma_2}\right)^2\right)\right),$$

т.е. $p_2(x_1,x_2)=p(x_1)\,p(x_2)$. Здесь $p(x_1)$ и $p(x_2)$ — плотности одномерного нормального распределения. Таким образом, если ξ_1 и ξ_2 – случайные величины, имеющие двумерное нормальное распределение с коэффициентом корреляции $\rho=0$, то плотность их совместного распределения равна произведению плотностей безусловного распределения каждой из этих случайных величин. Как известно, случайные величины ξ_1 и ξ_2 независимы в том и только в том случае, если их двумерная плотность распределения равна произведению одномерных.. Следовательно, для нормального распределения отсутствие корреляции означает независимость случайных величин. В общем случае отсутствие корреляции не означает независимости совместно распределенных случайных величин.

Установим вид связи между нормально распределенными случайными величинами ξ_1 и ξ_2 при стремлении их коэффициента корреляции ρ к ± 1.

Вначале будем считать, что $\rho \to 1$. Положим $\rho = 1-\varepsilon$, где малая величина $\varepsilon > 0$. При этом выражение (47) принимает вид

$$L^2 = \frac{\left(\frac{x_1}{\sigma_1} - \frac{x_2}{\sigma_2}\right)^2}{2-\varepsilon\ \varepsilon} + \frac{2}{2-\varepsilon}\left(\frac{x_1}{\sigma_1}\right)\left(\frac{x_2}{\sigma_2}\right).$$

Устремим ε к нулю (а, значит, ρ к единице). В пределе последнее выражение распадается на два:

$$L^2 = \begin{cases} \left(\frac{x_1}{\sigma_1}\right)^2 = \left(\frac{x_2}{\sigma_2}\right)^2 & при\ \frac{x_2}{\sigma_2} = \frac{x_1}{\sigma_1}, \\ \infty \quad при & \frac{x_2}{\sigma_2} \neq \frac{x_1}{\sigma_1}. \end{cases}$$

Это означает, что при $\rho = 1$ величина ξ_2 линейно функционально связана со случайной величиной ξ_1, а именно $\xi_2 = \frac{\sigma_2}{\sigma_1}\xi_1$. С ростом ξ_1 она также растет, с убыванием – убывает. Зависимость величин ξ_2 и ξ_1 здесь детерминирована, не случайна. Если $0<\rho<1$, то ξ_2 в среднем также растет с увеличением ξ_1, но не функционально, а в той или иной степени случайно, причем степень случайности тем выше, чем меньше ρ.

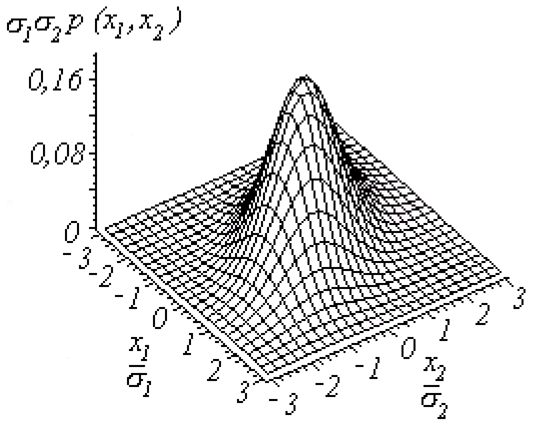


Рис.9 — Двумерное нормальное распределение плотности вероятности

Рассмотрим теперь случай $\rho \to -1$. Положим $\rho = -1 + \varepsilon$ и подставим в (47). Найдем

$$L^2 = \frac{\left(\frac{x_1}{\sigma_1} + \frac{x_2}{\sigma_2}\right)^2}{2-\varepsilon\;\varepsilon} - \frac{2}{2-\varepsilon}\left(\frac{x_1}{\sigma_1}\right)\left(\frac{x_2}{\sigma_2}\right).$$

При стремлении $\rho \to -1$ получаем

$$L^2 = \begin{cases} \left(\frac{x_1}{\sigma_1}\right)^2 = \left(\frac{x_2}{\sigma_2}\right)^2 & \text{\textit{при}}\ \frac{x_2}{\sigma_2} = -\frac{x_1}{\sigma_1}, \\ \infty & \text{\textit{при}}\ \frac{x_2}{\sigma_2} \neq -\frac{x_1}{\sigma_1}. \end{cases}$$

Здесь случайная величина ξ_2 также линейно детерминировано связана со случайной величиной ξ_1, но с ростом ξ_1 она убывает, и, наоборот, с убыванием ξ_1 растет. При $-1 < \rho < 0$ эта тенденция сохраняется, но уже имеет случайный характер.

Пример трехмерного изображения центрированного двумерного нормального распределения для $\rho = -0{,}5$ представлен на *Рис.9*.

На практике двумерное нормальное распределение довольно широко применяется для аппроксимации эмпирических распределений, полученных в процессе тех или иных натурных исследований. Совместное распределение роста и веса людей, распределение координат x и y пробоин в мишени и распределение действительной и мнимой составляющих суммы электрических импедансов - вот несколько примеров практического применения двумерного нормального распределения.

12 Многомерное нормальное распределение

В теоретических исследованиях и практических приложениях нередко приходится иметь дело с совокупностями k случайных непрерывных вещественных величин $\xi_1, \xi_2, \xi_3, ..., \xi_k$. Они могут быть распределены по многомерному нормальному закону (или многомерному закону Гаусса), который является обобщением одномерного и двумерного нормальных распределений, рассмотренных выше. Наиболее компактная запись многомерного распределения достигается при векторно-матричных обозначениях. Для этого необходимо исходную совокупность случайных величин рассматривать как случайный k - мерный вектор-столбец $\boldsymbol{\xi}^T = (\xi_1, \xi_2, ..., \xi_k)$. Знак ^{T}, как обычно, означает транспонирование вектора или матрицы. При этом k - мерная плотность распределения вероятностей такого вектора будет иметь вид [9]:

$$p_k(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k \det A_k}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{M}_1)^T A_k^{-1} (\boldsymbol{x} - \boldsymbol{M}_1)\right]. \qquad (48)$$

Здесь $\boldsymbol{M}_1$ — неслучайный вектор математических ожиданий соответственно величин $\xi_1, \xi_2, ..., \xi_k$, неслучайный вектор $\boldsymbol{x}^T = (x_1, x_2, ..., x_k)$ задает все k аргументов функции $p_k(\boldsymbol{x})$, соответствующих случайному вектору $\boldsymbol{\xi}^T$.

Если случайный вектор $\boldsymbol{\xi}^T$ центрирован, то для него $\boldsymbol{M}_1 = 0$ и формула (48) упрощается:

$$p_k(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k \det A_k}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T A_k^{-1} \boldsymbol{x}\right) \qquad (49)$$

В дальнейшем мы будем считать последнее условие выполненным. Так как энтропия и взаимная информация не зависят от постоянных составляющих исследуемых величин, то это предположение, не уменьшая общности получаемых результатов, позволяет заметно уменьшить объем вычислений.

В формулы (48) и (49) входит квадратная *корреляционная матрица* A_k размера $k \times k$

$$A_k = \begin{bmatrix} \mathbf{M}\left[\xi_1\xi_1\right] & \mathbf{M}\left[\xi_1\xi_2\right] & \mathbf{M}\left[\xi_1\xi_3\right] & \dots & \mathbf{M}\left[\xi_1\xi_k\right] \\ \mathbf{M}\left[\xi_2\xi_1\right] & \mathbf{M}\left[\xi_2\xi_2\right] & \mathbf{M}\left[\xi_2\xi_3\right] & \dots & \mathbf{M}\left[\xi_2\xi_k\right] \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{M}\left[\xi_k\xi_1\right] & \mathbf{M}\left[\xi_k\xi_2\right] & \mathbf{M}\left[\xi_k\xi_3\right] & \dots & \mathbf{M}\left[\xi_k\xi_k\right] \end{bmatrix}. \quad (50)$$

Ее элементы $\mathbf{M}\left[\xi_i \xi_j\right]$ являются корреляционными моментами случайных величин ξ_i и ξ_j (т.е. смешанными центральными моментами второго порядка компонент случайного вектора ξ). Эта матрица должна быть положительно определена. Положительная определенность матрицы означает, что квадратичная форма, построенная на ее основе, неотрицательна, т.е. $\boldsymbol{x}^T A_k^{-1}\boldsymbol{x} > 0$ для любого ненулевого вектора $\boldsymbol{x}$. Если матрица A_k положительно определена, то функция (48) с ростом $|\boldsymbol{x}|$ убывает. Это условие необходимо для обеспечения нормировки плотности вероятности (48), т.е. для выполнения равенства $\int\limits_{-\infty}^{\infty}\dots\int\limits_{-\infty}^{\infty} p_k\ \ \boldsymbol{x}\ \ d\boldsymbol{x} = 1$.

Матрица A_k^{-1}, *обратная* матрице A_k, называется *матрицей точности.* Так как A_k положительно определена, то для нее всегда существует обратная и $\det A_k > 0$.

Для того, чтобы матрица A_k была положительно определена, необходимо и достаточно, чтобы определители всех ее квадратных подматриц порядка $1 \le l \le k$, построенных на главной диагонали (т.е. главные миноры) были положительны [16].

Если матрица A_k диагональна и все ее диагональные элементы положительны, то k - мерная плотность распределения вероятностей (49) распадается на k одномерных нормальных плотностей. Следовательно, такой матрице соответствует совокупность k взаимно независимых гауссовских случайных величин $\xi_1, \xi_2, \dots, \xi_k$.

Убедимся, что из формулы (48) в одномерном случае следует формула (45), а из (49) в двумерном случае – формула (46) с учетом (47).

Положим $k=1$. При этом $\boldsymbol{\xi}=\xi$, $\boldsymbol{x}^T=\boldsymbol{x}=x$, $\boldsymbol{M}_1=M_1$, $A_1=\mathbf{M}\left[\xi^2\right]=\sigma^2$, $A_1^{-1}=1/\sigma^2$, $\det A_1=A_1=\sigma^2$ и распределение (48) обращается в (45).

Далее, при $k=2$ и $\boldsymbol{M}_1=0$ имеем: $\boldsymbol{\xi}^T=(\xi_1,\xi_2)$, $\boldsymbol{x}^T=(x_1,x_2)$ и

$$A_2=\begin{bmatrix}\mathbf{M}\left[\xi_1\xi_1\right] & \mathbf{M}\left[\xi_1\xi_2\right]\\ \mathbf{M}\left[\xi_2\xi_1\right] & \mathbf{M}\left[\xi_2\xi_2\right]\end{bmatrix}=\begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix},$$

где ρ - коэффициент корреляции между величинами ξ_1 и ξ_2. Определитель этой матрицы $\det A_2=\sigma_1^2\sigma_2^2\left(1-\rho^2\right)$, и обратная ей матрица

$$A_2^{-1}=\begin{bmatrix}\dfrac{1}{\sigma_1^2\left(1-\rho^2\right)} & \dfrac{-\rho}{\sigma_1\sigma_2\left(1-\rho^2\right)}\\ \dfrac{-\rho}{\sigma_1\sigma_2\left(1-\rho^2\right)} & \dfrac{1}{\sigma_2^2\left(1-\rho^2\right)}\end{bmatrix}.$$

Подставляя эти величины в формулу (49), найдем выражение для двумерной плотности вероятностей нормального распределения:

$$p_2\left(x_1,x_2\right)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{\left(x_1/\sigma_1\right)^2-2\rho\left(x_1x_2/\left(\sigma_1\sigma_2\right)\right)+\left(x_2/\sigma_2\right)^2}{2\left(1-\rho^2\right)}\right),$$

что совпадает с комбинацией формул (46) и (47).

Рассмотрим более подробно трехмерную плотность вероятности нормального распределения, что в дальнейшем позволит применить геометрическое толкование получаемых параметров и операций над ними.

