1. Системы счисления. Общие понятия.
2. Позиционные системы счисления.
3. Представления чисел в позиционных системах счисления.
4. Структура чисел в позиционных системах счисления.
5. Преимущества и недостатки двоичной системы счисления.
6. Задача хранения данных.
7. Понятие информационной емкости накопителя.
8. Мера Хартли.
9. Связь позиционной системы счисления со структурой накопителя данных.
10. Возможное количество перестановок символов в сообщении и применение его в теории информации.
11. Вывод формулы для энтропии дискретной случайной величины.
12. Понятие неопределенности результата наблюдения случайного явления.
13. Энтропия как числовая характеристика случайной величины.
14. Свойства энтропии.
15. Энтропия системы случайных величин.
16. Энтропия бинарного источника сообщений.
17. Графическое изображение соотношений между энтропиями систем случайных величин.
18. Прохождение дискретной случайной величины через канал связи.
19. Частная взаимная информация.
20. Свойства частной взаимной информации.

21. Свойства энтропии совокупности случайных величин.
22. Средняя взаимная информация двух систем дискретных случайных величин.
23. Свойства средней взаимной информации.
24. Энтропия непрерывных случайных величин.
25. Средняя взаимная информация двух систем непрерывных случайных величин.
26. Многомерное нормальное (гауссовское) распределение.
27. Энтропия и взаимная информация многомерных совокупностей гауссовских случайных величин.
28. Вычисление взаимной информации двух гауссовских случайных величин с известным коэффициентом корреляции
29. Вычисление взаимной информации двух отсчетов одного и того же гауссовского процесса, полученных в разные моменты времени.
30. Вычисление взаимной информации гауссовских случайных величины x и h=x+n, где n — не коррелированный с x шум.

Ответы

1. **Системы счисления. Общие понятия.**

Счисление — совокупность приемов представления натуральных чисел. 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.

Например, у древних вавилонян узловыми являлись числа 1, 10, 60; у маори узловыми были числа 1, 11, 112, 113. 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.(употреблялись славянами и римлянами).

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).

2. Позиционные системы счисления.

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).

3. Представления чисел в позиционных системах счисления.

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). Вводят p основных знаков, называемых цифрами. Эти знаки используют для образования цифровых последовательностей, которые служат для представления натуральных чисел. Будем цифры обозначать так: $b_0,\ldots,b_{p-1}$. Каждой цифровой последовательности, образованной только из одной цифры, однозначно сопоставим одно из p-1 первых натуральных чисел или нуль (Ноль 0 не является натуральным числом и в наших рассуждениях должен быть введен дополнительно). Тогда всякое натуральное число имеет точно одно представление в p-ичной системе счисления:

$$a=b_s p^s+b_{s-1}p^{s-1}+\ldots+b_0 p^0 \qquad (1)$$

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 $a=b_s b_{s-1} \ldots b_0$.

4. Структура чисел в позиционных системах счисления.

Чем больше основание 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 $29 = 1*2^4+1*2^3+1*2^2+0*2^1+1*2^0$

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:

$$y=p^k, \qquad (2)$$

Необходимо для этой структуры найти такое p, которое минимизирует величину

$$c=pk \qquad (3)$$

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.

5. Преимущества и недостатки двоичной системы счисления.[неоднозначно]

Преимущества:

- Для ее реализации нужны элементы с двумя возможными состояниями, а не с десятью.
- Представление информации посредством только двух состояний надежно и помехоустойчиво.

- Возможность применения алгебры логики для выполнения логических преобразований.
- Двоичная арифметика проще десятичной.

Недостатки:

- Простые десятичные числа записываются в виде бесконечных двоичных дробей.
- Двоичные числа имеют много разрядов.
- Человеку тяжело воспринимать такого рода информацию.
- Требуется много символов для записи небольшого числа
- не является идеальной(т.к. Идеальной является троичная)

6. Задача хранения информации

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.

7. Понятие информационной емкости накопителя.

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 $X = \{1, 2, 3, 4, 5\}$, во втором – состояния $Y = \{a, b, c, d, e\}$. Общее число состояний, в которых они могут пребывать совместно, равно 5^2=25. Действительно, эти состояния — элементы декартова произведения

$$X \times Y = \{(1,a),(1,b),K,(1,e),K,(2,a),K,(1,e),K,(5,e)\}.$$

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.

$$C = logN, \tag{4}$$

где N — число различимых состояний накопителя. Она носит название *мера Хартли информационной емкости накопителя*.

8. Мера Хартли.

Мера Хартли - перевод числа в логарифмический вид.

$$C = logN\,, \tag{4}$$

где N — число различимых состояний накопителя. Она носит название *мера Хартли информационной емкости накопителя*. 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 измеряется в двоичных единицах, которые имеют сокращенное название *бит.*

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

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

где p — число состояний одной элементарной ячейки, а n — число таких ячеек. При этом информационная емкость накопителя $C = nlogp$ 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 .

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

$$log_p N = n log_p p = n \, . \qquad (**)$$

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 5 ячеек.

Далеко не все числа представляют собой целые степени числа . Например, 5.129… 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.

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.

9. Связь позиционной системы счисления со структурой накопителя данных.

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 *рисунке ниже*.

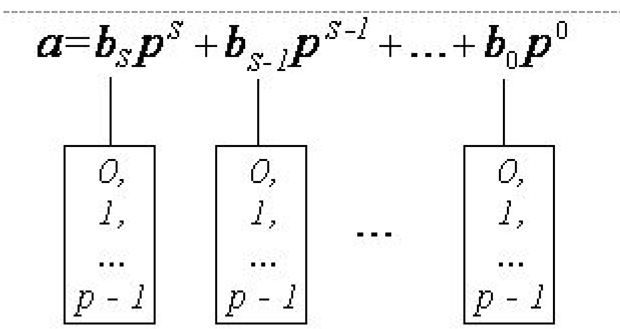

Рисунок - Запись числа в позиционной системе и соответствующее ему положение ячеек накопителя.

10. Возможное количество перестановок символов в сообщении и применение его в теории информации.

Последователи Шеннона использовали различные подходы к определению количественной меры информации. Так она вводится через логарифм отношения вероятностей, при помощи введения величины исчезнувшей неопределенности, 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 , применив аппарат комбинаторики. Рассмотрим этот подход более подробно.

Для хранения и передачи любого дискретного сообщения всегда используется некоторый (обычно конечный) алфавит $X = \{x_1, x_2, \ldots, x_k\}$. Величина k есть число элементов данного алфавита. В бинарном алфавите k=2, в десятичном 10, в кириллице 33, в латинице 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 букв можно составить максимум n! различных перестановок. 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 x1, x2 ...xk алфавита X. Если в сообщении содержится n1 элементов x1, n2 элементов x2,…, nk элементов xk и полная длина сообщения есть n=n1+n2+...nk, 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.*

11. Вывод формулы для энтропии дискретной случайной величины.

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

Понятие энтропии ввел Клод Шеннон в 1948 году

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

P - вероятность, xi - Дискретное значение

Логарифм может быть любого основания, но обычно используется 2, т.к. все железо у нас двоичное.

Если все буквы алфавита разные, то можно построить n! различных слов (еще это называется кол-во перестановок в комбинаторике), но если какие то буквочки повторяются, то количество перестановок считаем по этой формуле.

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

Где n - кол-во символов исходного алфавита, а то что в знаменателе - факториалы от кол-ва повторений букв в символе. Из этой формулы мы будем выводить далее энтропию.

N с шапочкой характеризует степень “разнообразия” сообщений, которые мы строим на основе алфавита.

Если переписать формулу в логарифмическом виде, то получим вот это:

$$\hat{H} = \log \hat{N} = \log\ n! - \sum_{i=1}^{k} \log\ n_i!$$

Здесь ничего особенного - логарифм от частного = разность логарифмов (св-во логарифма)

Она представляет собой число элементарных ячеек памяти, объем каждой из которых равен основанию используемого логарифма, необходимое и достаточное для хранения числа N С КРЫШЕЧКОЙ.