При $k=3$ и $\boldsymbol{M}_1=0$ имеем $\boldsymbol{\xi}^T=(\xi_1,\xi_2,\xi_3)$, $\boldsymbol{x}^T=(x_1,x_2,x_3)$ и корреляционную матрицу (50)

$$A_3 = \begin{bmatrix} \mathbf{M}[\xi_1\xi_1] & \mathbf{M}[\xi_1\xi_2] & \mathbf{M}[\xi_1\xi_3] \\ \mathbf{M}[\xi_2\xi_1] & \mathbf{M}[\xi_2\xi_2] & \mathbf{M}[\xi_2\xi_3] \\ \mathbf{M}[\xi_3\xi_1] & \mathbf{M}[\xi_3\xi_2] & \mathbf{M}[\xi_3\xi_3] \end{bmatrix}.$$

Учтем, что входящие сюда корреляционные моменты имеют вид

$$\mathbf{M}[\xi_1\xi_1] = \sigma_1^2,\ \mathbf{M}[\xi_2\xi_2] = \sigma_2^2,\ \mathbf{M}[\xi_3\xi_3] = \sigma_3^2,$$
$$\mathbf{M}[\xi_1\xi_2] = \mathbf{M}[\xi_2\xi_1] = \rho_{12}\sigma_1\sigma_2,\ \mathbf{M}[\xi_1\xi_3] = \mathbf{M}[\xi_3\xi_1] = \rho_{13}\sigma_1\sigma_3,$$
$$\mathbf{M}[\xi_2\xi_3] = \mathbf{M}[\xi_3\xi_2] = \rho_{23}\sigma_2\sigma_3,$$

где $\sigma_1, \sigma_2, \sigma_3$ — среднеквадратические отклонения случайных величин ξ_1, ξ_2, ξ_3, $\sigma_1^2, \sigma_2^2, \sigma_3^2$ — их дисперсии и $\rho_{12}, \rho_{13}, \rho_{23}$ — коэффициенты корреляции между этими величинами. Поэтому корреляционную матрицу можно записать как

$$A_3 = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}.$$

Входящий в формулу (49) определитель этой матрицы имеет вид

$$\det A_3 = \sigma_1^2\sigma_2^2\sigma_3^2\alpha\ ,$$

где обозначена безразмерная величина

$$\alpha = 1 - \rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}.$$

Она в силу требования $\det A_k > 0$ должна быть положительной. Это условие выполняется не при любых ρ_{12}, ρ_{13}, ρ_{23}. В частности, при $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$ имеем $\alpha = 0.500$, а при $\rho_{12} = \rho_{13} = 0.5$, $\rho_{23} = -0.6$ получим $\alpha = -0.160$. Следовательно, трехмерное гауссовское распределение плотности вероятности с первым набором коэффициентов корреляции может существовать, со вторым — не может.

Найдем трехмерную матрицу точности, т.е. обратную к A_3. Она имеет вид

$$A_3^{-1}=\begin{bmatrix} \dfrac{1-\rho_{23}^2}{\sigma_1^2\alpha} & \dfrac{\rho_{13}\rho_{23}-\rho_{12}}{\sigma_1\sigma_2\alpha} & \dfrac{\rho_{12}\rho_{23}-\rho_{13}}{\sigma_1\sigma_3\alpha} \\ \dfrac{\rho_{13}\rho_{23}-\rho_{12}}{\sigma_1\sigma_2\alpha} & \dfrac{1-\rho_{13}^2}{\sigma_2^2\alpha} & \dfrac{\rho_{12}\rho_{13}-\rho_{23}}{\sigma_2\sigma_3\alpha} \\ \dfrac{\rho_{12}\rho_{23}-\rho_{13}}{\sigma_1\sigma_3\alpha} & \dfrac{\rho_{12}\rho_{13}-\rho_{23}}{\sigma_2\sigma_3\alpha} & \dfrac{1-\rho_{12}^2}{\sigma_3^2\alpha} \end{bmatrix}.$$

Найдем квадратичную форму, входящую в аргумент функции (49)

$$L^2=\mathbf{x}^T A_3^{-1}\mathbf{x}= x_1,x_2,x_3\ A_3^{-1}\ x_1,x_2,x_3\ ^T.$$

Выполняя соответствующие операции, получим

$$L^2 = 1-\rho_{23}^2\left(\frac{x_1}{\sigma_1}\right)^2 + 1-\rho_{13}^2\left(\frac{x_2}{\sigma_2}\right)^2 + 1-\rho_{12}^2\left(\frac{x_3}{\sigma_3}\right)^2 +$$
$$+2\left[\ \rho_{13}\rho_{23}-\rho_{12}\ \frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2}+\ \rho_{12}\rho_{23}-\rho_{13}\ \frac{x_1}{\sigma_1}\frac{x_3}{\sigma_3}+\ \rho_{12}\rho_{13}-\rho_{23}\ \frac{x_2}{\sigma_2}\frac{x_3}{\sigma_3}\right]. \qquad (51)$$

Это уравнение описывает эллипсоид в трехмерном пространстве, ориентация которого и соотношения главных осей зависят от набора коэффициентов корреляции $\rho_{12},\ \rho_{13},\ \rho_{23}$. В частности, при $\rho_{12}=\rho_{13}=\rho_{23}=0$ оно превращается в уравнение сферы

$$\left(\frac{x_1}{\sigma_1}\right)^2+\left(\frac{x_2}{\sigma_2}\right)^2+\left(\frac{x_3}{\sigma_3}\right)^2=L^2.$$

На *Рис.10* приведен пример эллипсоида (51), имеющего следующие параметры: $L=3,\ \rho_{12}=0.4,\ \rho_{13}=0.6,\ \rho_{23}=0.8$. Видно, что оси этого эллипсоида не совпадают с осями системы координат, в которой описывается этот эллипсоид. Степень этого отличия определяется величинами коэффициентов корреляции $\rho_{12},\ \rho_{13},\ \rho_{23}$.

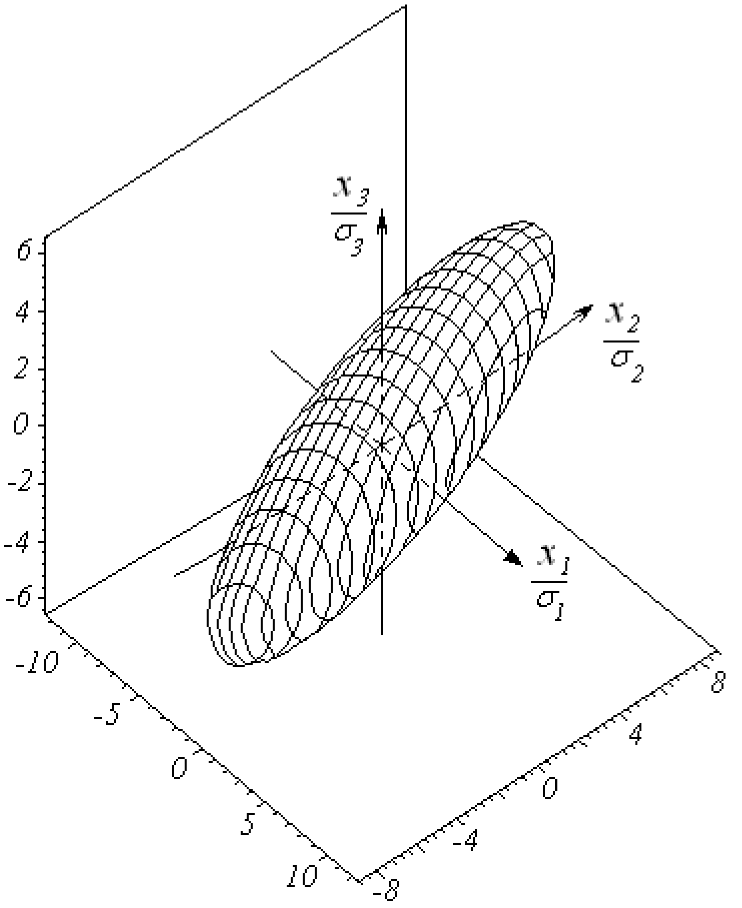


Рис.10 — Эллипсоид равной плотности вероятности трехмерного распределения

Если гауссовское распределение имеет больше трех измерений, то поверхность его равной плотности описывается эллипсоидом в соответствующем гиперпространстве, и графическое изображение его невозможно. Однако в таком эллипсоиде с целью его исследования можно выделить трехмерные или двумерные составляющие и изобразить их графически.

Из формул (48) и (49) можно получить выражения для плотностей вероятности меньшего, чем k, числа переменных, проинтегрировав их в бесконечных пределах по «лишним» переменным.

13 Энтропия и взаимная информация многомерных совокупностей гауссовских случайных величин

Явное вычисление величины энтропии и взаимной информации требует знания описывающих случайные векторы ξ, $\boldsymbol{\eta}$ и ζ многомерных распределений вероятностей. Такое вычисление не может быть проведено сразу в общем виде, а требует специального рассмотрения для каждого конкретного случая. Рассмотрим лишь наиболее важный (и одновременно один из простейших) такой случай. Будем считать, что векторы ξ, $\boldsymbol{\eta}$ и ζ – вещественные непрерывные гауссовские. Это означает, что все многомерные распределения вероятностей ξ, $\boldsymbol{\eta}$ и $\zeta = (\xi, \eta)$ являются многомерными распределениями Гаусса (многомерными нормальными распределениями) (см. раздел 12). Для таких распределений вычисление информации $I_{\xi\eta}$ может быть продвинуто существенно дальше, чем в общем случае. Получим этот результат.

Итак, пусть $\xi = (\xi_1, \ldots, \xi_k)^T$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_l)^T$ и $\zeta = (\xi, \eta) = (\xi_1, \ldots, \xi_k, \eta_1, \ldots, \eta_l)^T$ — случайные векторы-столбцы, распределенные по многомерным законам Гаусса. Будем считать, что математические ожидания всех компонент векторов ξ и ζ равны нулю (это предположение не является ограничением, так как информация $I_{\xi\eta}$ не зависит от значения математических ожидании, на что указывалось выше). В таком случае величины ξ, $\boldsymbol{\eta}$ и ζ будут описываться плотностями вида (49):

$$\left.\begin{array}{c} p_\xi(\boldsymbol{x}) = \dfrac{\exp[-\frac{1}{2}(\boldsymbol{x}^T A_\xi^{-1}\boldsymbol{x})]}{(2\pi)^{k/2}(\det A_\xi)^{1/2}}, \quad a \quad p_\eta(\boldsymbol{y}) = \dfrac{\exp[-\frac{1}{2}(\boldsymbol{y}^T A_\eta^{-1}\boldsymbol{y})]}{(2\pi)^{l/2}(\det A_\eta)^{1/2}}, \quad б \\ p_{\xi\eta}(\boldsymbol{z}) = \dfrac{\exp[-\frac{1}{2}(\boldsymbol{z}^T A_{\xi\eta}^{-1}\boldsymbol{z})]}{(2\pi)^{(k+l)/2}(\det A_{\xi\eta})^{1/2}}, \quad в \end{array}\right\} \quad (52)$$

где векторы-столбцы $\boldsymbol{x}$ и $\boldsymbol{y}$ — точки k- мерного и соответственно l- мерного пространств, и вектор $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) = (x_1, ..., x_k, y_1, ..., y_l)^T$ есть объединение векторов $\boldsymbol{x}$ и $\boldsymbol{y}$. Входящая в последнее из этих распределений квадратная блочная матрица размера $(k+l)\times(k+l)$

$$A_{\xi\eta} = \begin{bmatrix} A_\xi & D_{\xi\eta} \\ D_{\xi\eta}^T & A_\eta \end{bmatrix} \quad (53)$$

состоит из следующих матриц корреляционных моментов:

• входящей в формулу (52,*а*) квадратной матрицы A_ξ размера $k\times k$ вектора ξ, элементами которой являются корреляционные моменты компонент вектора ξ

$$a_{mn} = B_{\xi_m\xi_n} = \mathbf{M}[\xi_m\xi_n], \quad (1 \le m, n \le k), \quad (54)$$

• входящей в формулу (52,*б*) квадратной матрицы A_η размера $l\times l$ вектора $\boldsymbol{\eta}$, элементы которой — корреляционные моменты компонент вектора $\boldsymbol{\eta}$

$$b_{mn} = B_{\eta_m\eta_n} = \mathbf{M}[\eta_m\eta_n], \quad (1 \le m, n \le l), \quad (55)$$

• прямоугольной матрицы $D_{\xi\eta}$ размера $k\times l$ векторов ξ и $\boldsymbol{\eta}$, элементами которой являются взаимные корреляционные моменты компонент векторов ξ и $\boldsymbol{\eta}$

$$d_{ij} = B_{\xi_i\eta_j} = \mathbf{M}[\xi_i\eta_j], \quad (1 \le i \le k, 1 \le j \le l). \quad (56)$$

Здесь A^{-1} — матрица, обратная матрице A, ^{T} — как и ранее, знак транспонирования, $\det A$ — определитель матрицы A. Все входящие сюда определители, по условию, неотрицательны.

Вычислим вначале дифференциальную энтропию случайного k- мерного гауссовского вектора ξ. Для этого подставим плотность распределения вероятностей $p_\xi(\boldsymbol{x})$, определяемую формулой (52,*а*), в многомерный вариант формулы (41), считая в последней логарифм натуральным. Получим в натуральных единицах

$$H(\xi)=\frac{1}{2}\left[k\ln(2\pi)+\ln\det A_\xi\right]\int_{-\infty}^{\infty}p_\xi(\boldsymbol{x})\,d\boldsymbol{x}+\frac{1}{2}\int_{-\infty}^{\infty}(\boldsymbol{x}^T A_\xi^{-1}\boldsymbol{x})\,p_\xi(\boldsymbol{x})\,d\boldsymbol{x}. \qquad (57)$$

По условиям нормировки для любого распределения вероятностей значение первого входящего в это выражение интеграла есть $\int_{-\infty}^{\infty}p_\xi(\boldsymbol{x})\,d\boldsymbol{x}=1.$

Вычислим второй интеграл суммы (57). Обозначим его как

$$I_1=\int_{-\infty}^{\infty}(\boldsymbol{x}^T A_\xi^{-1}\boldsymbol{x})\,p_\xi(\boldsymbol{x})\,d\boldsymbol{x}.$$

Запишем это выражение с учетом формулы (52,*а*):

$$I_1=\frac{1}{(2\pi)^{k/2}\sqrt{\det A_\xi}}\int_{-\infty}^{\infty}(\boldsymbol{x}^T A_\xi^{-1}\boldsymbol{x})\exp\left[-\frac{1}{2}(\boldsymbol{x}^T A_\xi^{-1}\boldsymbol{x})\right]d\boldsymbol{x}. \qquad (58)$$

Любая матрица A корреляционных моментов вещественных случайных величин (а, следовательно, и обратная ей A^{-1}) вещественна и симметрична. Из теории матриц известно [3], что для такого типа матриц существует так называемая *ортогональная* матрица Q, которая приводит эту матрицу к диагональному виду с помощью следующей процедуры:

$$Q^T A_\xi^{-1} Q=\begin{bmatrix}\lambda_1 & 0 & 0 & \dots & 0\\ 0 & \lambda_2 & 0 & \dots & 0\\ \dots & \dots & \dots & \dots & 0\\ 0 & 0 & 0 & \dots & \lambda_k\end{bmatrix}. \qquad (59)$$

Входящие в полученную матрицу числа λ_i есть *собственные значения* матрицы A_ξ^{-1}. Они действительны и неотрицательны. В результате такого преобразования эллипсоид постоянной плотности вероятностей оказывается ориентированным своими осями вдоль осей избранной системы координат, а в описывающее его уравнение координаты входят только в виде квадратов и отсутствуют произведения этих координат.

Матрица Q является ортогональной, если она удовлетворяет соотношению $Q^T = Q^{-1}$.

Произведем в интеграле (58) замену переменных

$$\boldsymbol{x} = Q\,\boldsymbol{y}, \tag{60}$$

где Q - указанная ортогональная матрица (пока нам неизвестная), дополнительно считая, что она имеет единичный определитель:

$$\det Q = \det Q^{-1} = \det Q^T = 1. \tag{61}$$

Преобразование (60) с условием (61) приводит к повороту без зеркальных отражений в многомерном пространстве выбранной ортогональной системы координат, в которой измеряются векторы $\boldsymbol{x}$, без изменения ее масштабов по осям. Из этого следует, что элемент многомерного объема при таком повороте остается неизменным: $d\boldsymbol{x} = |\det Q|\,d\boldsymbol{y} = d\boldsymbol{y}$. Из формулы (60) имеем $\boldsymbol{x}^T = \boldsymbol{y}^T Q^T$. В результате такой замены с учетом равенства (59) получаем

$$\boldsymbol{x}^T A_\xi^{-1} \boldsymbol{x} = \boldsymbol{y}^T \; Q^T A_\xi^{-1} Q \; \boldsymbol{y} =$$

$$= \; y_1, y_2, \ldots, y_k \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \lambda_k \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_k \end{pmatrix} = \sum_{i=1}^{k} \lambda_i y_i^2 .$$

Поэтому после этой замены интеграл (58) принимает вид

$$I_1 = \frac{I_2}{2\pi^{k/2}\sqrt{\det A_\xi}}, \tag{62}$$

где обозначено

$$I_2 = \sum_{i=1}^{k} \lambda_i \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} y_i^2 \prod_{j=1}^{k} \exp\left(-\frac{1}{2}\lambda_j y_j^2\right) dy_1\, dy_2 .. dy_k . \tag{63}$$

Выделим в этой сумме первое слагаемое, соответствующее значению $i = 1$, т.е.

$$\lambda_1 \cdot \int_{-\infty}^{\infty} y_1^2 \exp\left(-\frac{1}{2}\lambda_1 y_1^2\right) dy_1 \cdot \prod_{j=2}^{k} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\lambda_j y_j^2\right) dy_j . \qquad (64)$$

Первый из входящих в него интегралов – табличный

$$\int_{-\infty}^{\infty} y_1^2 \exp\left(-\frac{1}{2}\lambda_1 y_1^2\right) dy_1 = \frac{\sqrt{2\pi}}{\lambda_1^{3/2}} .$$

Интегралы, входящие в последующее произведение формулы (2.50), также табличные:

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\lambda_j y_j^2\right) dy_j = \frac{\sqrt{2\pi}}{\lambda_j^{1/2}} .$$

Поэтому слагаемое (64) равно $(2\pi)^{k/2}(\lambda_1\lambda_2 ... \lambda_k)^{-1/2}$ и, следовательно, выражение (63) состоит из k таких одинаковых слагаемых, т.е.

$$I_2 = k\,(2\pi)^{k/2}(\lambda_1\lambda_2 ... \lambda_k)^{-1/2} . \qquad (65)$$

Далее найдем определитель левой и правой частей выражения (59). С учетом (61) имеем определитель его левой части $\det(Q^T A_\xi^{-1} Q) = \det A_\xi^{-1}$ и, в свою очередь, в силу диагональности матрицы определитель правой части равен произведению ее диагональных элементов:

$$\det\begin{bmatrix} \lambda_1 & 0 & 0 & ... & 0 \\ 0 & \lambda_2 & 0 & ... & 0 \\ ... & ... & ... & ... & 0 \\ 0 & 0 & 0 & ... & \lambda_k \end{bmatrix} = \lambda_1\lambda_2 ... \lambda_k .$$

Приравнивание этих определителей дает $\lambda_1\lambda_2 ... \lambda_k = \det A_\xi^{-1} = 1/\det A_\xi$. Учитывая это соотношение в (65), упростим его: $I_2 = k\,(2\pi)^{k/2}(\det A_\xi)^{1/2}$.

Подстановка этого выражения в формулу (62) дает $I_1 = k$. С учетом найденных соотношений выражение (57) принимает вид

$$H(\xi) = \frac{1}{2}\left[k \ln(2\pi) + \ln\det A_\xi + k\right]$$

или $H(\xi)=\ln\sqrt{(2\pi e)^k\det A_\xi}$. При ином основании логарифма его следует записать как

$$H(\xi)=\log\sqrt{(2\pi e)^k\det A_\xi}.$$

Аналогичные рассуждения дают для векторов $\boldsymbol{\eta}$ и $(\boldsymbol{\xi},\boldsymbol{\eta})$:

$$H(\boldsymbol{\eta})=\log\sqrt{(2\pi e)^l\det A_\eta}$$

и

$$H(\boldsymbol{\xi},\boldsymbol{\eta})=\log\sqrt{(2\pi e)^{k+l}\det A_{\xi\eta}}.$$

Подставляя последние соотношения в формулу (37), получим выражение для искомой взаимной информации [3]

$$I_{\xi\eta}=\frac{1}{2}\log\frac{\det A_\xi\det A_\eta}{\det A_{\xi\eta}}. \tag{66}$$

Отсюда следует, что взаимная информация систем гауссовских случайных величин определяется только матрицами собственных и взаимных корреляционных моментов составляющих их пар случайных векторов.

На практике, как правило, наибольшую ценность представляет знание величины именно взаимной информации $I_{\xi\eta}$, а не дифференциальных энтропий $H(\xi)$, $H(\boldsymbol{\eta})$ и $H(\xi,\boldsymbol{\eta})$. Такую роль она будет играть и в наших исследованиях.

Обратимся к неравенству $I(X,Y)\geq 0$ (п.2 Раздела 7). Применение его к формуле (66) дает неравенство для входящих в него определителей

$$\frac{\det A_\xi\det A_\eta}{\det A_{\xi\eta}}\geq 1$$

или

$$\det A_\xi\det A_\eta\geq\det A_{\xi\eta}. \tag{67}$$

Учитывая, что блочная матрица $A_{\xi\eta}$ дается выражением (53), $\det A_{\xi\eta}$ обращается в произведение $\det A_\xi\cdot\det A_\eta$ тогда и только тогда, когда матрица $D_{\xi\eta}$ взаимных корреляционных моментов векторов ξ и $\boldsymbol{\eta}$ состоит из нулевых элементов. Отсюда следует, что неравенство (67) обращается в равенство только при взаимной статистической независимости компонент векторов ξ и $\boldsymbol{\eta}$. При этом, как и следовало ожидать, взаимная информация этих векторов равна нулю.

14 Преобразование формулы для взаимной информации

Модифицируем формулу (66) для взаимной информации. Учтем, что корреляционные моменты Раздела 13 могут быть записаны в виде

$$\left.\begin{aligned} a_{mn} &= \sigma_{\xi m}\sigma_{\xi n}R_{mn}\ \ \xi \ \ , \quad 1 \le m,n \le k \ \ , \\ b_{mn} &= \sigma_{\eta m}\sigma_{\eta n}R_{mn}\ \ \boldsymbol{\eta} \ \ , \quad 1 \le m,n \le l \ \ . \end{aligned}\right\}, \tag{68}$$

$$d_{ij} = \sigma_{\xi i}\sigma_{\eta j}R_{ij}\ \ \xi,\boldsymbol{\eta} \ \ , \quad 1 \le i \le k,\ 1 \le j \le l \ \ . \tag{69}$$

Здесь $R_{mn}\ \ \xi\ \ = R_{nm}\ \ \xi$ — коэффициент корреляции случайных величин ξ_m и ξ_n, $R_{mn}\ \ \boldsymbol{\eta}\ \ = R_{nm}\ \ \boldsymbol{\eta}$ — коэффициент корреляции случайных величин η_m и η_n, $R_{ij}\ \ \xi,\boldsymbol{\eta}\ \ = R_{ji}\ \ \boldsymbol{\eta},\xi$ — коэффициент корреляции случайных величин ξ_i и η_j. Величины $\sigma_{\xi i}$ и $\sigma_{\eta j}$ - среднеквадратические отклонения случайных величин ξ_i и η_j. Обратим внимание, что имеют место равенства $R_{mm}\ \ \xi\ \ = R_{mm}\ \ \boldsymbol{\eta}\ \ = 1$ для любых индексов m.

С учетом соотношений (68) входящая в формулу (66) матрица A_ξ корреляционных моментов компонент вектора ξ согласно представлению (50) может быть записана как

$$A_\xi = \begin{bmatrix} \sigma_{\xi 1}\sigma_{\xi 1} & \sigma_{\xi 1}\sigma_{\xi 2}R_{12}\ \ \xi & \sigma_{\xi 1}\sigma_{\xi 3}R_{13}\ \ \xi & \ldots & \sigma_{\xi 1}\sigma_{\xi k}R_{1k}\ \ \xi \\ \sigma_{\xi 2}\sigma_{\xi 1}R_{21}\ \ \xi & \sigma_{\xi 2}\sigma_{\xi 2} & \sigma_{\xi 2}\sigma_{\xi 3}R_{23}\ \ \xi & \ldots & \sigma_{\xi 2}\sigma_{\xi k}R_{2k}\ \ \xi \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \sigma_{\xi k}\sigma_{\xi 1}R_{k1}\ \ \xi & \sigma_{\xi k}\sigma_{\xi 2}R_{k2}\ \ \xi & \sigma_{\xi k}\sigma_{\xi 3}R_{k3}\ \ \xi & \ldots & \sigma_{\xi k}\sigma_{\xi k} \end{bmatrix}. \tag{70}$$

Вынесем из первой строки этой матрицы общую для всех ее элементов величину $\sigma_{\xi 1}$, из второй $\sigma_{\xi 2}$ и т.д. и из последней строки $\sigma_{\xi k}$. Затем вынесем из первого столбца получившейся матрицы общую величину $\sigma_{\xi 1}$, из второго столбца - общую величину $\sigma_{\xi 2}$ и т.д. вплоть до последнего k - го столбца. Вычисляя определитель преобразованной таким образом матрицы, найдем

$$\det A_\xi = \sigma^2_{\xi 1}\sigma^2_{\xi 2}...\sigma^2_{\xi k} \det R_\xi , \qquad (71)$$

где

$$R_\xi = \begin{bmatrix} 1 & R_{12}\;\xi & R_{13}\;\xi & ... & R_{1k}\;\xi \\ R_{21}\;\xi & 1 & R_{23}\;\xi & ... & R_{2k}\;\xi \\ ... & ... & ... & ... & ... \\ R_{k1}\;\xi & R_{k2}\;\xi & R_{k3}\;\xi & ... & 1 \end{bmatrix} \qquad (72)$$

— матрица размера $k \times k$ коэффициентов корреляции компонент вектора ξ .

Аналогично найдем

$$\det A_\eta = \sigma^2_{\eta 1}\sigma^2_{\eta 2}...\sigma^2_{\eta l} \det R_\eta , \qquad (73)$$

где

$$R_\eta = \begin{bmatrix} 1 & R_{12}\;\eta & R_{13}\;\eta & ... & R_{1l}\;\eta \\ R_{21}\;\eta & 1 & R_{23}\;\eta & ... & R_{2l}\;\eta \\ ... & ... & ... & ... & ... \\ R_{l1}\;\eta & R_{l2}\;\eta & R_{l3}\;\eta & ... & 1 \end{bmatrix} \qquad (74)$$

— матрица размера $l \times l$ коэффициентов корреляции компонент вектора η .