Дальше для вывода энтропии нам понадобится формула Стирлинга

$$\log\ n! \approx n \log n.$$

Это выражение стремится к равенству при n->∞.

Подставим в формулу H с крышечкой значение логарифма

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

Так как мы имеем дело чаще с частотами:

$$q_i = n_i / n,$$

Выразим ni и подставим в формулу (***) и получим формулу, ту что ниже:

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

При n->∞ частота qi стремится к вероятности Pxi - появления в сообщении элемента xi . 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

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

ЭТО ЭНТРОПИЯ

Она характеризует степень разнообразия значений, приходящегося в среднем на один символ при условии, что длина сообщения стремится к бесконечности.

Если основание логарифма = 2, то

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

*

* - минимальное среднее количество двоичных единиц (бит), приходящихся на один символ, необходимое и достаточное для того, чтобы безошибочно хранить или передавать любое сообщение из данного ансам**бля** при условии, что n велико.

Если что, то (-) спереди компенсируется тем, что log от числа, меньшего чем основание, будет отрицательный

12. Понятие неопределенности результата наблюдения случайного явления.

Для понимания (из конспекта):

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 опустить дальнейшие рассуждения, то степень неопределенности тем ВЫШЕ, чем более равновероятно распределяется некоторая величина.

Энтропия является функционалом распределения вероятностей (в общем случае)

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.

Пример из конспекта:

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

$$H_4(X) = 2$$

Здесь, как говорят, новое знание, позволило уменьшить (снять) начальную неопределенность.

13. Энтропия как числовая характеристика 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

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

$\mathrm{M}[\xi]$ — обозначение математического ожидания случайной величины

Сравнивая это выражение с формулой ниже:

$$H\ X\ =-\sum_{i=1}^{k}P\ x_i\ \log P\ x_i$$

Последнее можно записать как

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

где обозначена

$$I\ x_i\ =-\log P\ x_i$$

величина, осредняемая в предыдущей формуле

Эта величина называется ЧАСТНАЯ информация или СОБСТВЕННАЯ информация, она представляет собой количество информации, доставляемое самим событием x_i

14. Свойства энтропии.

1. Энтропия является вещественной и неотрицательной величиной, так как для любого i вероятность P(xi) изменяется в интервале от 0 до 1, logP(xi) отрицателен и, следовательно, каждое ее слагаемое -P(xi)logP(xi) неотрицательно. Следовательно H(X)>=0.

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*log(1)=0, а слагаемые, соответствующие остальным событиям, также равны

нулю: $\lim\limits_{P(x_i)\to 0}\left(-P\left(x_i\right)\log P\left(x_i\right)\right)=0$. Отсюда следует, что энтропия достоверного (неслучайного) события минимальна и равна нулю.

3. Энтропия максимальна, когда все элементы xi 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. В нашем случае такой функцией будет

$$H\left(P_1,P_2,\ldots,P_k\right)=-\sum_{i=1}^{k}P_i\ln P_i, \qquad (18)$$

где для краткости обозначено Pi=P(xi). Здесь выбран натуральный логарифм с целью упрощения решения поставленной задачи. Переменные Pi, от которых зависит энтропия, связаны известным условием

$$\sum_{i=1}^{k}P_i-1=0. \qquad (19)$$

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

$$\Phi\left(P_1,P_2,\ldots,P_k,\lambda\right)=H\left(P_1,P_2,\ldots,P_k\right)+\lambda\left(\sum_{i=1}^{k}P_i-1\right), \qquad (20)$$

состоящую из суммы выражения (18) и условия (19), умноженного на λ. Как известно, в методе Лагранжа нахождение условного экстремума функции (18) при условии (19) заменяется более простой задачей определения безусловного экстремума функции (20).