Учитывая представление (69), прямоугольную матрицу $D_{\xi\eta}$ размера $k \times l$ взаимных корреляционных моментов компонент векторов ξ и η запишем в виде

$$D_{\xi\eta} = \begin{bmatrix} \sigma_{\xi 1}\sigma_{\eta 1}R_{11}\;\xi,\eta & \sigma_{\xi 1}\sigma_{\eta 2}R_{12}\;\xi,\eta & \dots & \sigma_{\xi 1}\sigma_{\eta l}R_{1l}\;\xi,\eta \\ \sigma_{\xi 2}\sigma_{\eta 1}R_{21}\;\xi,\eta & \sigma_{\xi 2}\sigma_{\eta 2}R_{22}\;\xi,\eta & \dots & \sigma_{\xi 2}\sigma_{\eta k}R_{2l}\;\xi,\eta \\ \dots & \dots & \dots & \dots \\ \sigma_{\xi k}\sigma_{\eta 1}R_{k1}\;\xi,\eta & \sigma_{\xi k}\sigma_{\eta 2}R_{k2}\;\xi,\eta & \dots & \sigma_{\xi k}\sigma_{\eta l}R_{kl}\;\xi,\eta \end{bmatrix}. \qquad (75)$$

Подставляя матрицу (70), аналогичным образом записанную матрицу A_η и (75) в блочную матрицу (53) и вынося из ее строк и столбцов одинаковые сомножители способом, описанным выше, найдем

$$\det A_{\xi\eta} = \sigma^2_{\xi 1}\sigma^2_{\xi 2}...\sigma^2_{\xi k} \quad \sigma^2_{\eta 1}\sigma^2_{\eta 2}...\sigma^2_{\eta\, l} \ \det R_{\ \xi,\eta}\ , \qquad (76)$$

где блочная матрица собственных и взаимных коэффициентов корреляции компонент обоих векторов ξ и η имеет вид

$$R_{\ \xi,\eta} = \begin{bmatrix} R_{\xi} & R_{\xi\eta} \\ R^T_{\xi\eta} & R_{\eta} \end{bmatrix}. \qquad (77)$$

Она состоит из блоков, представляющих собой корреляционные матрицы (72) и (74) и матрицу взаимных коэффициентов корреляции компонент векторов ξ и η, получаемую из матрицы (75)

$$R_{\xi\eta} = \begin{bmatrix} R_{11}\ \xi,\eta & R_{12}\ \xi,\eta & R_{13}\ \xi,\eta & ... & R_{1l}\ \xi,\eta \\ R_{21}\ \xi,\eta & R_{22}\ \xi,\eta & R_{23}\ \xi,\eta & ... & R_{2l}\ \xi,\eta \\ ... & ... & ... & ... & ... \\ R_{k1}\ \xi,\eta & R_{k2}\ \xi,\eta & R_{k3}\ \xi,\eta & ... & R_{kl}\ \xi,\eta \end{bmatrix}. \qquad (78)$$

Учитывая определители (71), (73) и (76) в формуле (66), получим искомое выражение:

$$I_{\xi\eta} = \frac{1}{2}\log\frac{\det R_{\xi}\det R_{\eta}}{\det R_{(\xi,\eta)}}. \qquad (79)$$

Отсюда следует, что взаимная информация случайных векторов ξ и η, распределенных по нормальному закону, зависит только от коэффициентов корреляции как между компонентами самих векторов ξ и η, так и от их взаимных коэффициентов корреляции, но не зависит ни от математических ожиданий, ни от дисперсий этих векторов.

Полученная формула в некоторых случаях более удобна для применения, чем формула (66).

Обобщая приведенные выше свойства взаимной информации, можно записать

$$I\ \ \xi,\eta\ = I\ \ a_1\xi + \boldsymbol{b}_1, a_2\eta + \boldsymbol{b}_2\ \ ,$$

где a_1, a_2 — вещественные неслучайные числа, $\boldsymbol{b}_1, \boldsymbol{b}_2$— неслучайные вещественные векторы. Таким образом, взаимная информация не зависит ни от средних значений, ни от амплитуд случайных векторов. Это позволяет упрощать исходные формулы уже при постановке задач нахождения взаимной информации.

15 Примеры вычисления взаимной информации непрерывных случайных величин

Найдем и проанализируем выражения для взаимной информации в некоторых частных случаях.

Пример 6. Пусть имеются две одномерных гауссовских случайных величины ξ и η, коэффициент корреляции которых равен $R_{\xi\eta} = r$. Для них, согласно формулам (72) и (73), имеют место значения $R_\xi = R_\eta = 1$ и в соответствии с формулой (78) нормированный взаимный корреляционный момент есть $R_{\xi\eta} = r$.

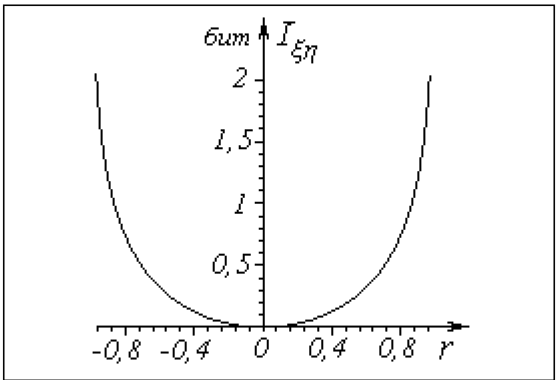


Рис.11 — Зависимость взаимной информации двух гауссовских случайных величин от их коэффициента корреляции

Следовательно, блочная матрица (77) здесь имеет вид $R_{\xi,\eta} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$. Отсюда имеем $\det R_\xi = \det R_\eta = 1$, $\det R_{\xi,\eta} = 1 - r^2$, и формула (79) упрощается:

$$I_{\xi\eta} = -\frac{1}{2}\log\ 1 - r^2\ . \qquad (80)$$

Эта зависимость представлена на *Рис.11*. Из нее следует, что взаимная информация двух нормальных случайных величин полностью определяется модулем коэффициента их корреляции. Это означает, что, зная значение r для нормального распределения, можно по формуле (2.78) вычислить величину $I_{\xi\eta}$. В то же время в силу четного характера функции $I_{\xi\eta}$ r найти значение r по величине $I_{\xi\eta}$ можно только с точностью до знака.

Из *Рис.11* видно, что большие значения $I_{\xi\eta}$ имеют место при величинах r, близких к единице. □

Пример 7. Допустим, что две случайные величины являются отсчетами одного и того же гауссовского процесса, полученные в разные моменты времени: $\xi=\xi(t_1)$ и $\eta=\xi(t_2)$. Величина коэффициента корреляции между ними представляет собой нормированную корреляционную функцию процесса $\xi(t)$, аргументом которой является временной сдвиг $\tau=t_2-t_1$, т.е. $r=R(\tau)$. Поэтому здесь формула (80) принимает вид

$$I_{\xi(t_1)\xi(t_2)}=-\frac{1}{2}\log\left(1-R^2(\tau)\right). \tag{81}$$

Конкретизируем это выражение для двух разных корреляционных функций: экспоненциальной

$$R_{эксп}(\tau)=\exp\left(-|\tau|/\tau_x\right) \tag{82}$$

и колоколообразной

$$R_{кол}(\tau)=\exp\left(-\frac{\pi}{4}\left(\frac{\tau}{\tau_x}\right)^2\right). \tag{83}$$

Им соответствуют величины взаимной информации

$$I_{эксп}=-\frac{1}{2}\log\left[1-\exp\left(-2|\tau|/\tau_x\right)\right] \text{ и } I_{кол}=-\frac{1}{2}\log\left[1-\exp\left(-\frac{\pi}{2}\left(\frac{\tau}{\tau_x}\right)^2\right)\right].$$

Здесь τ_x — характерный масштаб процесса.

Результаты расчетов этих функций представлены на *Рис.12*. При $\tau=0$ они обращаются в бесконечность, и с ростом τ монотонно убывают, причем для экспоненциальной корреляционной функции круче, чем для колоколообразной.

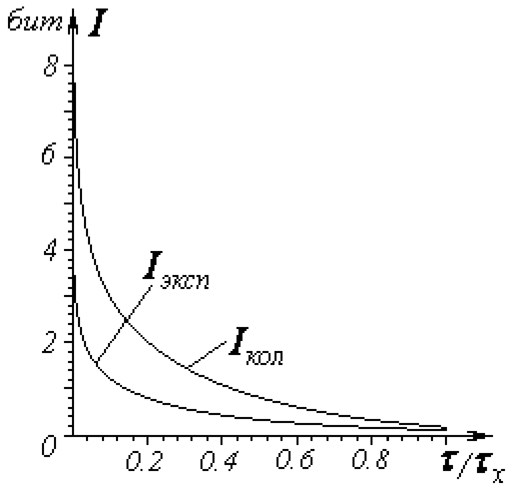


Рис.12. Зависимости взаимной информации пары случайных отсчетов процессов от величины сдвига

Из анализа рисунка следует, что независимо от вида корреляционной функции процесса взаимная информация между его отсчетами на интервале, превышающем τ_x, отсутствует. Это означает, что зная величину $\xi\ \ t_1$, ничего нельзя сказать о величине $\xi\ \ t_1+\tau_x$.□

Пример 8. Пусть случайная величина η является суммой случайной величины ξ и не коррелированного с ней шума n, т.е. $\eta=\xi+n$. Все величины ξ, η и n предполагаются гауссовскими и центрированными.

Найдем взаимную информацию величин ξ и η. Для этого воспользуемся формулой (66).

Согласно соотношению (54), входящая в нее матрица A_ξ содержит единственную составляющую $a_{11}=\mathbf{M}\left[\xi\xi\right]=\sigma_\xi^2$. Аналогично, единственная составляющая матрицы A_η

$$b_{11}=\mathbf{M}\left[\eta\eta\right]=\mathbf{M}\left[\ \xi+n\ ^{2}\right]=\mathbf{M}\left[\xi^2\right]+2\mathbf{M}\left[\xi n\right]+\mathbf{M}\left[n^2\right]=\sigma_\xi^2+\sigma_ш^2,$$

где $\sigma_ш^2=\mathbf{M}\left[n^2\right]$ — дисперсия шума. Величина $\mathbf{M}\left[\xi n\right]=0$ в силу некоррелированности ξ и n. При выводе этой формулы использовано следующее свойство: *Математическое ожидание суммы случайных величин равно сумме их математических ожиданий независимо от законов распределения*

Единственная составляющая матрицы $D_{\xi\eta}$ есть

$$d_{11}=\mathbf{M}\left[\xi\eta\right]=\mathbf{M}\left[\xi(\xi+n)\right]=\mathbf{M}\left[\xi^2\right]+\mathbf{M}\left[\xi n\right]=\sigma_\xi^2$$

и этом блочная матрица (53) приобретает вид

$$A_{\xi\eta}=\begin{bmatrix}\sigma_\xi^2 & \sigma_\xi^2\\ \sigma_\xi^2 & \sigma_\xi^2+\sigma_ш^2\end{bmatrix}.$$

Поэтому имеем $\det A_\xi=\sigma_\xi^2$, $\det A_\eta=\sigma_\xi^2+\sigma_ш^2$, $\det A_{\xi\eta}=\sigma_\xi^2\sigma_ш^2$. Подстановка этих значений в формулу (53) и простейшие преобразования дают формулу

$$I_{\xi\eta}=\frac{1}{2}\log\left(1+\frac{\sigma_\xi^2}{\sigma_ш^2}\right), \tag{84}$$

впервые полученную К.Шенноном [17].

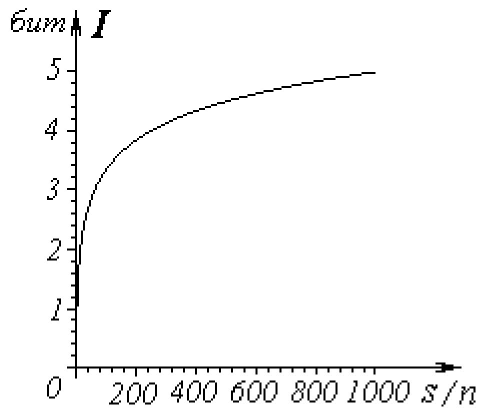


Рис.13. Зависимость взаимной информации от отношения сигнал / шум

Эта зависимость представлена на *Рис.13*. Ее аргументом является отношение дисперсий (мощностей) сигнала и шума $s/n=\sigma_\xi^2/\sigma_ш^2$, которое обычно называют «отношением сигнал-шум». Обратим внимание, что это отношение безразмерно. Если одновременно увеличить мощность сигнала и в то же количество раз мощность шума, то величина взаимной информации не изменится. □

Пример 9. Допустим, что случайная величина ξ является отсчетом гауссовского процесса, полученным в момент времени t, т.е. $\xi=\xi(t)$, а случайная величина η — суммой отсчетом того же процесса в момент времени $t+\tau$ и не коррелированного с ним шума , т.е. $\eta=\xi(t-\tau)+n(t-\tau)$.

Найдем взаимную информацию величин ξ и η. Для этого воспользуемся формулой (66).