Необходимые условия экстремума функции $\Phi\left(P_1,P_2,\ldots,P_k,\lambda\right)$ находятся из системы уравнений $\partial\Phi/\partial P_i=0,\ \left(i=1,2,\ldots,k\right)$ и (19). Продифференцировав (20) по Pi с учетом (18), найдем, что экстремум обеспечивает Pi, удовлетворяющее равенству ln(Pi)=λ-1. Обратим внимание, что правая часть этого равенства не зависит от номера символа i. Следовательно, все Pi одинаковы и, согласно (19), равны 1/k независимо от основания логарифма в формуле (18). Соответствующая энтропия максимальна и равна.

$$H_{max}=\log k.$$

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

15. Энтропия системы 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 рамках одной системы, так и между состояниями, последовательно выбираемыми одним источником.

Определим энтропию объединения двух статистически связанных ансамблей $X = \{x_1, x_2, \dots, x_m\}$ и $Y = \{y_1, y_2, \dots, y_n\}$. *Объединение ансамблей* описывается в общем случае прямоугольной матрицей $\mathbf{P}(X,Y)$ вероятностей $P(x_i, y_j)$ всех возможных комбинаций состояний xi ансамбля X и состояний yi ансамбля Y:

$$\mathbf{P}(X,Y) = \begin{pmatrix} P(x_1,y_1) & \dots & P(x_1,y_j) & \dots & P(x_1,y_n) \\ \dots & \dots & \dots & \dots & \dots \\ P(x_i,y_1) & \dots & P(x_i,y_j) & \dots & P(x_i,y_n) \\ \dots & \dots & \dots & \dots & \dots \\ P(x_m,y_1) & \dots & P(x_m,y_j) & \dots & P(x_m,y_n) \end{pmatrix}. \qquad (21)$$

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

$$H(X) = -\sum_{i=1}^{m} P(x_i) \log P(x_i), \qquad (22)$$

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

$$P\left(x_i\right)=\sum_{j=1}^{n} P\left(x_i, y_j\right), \quad P\left(y_j\right)=\sum_{i=1}^{m} P\left(x_i, y_j\right). \qquad (23)$$

Совокупность случайных величин ξ и η представляет собой кортеж (упорядоченную двойку) $\varsigma=\left(\xi, \eta\right)$. 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 $P\left(x_i, y_j\right)$.

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 (22) можно определить *совместную энтропию* совокупности величин ξi и ηi

$$\begin{aligned} H\left(Z\right)=H\left(X, Y\right)&=-\sum_{i=1}^{m}\sum_{j=1}^{n} P\left(z_{ij}\right)\log P\left(z_{ij}\right)= \\ &=-\sum_{i=1}^{m}\sum_{j=1}^{n} P\left(x_i, y_j\right)\log P\left(x_i, y_j\right). \end{aligned} \qquad (24)$$

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

$$P\left(x_i, y_j\right)=P\left(x_i\right)P\left(y_j / x_i\right), \qquad (25)$$

где $P\left(y_j / x_i\right)$ - условная вероятность исхода yi при условии, что реализовался исход xi. Подставим (25) в (24) и выполним простые преобразования, учитывая, что логарифм произведения чисел равен сумме их логарифмов:

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

Учитывая, что для всех xi справедливо условие нормировки $\sum_{j=1}^{n}P(y_j/x_i)=1$. (т.е. при любом xi обязательно реализуется хотя бы одно yj), последнюю формулу преобразуем следующим образом:

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

где H(X) - энтропия случайной величины ξi. В полученное выражение входит так называемая *частная условная энтропия*

$$H(Y/x_i)=-\sum_{j=1}^{n}P(y_j/x_i)\log P(y_j/x_i). \quad (27)$$

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

Второй член в последнем выражении (26) есть результат осреднения $H(Y/x_i)$ по всем значениям X величины ξ и называется *средней условной энтропией* величины η при условии реализации любой из величин ξ. Обозначив

$$H(Y/X)=\sum_{i=1}^{m}H(Y/x_i)P(x_i), \quad (28)$$

окончательно получим

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

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 (25), формулу (28) с учетом (27) можно также преобразовать к виду

$$H\left(Y / X\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n} P\left(x_i, y_j\right) \log P\left(y_j / x_i\right) \qquad (30)$$