Согласно соотношению (54), входящая в нее матрица A_ξ содержит единственную составляющую $a_{11} = \mathbf{M}\left[\xi\ t\ \ \xi\ t\ \right] = \sigma_\xi^2$. Аналогично, единственная составляющая матрицы A_η

$$b_{11} = \mathbf{M}\left[\eta\eta\right] = \mathbf{M}\left[\ \xi\ t-\tau\ + n\ t-\tau\ \ ^2\right] =$$

$$= \mathbf{M}\left[\xi^2\ t-\tau\ \right] + 2\mathbf{M}\left[\xi\ t-\tau\ \ n\ t-\tau\ \right] + \mathbf{M}\left[n^2\ t-\tau\ \right] = \sigma_\xi^2 + \sigma_ш^2.$$

Единственная составляющая матрицы $D_{\xi\eta}$ есть

$$d_{11} = \mathbf{M}\left[\xi\eta\right] = \mathbf{M}\left[\xi\ t\ \ \ \xi\ t-\tau\ + n\ t-\tau\ \ \right] =$$

$$= \mathbf{M}\left[\xi\ t\ \ \xi\ t-\tau\ \right] + \mathbf{M}\left[\xi\ t\ \ n\ t-\tau\ \right] = \sigma_\xi^2 R_\xi\ \tau\ .$$

При этом блочная матрица (53) приобретает вид

$$A_{\xi\eta} = \begin{bmatrix} \sigma_\xi^2 & \sigma_\xi^2 R_\xi\ \tau \\ \sigma_\xi^2 R_\xi\ \tau & \sigma_\xi^2 + \sigma_ш^2 \end{bmatrix}.$$

Поэтому имеем

$$\det A_\xi = \sigma_\xi^2, \quad \det A_\eta = \sigma_\xi^2 + \sigma_ш^2,$$

$$\det A_{\xi\eta} = \sigma_\xi^4 + \sigma_\xi^2\sigma_ш^2 - \sigma_\xi^4 R_\xi^2\ \ \tau\ = \sigma_\xi^4\left[1 - R_\xi^2\ \ \tau\ + \sigma_ш^2/\sigma_\xi^2\right].$$

Подстановка этих значений в формулу (66) и простейшие преобразования дают

$$I_{\xi\eta} = \frac{1}{2}\log\frac{\sigma_\xi^2\ \ \sigma_\xi^2 + \sigma_ш^2}{\sigma_\xi^4\left[1 - R_\xi^2\ \ \tau\ + \sigma_ш^2/\sigma_\xi^2\right]} = \frac{1}{2}\log\frac{1 + \sigma_ш^2/\sigma_\xi^2}{1 + \sigma_ш^2/\sigma_\xi^2 - R_\xi^2\ \ \tau} =$$

$$= \frac{1}{2}\log\frac{1}{1 - \dfrac{R_\xi^2\ \ \tau}{1 + \sigma_ш^2/\sigma_\xi^2}} = -\frac{1}{2}\log\left[1 - \frac{R_\xi^2\ \ \tau}{1 + \sigma_ш^2/\sigma_\xi^2}\right]. \qquad (85)$$

Найденная взаимная информация зависит не только от корреляционной функции, но и от отношения $\sigma_ш^2/\sigma_\xi^2$ мощности шума к мощности сигнала.

При отсутствии шума, т.е. при $\sigma_ш^2 = 0$ это выражение обращается в формулу (81). При $\tau = 0$ отсюда получим формулу (84).

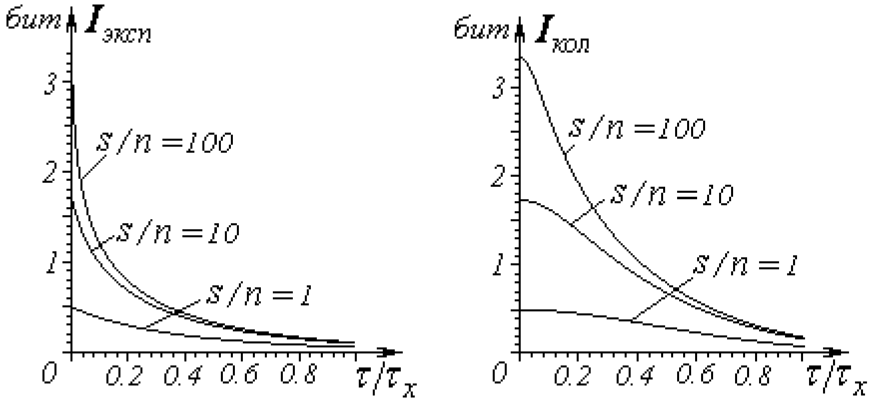


Рис.14 Зависимость взаимной информации от временного сдвига при разных величинах отношения сигнал / шум для экспоненциальной и колоколообразной корреляционных функций

Зависимость (85) для процессов с обеими рассмотренными выше корреляционными функциями представлена на *Рис.14* Из него следует, при одинаковом отношении сигнал-шум в процессе с колоколообразной корреляционной функцией информационная взаимосвязанность распространяется на больший временной интервал, чем в процессе с экспоненциальной корреляционной функцией. □

Пример 10. Обобщим полученные результаты. Для этого вернемся к формулам (53) - (56). Будем считать, что переменная ξ является отсчетом гауссовского процесса $\xi = \xi\ t$, а переменная $\eta = \eta\ t$ — отсчет связанного с ним другого гауссовского процесса. Найдем общее выражение для взаимной информации такой пары процессов.

В данном случае матрица A_ξ состоит из одного элемента — дисперсии процесса $\xi\ t$: $a_{11} = \mathbf{M}[\xi\ t\ \xi\ t\ ] = \sigma_\xi^2$. Ее определитель $\det A_\xi = \sigma_\xi^2$. Матрица A_η также состоит из одного элемента — дисперсии процесса $\eta\ t$: $b_{11} = \mathbf{M}[\eta\ t\ \eta\ t\ ] = \sigma_\eta^2$. Ее определитель $\det A_\eta = \sigma_\eta^2$.

Прямоугольная матрица $D_{\xi\eta}$ размера 1×1 процессов $\xi\; t$ и $\eta\; t$ состоит из одного элемента, представляющего собой взаимную корреляционную функцию этих процессов $d_{11} = \mathbf{M}[\xi\; t\; \eta\; t\,]$.

Учитывая значения полученных величин, найдем квадратную блочную матрицу (53) размера 2×2

$$A_{\xi\eta} = \begin{bmatrix} A_\xi & D_{\xi\eta} \\ D_{\xi\eta}^T & A_\eta \end{bmatrix} = \begin{bmatrix} \sigma_\xi^2 & \mathrm{M}[\xi\; t\; \eta\; t\,] \\ \mathrm{M}[\xi\; t\; \eta\; t\,] & \sigma_\eta^2 \end{bmatrix}.$$

Определитель этой матрицы $\det A_{\xi\eta} = \sigma_\xi^2 \sigma_\eta^2 - \mathrm{M}^2[\xi\; t\; \eta\; t\,]$.

Подставляя найденные определители в формулу (66), получаем искомое выражение для взаимной информации рассматриваемой пары случайных процессов:

$$\begin{aligned} I_{\xi\; t\; \eta\; t} &= \frac{1}{2}\log \frac{\sigma_\xi^2 \sigma_\eta^2}{\sigma_\xi^2 \sigma_\eta^2 - \mathrm{M}^2[\xi\; t\; \eta\; t\,]} = \\ &= -\frac{1}{2}\log\; 1 - \mathrm{M}^2[\xi\; t\; \eta\; t\,] \big/ \sigma_\xi^2 \sigma_\eta^2 \end{aligned}. \qquad (86)$$ □

Пример 11. Воспользуемся найденным соотношением для определения величины взаимной информации между гауссовским стационарным процессом $\xi\; t$ и процессом, полученным из него путем линейного преобразования (линейной фильтрации)

$$4.\;\; \eta\; t = \sum_{i=1}^{k} c_i \xi\; t - i\tau\; + n\; t\;. \qquad (87)$$

Здесь c_i - постоянные коэффициенты, обеспечивающие требуемый тип фильтрации, а τ - заданный интервал между соседними отсчетами процесса $\xi\; t$, $n\; t$ — аддитивный белый гауссовский шум, не коррелированный с процессом $\xi\; t$. Известно, что процесс, полученный в результате линейного преобразования стационарного гауссовского процесса, также является стационарным гауссовским. Поэтому для решения нашей задачи применима формула (86).

Найдем составляющие формулы (86), соответствующие процессу $\eta\ t$. Дисперсия этого процесса есть

$$\sigma_\eta^2 = \mathrm{M}[\eta\ t\ \eta\ t\ ] = \mathrm{M}[\left\{\sum_{i=1}^{k} c_i \xi\ t - i\tau\ + n\ t\right\}\left\{\sum_{j=1}^{k} c_j \xi\ t - j\tau\ + n\ t\right\}] =$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j \mathrm{M}[\xi\ t - i\tau\ \xi\ t - j\tau\ ] + \sigma_ш^2 = \sigma_\xi^2 \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j R_\xi\left[\ i - j\ \tau\right] + \sigma_ш^2.$$

Далее, имеем

$$\mathrm{M}[\xi\ t\ \eta\ t\ ] = \mathrm{M}[\xi\ t\ \left\{\sum_{i=1}^{k} c_i \xi\ t - i\tau\ + n\ t\right\}] =$$

$$= \sum_{i=1}^{k} c_i \mathrm{M}[\xi\ t\ \xi\ t - i\tau\ ] = \sigma_\xi^2 \sum_{i=1}^{k} c_i R_\xi\ i\tau\ .$$

Обозначим функции

$$f_1\ \tau\ = \sum_{i=1}^{k} c_i R_\xi\ i\tau \quad \text{и} \quad f_2\ \tau\ = \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j R_\xi\left[\ i - j\ \tau\right].$$

При этом формула (86) принимает вид

$$I_{\xi\ t\ \eta\ t}\ \tau\ = -\frac{1}{2}\log\ 1 - f_1^2\ \tau\ \Big/\ f_2\ \tau\ + \sigma_ш^2 / \sigma_\xi^2. \quad . \qquad (88)$$

Обратим внимание на то, что полученное выражение зависит не от дисперсий сигнала и шума порознь, а от отношения сигнал-шум.

Преобразуем функцию $f_2\ \tau$. Для этого выпишем ее слагаемые с учетом симметрии корреляционной функции в виде элементов матрицы:

$$\begin{bmatrix} c_1^2 & c_1 c_2 R\ \tau & \dots & c_1 c_{k-1} R\left[\ k-2\ \tau\right] & c_1 c_k R\left[\ k-1\ \tau\right] \\ c_2 c_1 R\ \tau & c_2^2 & \dots & c_2 c_{k-1} R\left[\ k-3\ \tau\right] & c_2 c_k R\left[\ k-2\ \tau\right] \\ \dots & \dots & \dots & \dots & \dots \\ c_{k-1} c_1 R\left[\ k-2\ \tau\right] & c_{k-1} c_2 R\left[\ k-3\ \tau\right] & \dots & c_{k-1}^2 & c_{k-1} c_k R\ \tau \\ c_k c_1 R\left[\ k-1\ \tau\right] & c_k c_2 R\left[\ k-2\ \tau\right] & \dots & c_k c_{k-1} R\ \tau & c_k^2 \end{bmatrix},$$

где под R понимается R_ξ.

Суммируя члены матрицы по диагоналям, получим

$$f_2(\tau)=\sum_{i=1}^{k} c_i^2+2R_\xi(\tau)\sum_{i=1}^{k-1} c_i c_{i+1}+2R_\xi\ 2\tau\ \sum_{i=1}^{k-2} c_i c_{i+2}+$$

$$+2R_\xi\ 3\tau\ \sum_{i=1}^{k-3} c_i c_{i+3}+\ldots+2R_\xi\left[(k-1)\tau\right]c_1 c_k.$$

Применяя найденные соотношения, рассмотрим несколько примеров вычисления информации об исходном процессе, содержащейся в результате его линейного преобразования. □

Пример 12.. Выберем $k=1$ и $C_1=1$. При этом из (87) имеем

$$\eta\ t\ =\xi\ t-\tau\ +n\ t\ ,$$

т.е. выходной процесс получается из входного сдвигом на интервал τ и добавлением «белого» шума $n\ t$. Далее, имеем $f_1\ \tau\ =R_\xi\ \tau$ и $f_2\ \tau\ =1$. Подставив эти значения в формулу (88), находим

$$I_{\xi\ t\ \eta\ t}\ \tau\ =-\frac{1}{2}\log\left[1-R_\xi^2\ \tau\ /\ 1+\sigma_ш^2/\sigma_\xi^2.\ \right],$$

что, как и следовало ожидать, совпадает с формулой (85). □

Пример 13.. Выберем $k=2$. Формула (87) принимает вид

$$\eta\ t\ =c_1\xi\ t-\tau\ +c_2\xi\ t-2\tau\ +n\ t\ . \tag{89}$$

Далее,

$$f_1\ \tau\ =c_1R_\xi\ \tau\ +c_2R_\xi\ 2\tau\ ,\ f_2(\tau)=c_1^2+c_2^2+2c_1c_2R_\xi(\tau)$$

и взаимная информация определяется формулой

$$I_{\xi\ t\ \eta\ t}\ \tau\ =-\frac{1}{2}\log\left\{1-\frac{c_1^2R_\xi^2\ \tau\ +2c_1c_2R_\xi\ \tau\ R_\xi\ 2\tau\ +c_2^2R_\xi^2\ 2\tau}{c_1^2+c_2^2+2c_1c_2R_\xi\ \tau\ +\sigma_ш^2/\sigma_\xi^2}\right\}. \tag{90}$$

Рассмотрим частные случаи этого выражения. □

Пример 14. Положив в формуле (89) $c_1=c_2=1/2$, получим выходной сигнал в виде осредненного по двум отсчетам входного сигнала

$$\eta(t)=\frac{1}{2}\left[\xi(t-\tau)+\xi(t-2\tau)\right]+n(t).$$

При этом формула (91) для взаимной информации исходного сигнала и осредненного принимает вид

$$I_{\xi(t)\eta(t)}(\tau)=-\frac{1}{2}\log\left\{1-\frac{\frac{1}{2}\left[R_\xi(\tau)+R_\xi(2\tau)\right]^2}{1+R_\xi(\tau)+2\sigma_ш^2/\sigma_\xi^2}\right\}.$$

Расчеты показывают, что полученные результаты практически не отличаются от представленных на *Рис.14.*

Совсем иной результат получается, если выбрать коэффициенты следующим образом: $c_1=1, c_2=-1$, что соответствует приближенному дифференцированию [5] функции $\xi(t)$:

$$\eta(t)=\xi(t-\tau)-\xi(t-2\tau)+n(t).$$

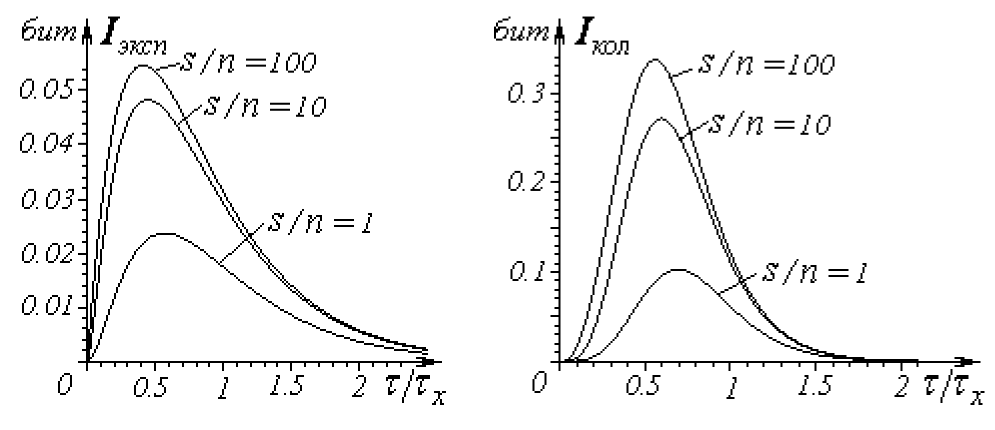


Рис.15 - Зависимость от временного сдвига взаимной информации исходного процесса и его приближенной производной

Ему соответствует формула

$$I_{\xi(t)\eta(t)}(\tau)=-\frac{1}{2}\log\left\{1-\frac{R_\xi^2(\tau)-2R_\xi(\tau)R_\xi(2\tau)+R_\xi^2(2\tau)}{2\left[1-R_\xi(\tau)\right]+\sigma_ш^2/\sigma_\xi^2}\right\}.$$

Результаты вычисления по ней представлены на *Рис.15.*

Рассмотрим этот рисунок подробнее.

При малых сдвигах τ процесс $\eta(t)$ близок к производной процесса $\xi(t)$. Из теории вероятностей известно, что случайный процесс не коррелирует со своей производной, а отсутствие корреляции гауссовских процессов ведет к их статистической независимости. Все это приводит к их информационной независимости. Поэтому при малых τ величина $I_{\xi(t)\eta(t)}(\tau)$ близка к нулю. При увеличении τ процесс $\eta(t)$ все больше отличается от производной процесса $\xi(t)$, и взаимная информация между ними растет, что видно на рисунке. При дальнейшем увеличении τ корреляция между $\xi(t)$ и $\eta(t)$ убывает, что ведет к уменьшению $I_{\xi(t)\eta(t)}(\tau)$.

Этот эффект имеет место при любом виде корреляционных функций процесса $\xi(t)$. Однако при одинаковом соотношении $|\tau|/\tau_x$ величина взаимной информации существенным образом зависит от вида корреляционной функции процесса $\xi(t)$. Для дифференцируемых процессов она намного больше, чем для недифференцируемых. □

Пример 15. В Примере 13 была рассмотрена информационная связь между процессом $\xi(t)$ и взвешенной суммой его величин, сдвинутых соответственно на временные интервалы τ и 2τ. Найдем величину информационной связи между $\xi(t)$ и двумя его значениями

$$\eta_1(t)=\xi(t-\tau)+n(t-\tau), \quad \eta_2(t)=\xi(t-2\tau)+n(t-2\tau),$$

рассматриваемым порознь, без их суммирования. Иными словами, получим формулу $I_{\xi\boldsymbol{\eta}}(\tau)$, позволяющую вычислить величину информационной связи скаляра $\xi(t)$ с вектором $\boldsymbol{\eta}(t)=(\eta_1(t),\eta_2(t))^T$.

Воспользуемся формулой (66). Входящий в нее вектор ξ имеет единственную составляющую $\xi(t)$. Соответствующая ему матрица согласно (54) имеет определитель $\det A_\xi=\sigma_\xi^2$.

Квадратная матрица $A_{\boldsymbol{\eta}}$ размера 2×2 вектора $\boldsymbol{\eta}$ имеет своими элементами корреляционные моменты компонент вектора $\boldsymbol{\eta}$

$$b_{mn} = B_{\eta_m \eta_n} = \mathbf{M}[\eta_m \eta_n], \ \ (1 \le m,n \le 2),$$

т.е.

$$A_{\boldsymbol{\eta}} = \begin{bmatrix} \sigma_\xi^2 + \sigma_ш^2 & \sigma_\xi^2 R_\xi(\tau) \\ \sigma_\xi^2 R_\xi(\tau) & \sigma_\xi^2 + \sigma_ш^2 \end{bmatrix}.$$

Прямоугольная матрица $D_{\xi\boldsymbol{\eta}}$ размера 1×2 скаляра ξ и вектора $\boldsymbol{\eta}$ имеет элементами взаимные корреляционные моменты компонент векторов ξ и $\boldsymbol{\eta}$:
$d_{1j} = B_{\xi\,\eta_j} = \mathbf{M}[\xi \eta_j], \ (1 \le j \le 2)$, т.е.

$$D_{\xi\boldsymbol{\eta}} = \begin{bmatrix} \sigma_\xi^2 R_\xi(\tau) & \sigma_\xi^2 R_\xi(2\tau) \end{bmatrix}.$$

Поэтому имеем блочную матрицу (53)

$$A_{\xi\boldsymbol{\eta}} = \begin{bmatrix} \sigma_\xi^2 & \sigma_\xi^2 R_\xi(\tau) & \sigma_\xi^2 R_\xi(2\tau) \\ \sigma_\xi^2 R_\xi(\tau) & \sigma_\xi^2 + \sigma_ш^2 & \sigma_\xi^2 R_\xi(\tau) \\ \sigma_\xi^2 R_\xi(2\tau) & \sigma_\xi^2 R_\xi(\tau) & \sigma_\xi^2 + \sigma_ш^2 \end{bmatrix}.$$

Вычисления дают

$$\det A_\xi = \sigma_\xi^2, \ \det A_{\boldsymbol{\eta}} = \sigma_\xi^4 \left[(1+\lambda)^2 - R_\xi^2(\tau)\right],$$

$$\det A_{\xi\boldsymbol{\eta}} = \sigma_\xi^6 \left\{(1+\lambda)^2 - \left[\lambda + 2\left(1 - R_\xi(2\tau)\right)\right] R_\xi^2(\tau) - (1+\lambda) R_\xi^2(2\tau)\right\},$$

где обозначено отношение шум-сигнал $\lambda = \sigma_ш^2 / \sigma_\xi^2$. Подставляя эти величины в формулу (66), найдем величину искомой взаимной информации

$$I_{\xi\boldsymbol{\eta}} = \frac{1}{2}\log \frac{(1+\lambda)^2 - R_\xi^2(\tau)}{(1+\lambda)^2 - \left[\lambda + 2\left(1 - R_\xi(2\tau)\right)\right] R_\xi^2(\tau) - (1+\lambda) R_\xi^2(2\tau)}. \qquad (91)$$

При $\tau = 0$ отсюда имеем

$$I_{\xi\boldsymbol{\eta}} = \frac{1}{2}\log\left(1 + 2\frac{\sigma_\xi^2}{\sigma_ш^2}\right)$$

и при $\sigma_{ш}^{2}/\sigma_{\xi}^{2}=0$

$$I_{\xi\eta}=-\frac{1}{2}\log\left(1-\frac{\left[R_{\xi}(\tau)-R_{\xi}(2\tau)\right]^{2}}{1-R_{\xi}^{2}(\tau)}\right).$$

Результаты вычислений по формуле (91) представлены на *Рис.16*, Полученные кривые качественно похожи на соответствующие кривые *Рис.14*, дающие величину взаимной информации между процессом $\xi=\xi(t)$ и его единственным отсчетом $\xi(t-\tau)$, сдвинутым на временной интервал τ с аддитивным шумом $\eta=\xi(t-\tau)+n(t-\tau)$.

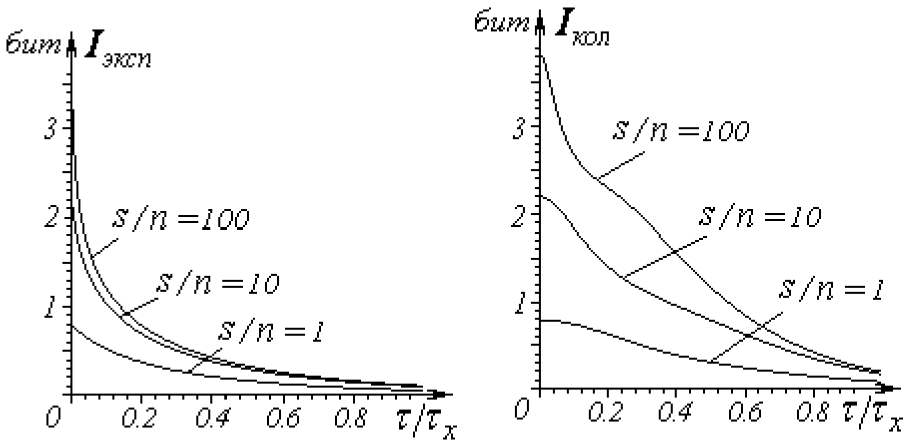


Рис.16 - Зависимость (2.102) для экспоненциальной и колоколообразной корреляционных функций

В рассматриваемом примере в вектор $\eta(t)$ добавлен отсчет $\xi(t-2\tau)$ исходного процесса, что должно привести к увеличению взаимной информации. Именно этот эффект и наблюдается при сравнении *Рис.14* и *Рис.16* Величина ΔI такого приращения взаимной информации изображена на *Рис.17*-

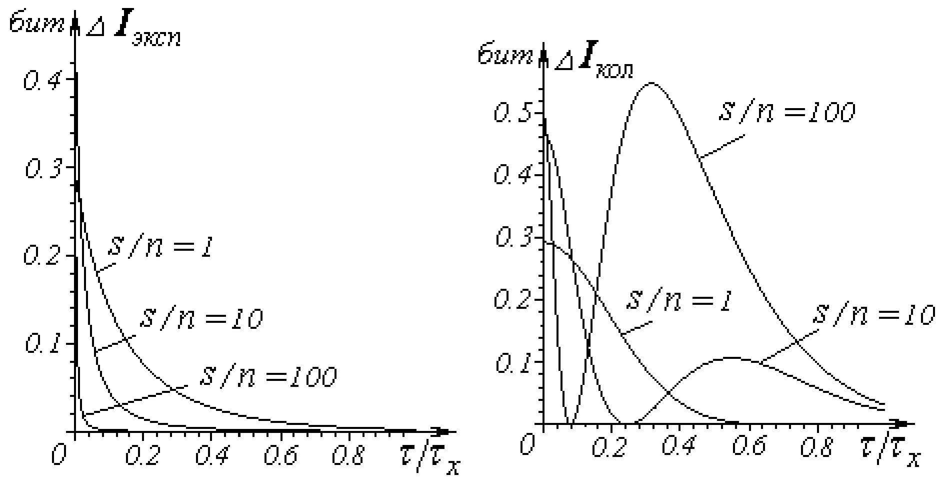


Рис.17- Приращение взаимной информации для экспоненциальной и колоколообразной корреляционных функций

Эта величина невелика и зависит от вида корреляционной функции. □

Пример 16. Рассмотрим пример применения найденных соотношений для определения взаимной информации между элементами двумерного гауссовского статистически однородного изотропного поля. Статистическая однородность является расширением понятия стационарности случайного процесса на поле, в нашем случае двумерное. Статистические характеристики такого поля одинаковы по всему полю. Корреляционная функция между двумя любыми точками изотропного поля зависит только от абсолютной величины пространственного сдвига между ними и не зависит от угла ориентации этого сдвига в поле.