Из очевидного равенства $H\left(X,Y\right) = H\left(Y,X\right)$ получим, кроме (29), еще одно выражение для совместной энтропии

$$H\left(X,Y\right) = H\left(Y\right) + H\left(X / Y\right), \qquad (31)$$

где средняя условная энтропия

$$H\left(X / Y\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n} P\left(x_i, y_j\right) \log P\left(x_i / y_j\right). \qquad (32)$$

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

16. Энтропия бинарного источника сообщений.

Пусть алфавит источника информации имеет два символа: $X = \{x_1, x_2\}$. Такой источник называется *бинарным* (*двоичным*). Обозначим через p вероятность символа x_i, т.е. $P\left(x_1\right) = p$. Тогда вероятность второго символа $P\left(x_2\right) = 1 - p$. 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 *рисунке* 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 $p = 0$). 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.

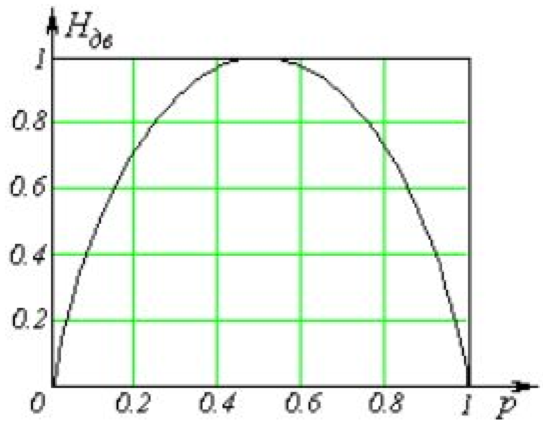

Рис.3.. - Энтропия двоичного источника

17 Графическое изображение соотношений между энтропиями систем случайных величин.

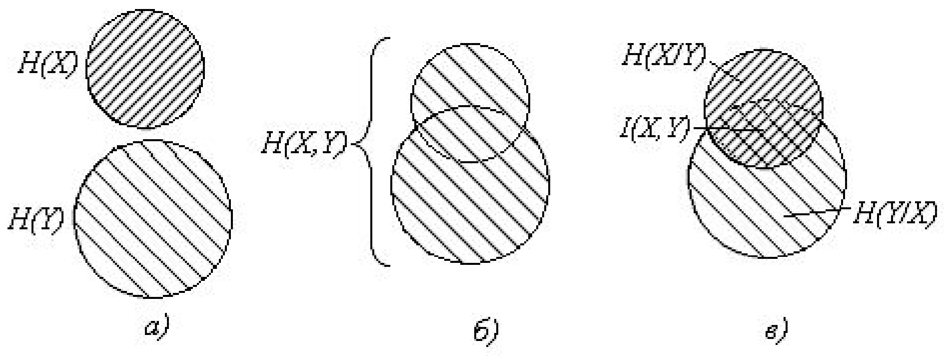

Рисунок 2. Графическое изображение соотношений между энтропиями

Уяснению соотношений между рассмотренными энтропиями дискретных источников информации способствует их графическое изображение в виде диаграмм, похожих на диаграммы Эйлера-Венна (рисунок 2). 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.

18. Прохождение дискретной случайной величины через канал связи.

На вход канала подается символ или дискретная случайная величина (или символ) , принимающая значения из множества X = {x1, x2, x3 … xN}, а на его выходе образуется также дискретная случайная величина, принимающая значения из множества Y = {y1, y2, y3 … yN} В канале связи обычно присутствуют шумы. Поэтому при поступлении на его вход величины xi (что можно также трактовать как осуществление некоторого события xi), на его выходе не обязательно будет зарегистрирована величина yi (т.е. событие yi) соответствующая именно данному xi. Отличие принятой величины от переданной характеризует ошибки канала связи.