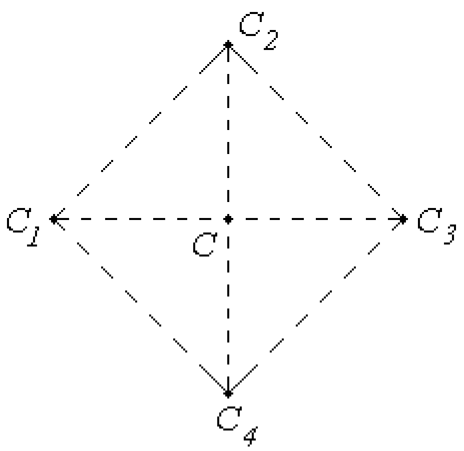


Рис.18. Схема наблюдения случайного поля

Обратимся к схеме наблюдения такого поля, представленной на *Рис.18.* Допустим, что в точках C_1, C_2, C_3 и C_4 производится измерение поля в сумме с гауссовским статистически однородным изотропным некоррелированным шумом: $\eta_1 = \xi_1 + n_1$, $\eta_2 = \xi_2 + n_2$, $\eta_3 = \xi_3 + n_3$, $\eta_4 = \xi_4 + n_4$. В совокупности они составляют вектор $\boldsymbol{\eta} = \eta_1, \eta_2, \eta_3, \eta_4{}^T$. В точке C нас интересует истинное значение поля ξ.

Обозначим одинаковые расстояния между точкой C и любой из точек наблюдения поля C_1, C_2, C_3 и C_4 как d. При этом расстояния между парами точек наблюдения имеют величины

$$\left.\begin{aligned} CC_1 = CC_2 = CC_3 = CC_4 = d, \\ C_1C_2 = C_1C_4 = C_2C_3 = C_3C_4 = \sqrt{2}\,d, \\ C_1C_3 = C_2C_4 = 2d. \end{aligned}\right\}$$

Построим матрицы, входящие в формулу (66).

Как и ранее, матрица A_ξ состоит из одного элемента — дисперсии поля ξ: $a_{11} = \mathbf{M}[\xi\xi] = \sigma_\xi^2$. Ее определитель $\det A_\xi = \sigma_\xi^2$.

Элементами квадратной матрицы $A_{\boldsymbol{\eta}}$ размера 4×4 вектора $\boldsymbol{\eta}$ являются корреляционные моменты компонент вектора $\boldsymbol{\eta}$, т.е. $\mathbf{M}[\eta_m \eta_n]$, где $(1 \le m, n \le 4)$. Поэтому матрица $A_{\boldsymbol{\eta}}$ имеет вид

$$A_{\boldsymbol{\eta}} = \begin{bmatrix} \mathbf{M}[\eta_1\eta_1], & \mathbf{M}[\eta_1\eta_2] & \mathbf{M}[\eta_1\eta_3] & \mathbf{M}[\eta_1\eta_4] \\ \mathbf{M}[\eta_2\eta_1] & \mathbf{M}[\eta_2\eta_2] & \mathbf{M}[\eta_2\eta_3] & \mathbf{M}[\eta_2\eta_4] \\ \mathbf{M}[\eta_3\eta_1] & \mathbf{M}[\eta_3\eta_2] & \mathbf{M}[\eta_3\eta_3] & \mathbf{M}[\eta_3\eta_4] \\ \mathbf{M}[\eta_4\eta_1] & \mathbf{M}[\eta_4\eta_2] & \mathbf{M}[\eta_4\eta_3] & \mathbf{M}[\eta_4\eta_4] \end{bmatrix}.$$

Учитывая приведенные выше соотношения, получим

$$A_{\boldsymbol{\eta}} = \begin{bmatrix} \sigma_\xi^2 + \sigma_{\text{ш}}^2 & \sigma_\xi^2 R_\xi\ \sqrt{2}d & \sigma_\xi^2 R_\xi\ 2d & \sigma_\xi^2 R_\xi\ \sqrt{2}d \\ \sigma_\xi^2 R_\xi\ \sqrt{2}d & \sigma_\xi^2 + \sigma_{\text{ш}}^2 & \sigma_\xi^2 \sigma_\xi^2 R_\xi\ \sqrt{2}d & \sigma_\xi^2 R_\xi\ 2d \\ \sigma_\xi^2 R_\xi\ 2d & \sigma_\xi^2 \sigma_\xi^2 R_\xi\ \sqrt{2}d & \sigma_\xi^2 + \sigma_{\text{ш}}^2 & \sigma_\xi^2 \sigma_\xi^2 R_\xi\ \sqrt{2}d \\ \sigma_\xi^2 R_\xi\ \sqrt{2}d & \sigma_\xi^2 R_\xi\ 2d & \sigma_\xi^2 \sigma_\xi^2 R_\xi\ \sqrt{2}d & \sigma_\xi^2 + \sigma_{\text{ш}}^2 \end{bmatrix}$$

Определитель этой матрицы есть

$$\det A_{\boldsymbol{\eta}} = \sigma_\xi^8 \left[1+\lambda - R_\xi\ 2d\ \right]^2 \left[1+\lambda + R_\xi\ 2d\ - 2R_\xi\ \sqrt{2}d\ \right]\times$$
$$\times\left[1+\lambda + R_\xi\ 2d\ + 2R_\xi\ \sqrt{2}d\ \right].$$

Элементами прямоугольной матрицы $D_{\xi\boldsymbol{\eta}}$ размера 1×4 скаляра ξ и вектора $\boldsymbol{\eta}$ являются взаимные корреляционные моменты ξ и компонент вектора $\boldsymbol{\eta}$, т.е. $\mathbf{M}[\xi\eta_j]$, ($1\le j\le 4$). Эта матрица имеет вид

$$D_{\xi\boldsymbol{\eta}} = \left[\xi\eta_1 \quad \xi\eta_2 \quad \xi\eta_3 \quad \xi\eta_4\right].$$

Учет приведенных выше соотношений дает

$$D_{\xi\boldsymbol{\eta}} = \left[\sigma_\xi^2 R_\xi\ d \quad \sigma_\xi^2 R_\xi\ d \quad \sigma_\xi^2 R_\xi\ d \quad \sigma_\xi^2 R_\xi\ d\ \right].$$

Теперь можно найти блочную матрицу (53):

$$A_{\xi\boldsymbol{\eta}} = \begin{bmatrix} \sigma_\xi^2 & \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2 R_\xi\ d \\ \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2+\sigma_{\boldsymbol{ш}}^2 & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} & \sigma_\xi^2 R_\xi\ 2\mathbf{d} & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} \\ \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} & \sigma_\xi^2+\sigma_{\boldsymbol{ш}}^2 & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} & \sigma_\xi^2 R_\xi\ 2\mathbf{d} \\ \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2 R_\xi\ 2\mathbf{d} & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} & \sigma_\xi^2+\sigma_{\boldsymbol{ш}}^2 & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} \\ \sigma_\xi^2 R_\xi\ d & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} & \sigma_\xi^2 R_\xi\ 2\mathbf{d} & \sigma_\xi^2 R_\xi\ \sqrt{2}\mathbf{d} & \sigma_\xi^2+\sigma_{\boldsymbol{ш}}^2 \end{bmatrix}.$$

Определитель этой матрицы

$$\det A_{\xi\boldsymbol{\eta}} = \sigma_\xi^{10} \left[1+\lambda - R_\xi\ 2d\ \right]^2 \left[1+\lambda + R_\xi\ 2d\ - 2R_\xi\ \sqrt{2}d\ \right]\times$$
$$\times\left[1+\lambda + R_\xi\ 2d\ + 2R_\xi\ \sqrt{2}d\ - 4R_\xi^2\ d\ \right].$$

Подставив найденные выражения для $\det A_\xi$, $\det A_{\boldsymbol{\eta}}$ и $\det A_{\xi\boldsymbol{\eta}}$ в формулу (66), получаем искомое выражение для взаимной информации

$$I_{\xi\eta}=\frac{1}{2}\log\frac{1+\lambda+R_\xi\;2d\;+2R_\xi\;\sqrt{2}d}{1+\lambda+R_\xi\;2d\;+2R_\xi\;\sqrt{2}d\;-4R_\xi^2\;d}$$

или

$$I_{\xi\eta}=-\frac{1}{2}\log\left[1-\frac{4R_\xi^2\;d}{1+\lambda+R_\xi\;2d\;+2R_\xi\;\sqrt{2}d}\right]. \qquad (92)$$

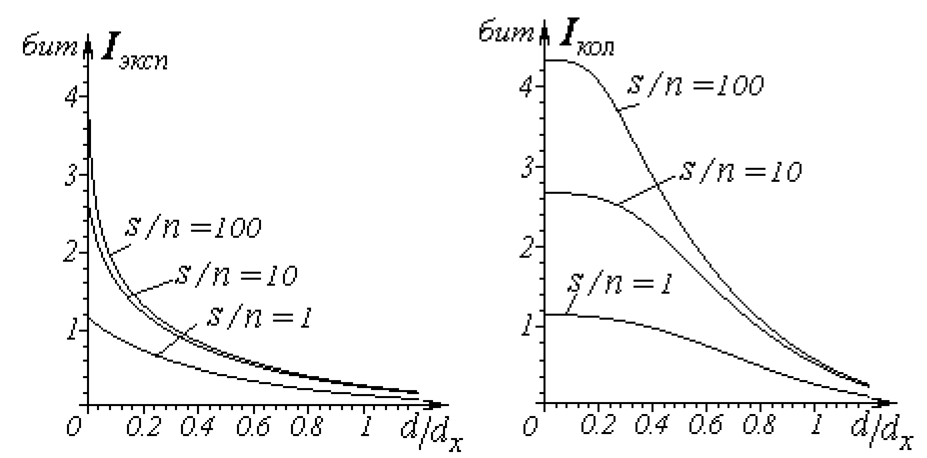


Рис.19 - Взаимная информация между значениями поля схемы Рис.18.

На *Рис.19* представлены результаты расчетов взаимной информации по этой формуле для полей с корреляционными функциями

$$R_{эксп}(d)=\exp\;-d/d_x\;,\;R_{кол}(d)=\exp\left(-\frac{\pi}{4}\left(\frac{d}{d_x}\right)^2\right),$$

которые являются пространственными аналогами функций (82) и (83). В них d_x - характерный пространственный масштаб. Качественно эти зависимости сходны с представленными на *Рис.14*, но отличаются от них количественно.

Применяя формулу (92), следует иметь ввиду, что не все корреляционные функции случайных процессов могут быть корреляционными функциями случайных полей. □

16 Примеры вычисления взаимной информации дискретных случайных величин

Зададим для двух дискретных случайных величин ξ_i и η_j множества их возможных значений $X = \{x_1, x_2, \ldots, x_m\}$ и $Y = \{y_1, y_2, \ldots, y_n\}$ и с помощью матрицы (18) совместные двумерные распределения вероятностей $P(x_i, y_j)$ всех возможных комбинаций состояний x_i $1 \le i \le m$ ансамбля X и состояний y_j $1 \le j \le n$ ансамбля Y, соответствующие этим значениям.

Рассмотрим пару примеров, соответствующих таким величинам.

Пример 17. Пусть даны две дискретные случайные величины, которые могут принимать значения

$$X = \{-2, -1, 0, 3, 4\}, \qquad Y = \{2, 3, 4, 5, 6, 7\}$$

с вероятностями

$$P(x_i, y_j) = \begin{bmatrix} 0.01 & 0.01 & 0.02 & 0.02 & 0.01 & 0.01 \\ 0.03 & 0.01 & 0.04 & 0.01 & 0.04 & 0.02 \\ 0.01 & 0.07 & 0.05 & 0.06 & 0.07 & 0.09 \\ 0.04 & 0.02 & 0.02 & 0.03 & 0.02 & 0.03 \\ 0.05 & 0.06 & 0.04 & 0.02 & 0.04 & 0.05 \end{bmatrix},$$

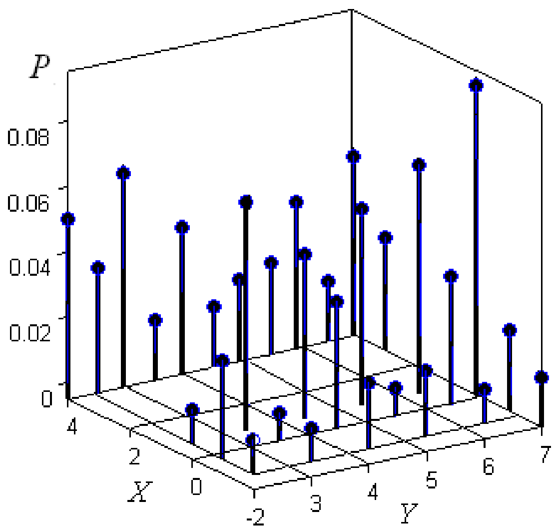


Рис.20 - Распределение двумерной вероятности

Это распределение изображено на *Рис.20*. Из него получим одномерные распределения вероятностей

$$P(x_i) = 0.08,\ 0.15,\ 0.35,\ 0.16,\ 0.26\ ,$$

$$P(y_j) = 0.14,\ 0.17,\ 0.17,\ 0.14,\ 0.18,\ 0.20\ .$$

Они представлены на *Рис.21*.

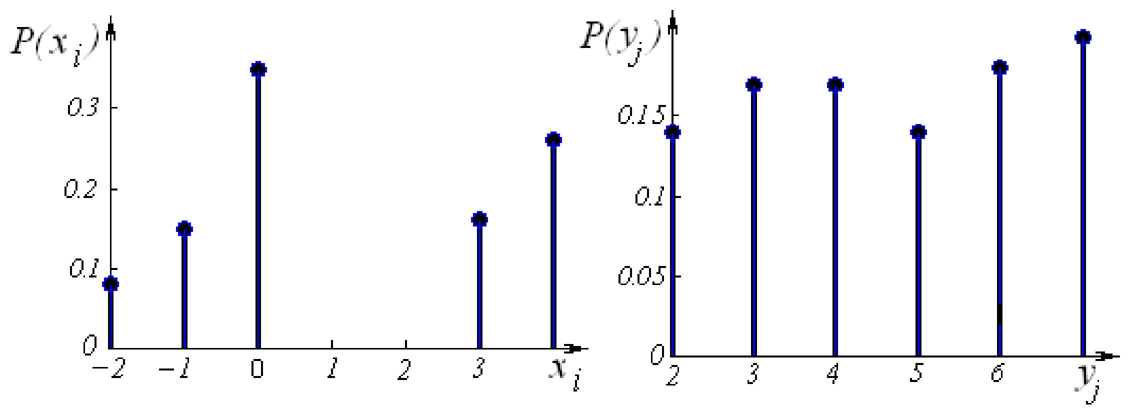


Рис.21. - Распределения одномерных вероятностей

Совместная энтропия этого распределения $H(X,Y) = 4.6244$, энтропия дискретной величины X равна $H(X) = 2.1605$, случайной величины Y равна $H(Y) = 2.5731$. Взаимная информация между этими величинами есть

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = 0.1091. \ \square$$

Пример 18. Найдем информационные характеристики цветного цифрового изображения пейзажа (*Рис.22*).

а)

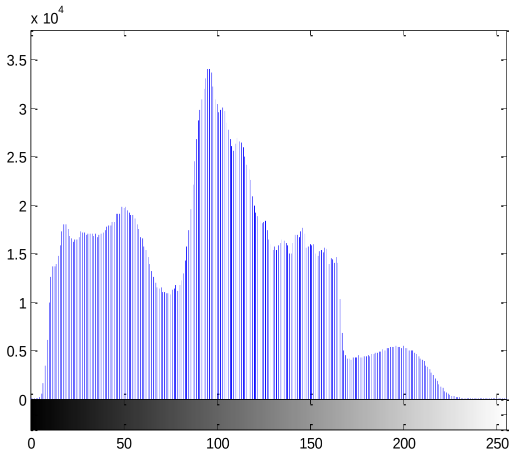
б)

в)

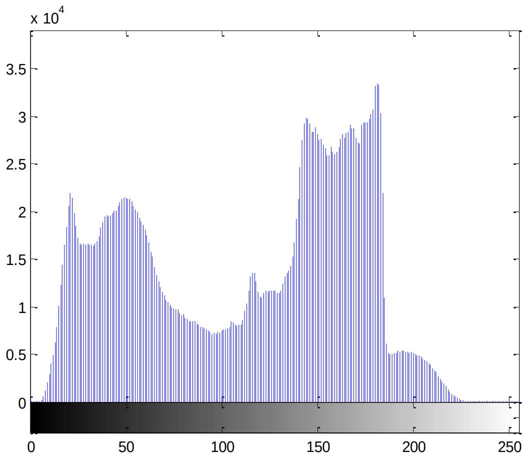
г)

Рис.22 -. Цифровое изображение пейзажа: а) полноцветное, б) красная, в) зеленая и г) синяя составляющие

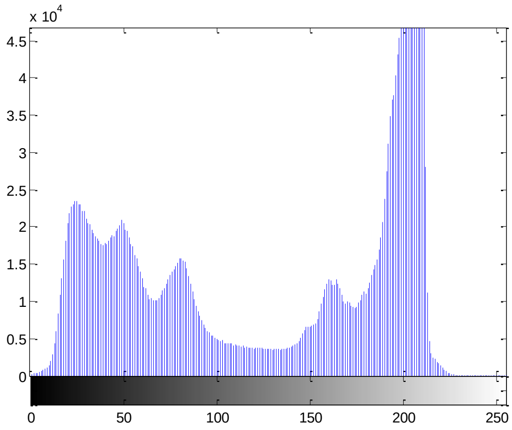
На *Рис.22,а* представлено полноцветное изображение этого пейзажа, а на рисунках *б), в)* и *г)* — соответственно его красная, зеленая и синяя составляющие в монохромной форме в восьмиразрядном коде. Визуальное сопоставление этих составляющих позволяет отметить, что они имеют в некоторых деталях определенное сходство друг с другом и довольно значительно отличаются в других деталях. Отсюда можно сделать вывод, что взаимная информация между различными составляющими должна быть довольно большой, но заметно меньшей, чем их энтропии.



а)



б)



в)

Рис.23 . Гистограммы а) красной, б) зеленой и в) синей составляющих изображения пейзажа

Найдем количественные величины этих характеристик.

На *Рис.23* представлены гистограммы указанных составляющих. Видно, что распределения вероятностей этих составляющих существенно отличаются как от гауссовского распределения, так и друг от друга. По полученным на их основании одномерным распределениям вероятностей найдены энтропии этих составляющих:

Красной составляющей $H(R) = 7.53$ *бит*;

Зеленой составляющей $H(G) = 7.48$ *бит;*

Синей составляющей $H(B) = 7.24$ *бит.*

Полученные результаты близки к 8 *битам* — емкости каждой ячейки фотоизображения, что свидетельствует об эффективном использовании памяти этого изображения.

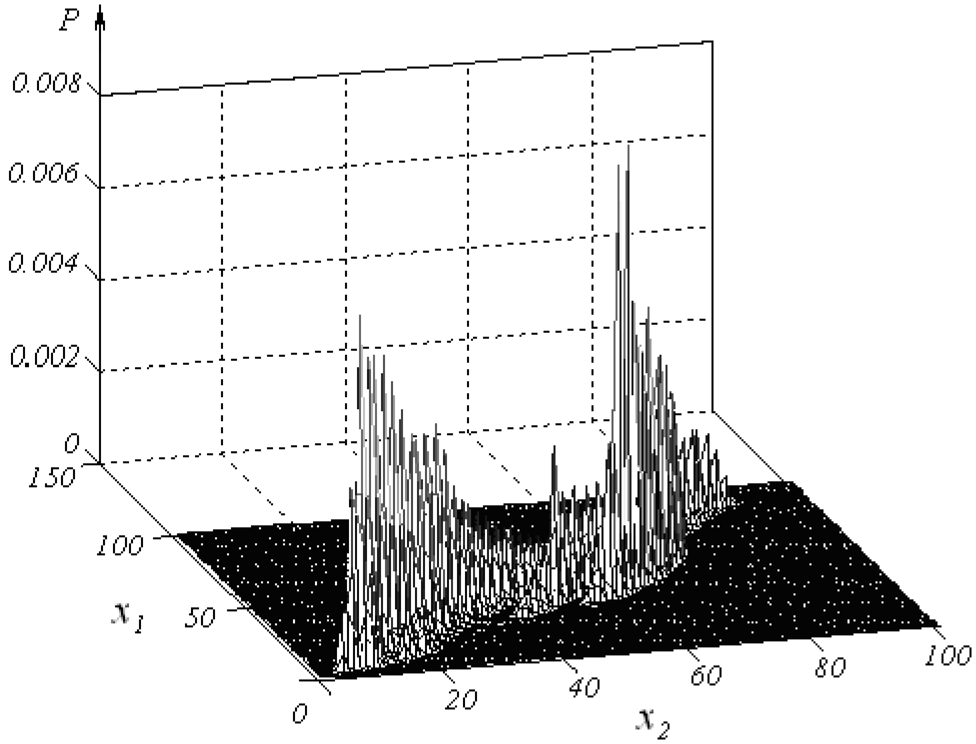


Рис.24 – Двумерное распределение вероятностей для красного и зеленого цветов

Для нахождения величины взаимной информации различных составляющих изображения требуется рассчитать двумерное распределения вероятности для этих составляющих. Пример такого распределения, полученного для рассматриваемого изображения, представлен на *Рис.24*. Расчет показывает, что взаимная информация красной и зеленой составляющих изображения пейзажа равна $I\;\;R,G\;\;=2.303$ *бит*, что составляет 31% от энтропии.

Расчеты, выполненные аналогичным образом для других пар составляющих, дают следующие результаты: взаимная информация красной и синей составляющих равна $I\;\;R,B\;\;=0.962$ *бит*, что составляет13% от энтропии, и взаимная информация зеленой и синей составляющих равна $I\;\;G,B\;\;=1.264$ *бит*, что составляет17% от энтропии. Таким образом, в рассматриваемом изображения пейзажа согласно информационному критерию наиболее тесно взаимосвязаны красная и зеленая составляющие. □

Заключение

Особую известность и популярность теория информации получила в связи следующей из нее теории помехоустойчивого кодирования, которая сформулирована К.Шенноном в виде теоремы [17]:

« 1. При любой производительности источника сообщений, меньшей, чем пропускная способность канала, существует такой способ кодирования, который позволяет обеспечить передачу всей информации, создаваемой источником сообщений, со сколь угодно малой вероятностью ошибки.

2. Не существует способа кодирования, позволяющего вести передачу информации со сколь угодно малой вероятностью ошибки, если производительность источника сообщений больше пропускной способности канала.»

Эта теорема не конструктивна: ею доказывается только существование искомого способа кодирования, но не дается метод нахождения этого способа. Тем не менее полученный в ней вывод дал понять, что такой способ существует, и это привело к исследованию различных подходов к разработке теории помехоустойчивого кодирования и созданию эффективных помехоустойчивых кодов.

Результаты, полученные в теории информации, далеко не исчерпываются указанной теоремой. В частности, как показано выше, большую роль играет взаимная информация между парами случайных объектов. В теории вероятности и математической статистике обычно степень взаимосвязи между случайными объектами определяют средствами корреляционного или спектрального анализа. Однако при этом указанные объекты должны удовлетворять некоторым свойствам, например, стационарности или статистической однородности. Ничего этого не требуется при нахождении взаимной информации. Она может быть найдена для объектов самой различной степени сложности.

Взаимную информацию нельзя измерить инструментально. Однако ее можно вычислить на основании результатов эксперимента. Зачастую эти результаты оказываются весьма объемными, как это показано выше применительно к цифровым изображениям. Тем не менее современная вычислительная техника позволяет произвести соответствующие вычисления в весьма короткие сроки, что делает применение методов теории информации весьма привлекательным и эффективным при теоретических и экспериментальных исследованиях случайных объектов и предназначенных для их обработки информационных систем.

Библиография.

1. Виленкин Н.Я. Комбинаторика./Н.Я.Виленкин. – М.: Наука, 1969. – 328 с.
2. Виноградов И.М. (гл. редактор) Математическая энциклопедия. Т.5. – М.: «Советская Энциклопедия», 1984. – 1248 стб.
3. Гельфанд И.М. О вычислении количества информации о случайной функции, содержащейся в другой такой функции/ И.М.Гельфанд, А.М.Яглом.//Успехи математических наук, том XII, вып.1(73), 1957. – С.3-52.
4. Голдман С. Теория информации. Пер. с англ. – М.: ИЛ, 1957. - 448 с.
5. Демидович Б.П., Марон И.А. Основы вычислительной математики. – М.: ФМ, 1960. – 659 с.
6. Дмитриев В.И. Прикладная теория информации/В.И.Дмитриев. – М.: Высшая школа, 1989. – 320 с.
7. Колесник В.Д., Полтырев Г.Ш. Курс теории информации.- М.: «Наука», 1982. – 416 с.
8. Колмогоров А.Н. - Теория информации и теория алгоритмов.- М.: Наука, 1987 - 304 с.
9. Королюк В.С. Справочник по теории вероятностей и математической статистике. /В.С.Королюк, Н.И.Портенко, А.В. Скороход, А.Ф.Турбин. – М.: Наука, 1985. – 640 с.

10. Куликовский Л.Ф., Мотов В.В. Теоретические основы информационных процессов.- М. «Высш. школа», 1987. – 248 с.
11. Орлов В.А. Теория информации в упражнениях и задачах/ В.А. Орлов, Л.И.Филиппов.– М.: Высш. школа, 1976.– 136 с.
12. Солодов А.В. Теория информации и ее применение к задачам автоматического управления и контроля. – М.: « Наука», 1967. – 432 с.
13. Стратонович Р.Л. Теория информации.- М.: «Сов. радио», 1975.- 424 с.
14. Тихонов В.И. Статистическая радиотехника. - М., Сов. радио, 1966. – 679 с.
15. Фихтенгольц Г.М. Основы математического анализа (2) – СПб: Лань, 2001. – 463 с.
16. Хорн Р. Матричный анализ./ Р.Хорн, Ч.Джонсон. – М.: Мир, 1989. – 655 с.
17. Шеннон К. Математическая теория связи. В кн. Работы по теории информации и кибернетике. Пер. с англ. – М.: Изд. иностр. лит-ры, 1963. – 829 с.
18. Яглом и А.М., Яглом И.М. Вероятность и информация.- М.: КомКнига, 2007.- 512 с.