Пусть величина xi имеет априорную вероятность P(xi). 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 {xi} и {yi} в точности соответствовали бы друг другу. Это означает, что при передаче величины xi на выходе канала была бы получена обязательно величина yi.

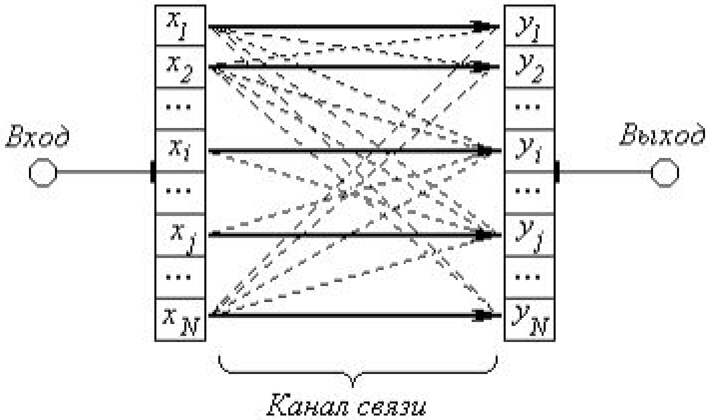

Рис.5. — Прохождение дискретной случайной величины через канал связи

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

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

Где P(xi,yj) - совместная двумерная вероятность осуществления событий xi на входе и yi на выходе канала.

19. Частная взаимная информация.

Введем математическое определение количества *взаимной информации*. Следуя К. Шеннону, определим количество *частной* информации $I\left(x_i;y_j\right)$ относительно события xi, доставляемое событием yi, как разницу между априорной информацией и апостериорной информацией, т.е. согласно формуле $I\left(x_i\right)=-\log P\left(x_i\right)$

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

Это можно представить как логарифм отношения апостериорной вероятности к априорной:

$$I\left(x_i;y_j\right)=\log\frac{P\left(x_i / y_j\right)}{P\left(x_i\right)}. \qquad (36)$$

Выбор основания логарифма определяет единицу измерения информации и осуществляется точно так же, как при определении информационной емкости накопителя.

20. Свойства частной взаимной информации.

Рассмотрим основные свойства частичной информации, следующие из ее определения.

1. Информация, содержащаяся в событии yi о событии xi, равна информации, содержащейся в событии xi о событии yi. Для док-ва учтем в выражении

$$I\left(x_i;y_j\right)=\log\frac{P\left(x_i / y_j\right)}{P\left(x_i\right)}$$ (частная взаимная инф.)

и формулу

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

И получим:

$$I\left(x_i;y_j\right)=\log\frac{P\left(x_i,y_j\right)}{P\left(x_i\right)P\left(y_j\right)} \qquad (*)$$

Отсюда видно, что мера количества информации симметрична относительно xi и yi т.е $I(x_i;y_j)=I(y_j;x_i)$. 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 (*) характеризует меру статистической связи между событиями xi и yi. Так, если события xi и yi статистически независимы, то $I(x_i;y_j)=0.$ При независимых событиях результат наблюдения или измерения yi не увеличивает знаний относительно xi

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

$$I(x_i;y_j)=\log\frac{P(x_i/y_j)}{P(x_i)}.$$ (частная взаимная инф.)

следует, что при фиксированной вероятности $P(x_i)$ взаимная информация $I(x_i;y_i)$ будет наибольшей, если $P(x_i/y_i)=1$ т.е когда отсутствуют ошибки перехода xi в yi и yi достоверно и однозначно определяет xi. Это максимальное значение

совпадает с $I(x_i)=-\log P(x_i)$ и называется *собственной информацией события* xi. Оно представляет количество информации, доставляемое самим событием xi или любым другим, однозначно связанным с xi. Это количество тем больше, чем меньше вероятность появления события xi.

Так как в общем случае $P\ x_i / y_j\ \leq 1$ то из формулы

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

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 (дезинформации) в противоположном случае.

4. Предположим, что величина ξ может с равной вероятностью принимать N возможных состояний. Очевидно, вероятность появления любого из ее состояний xi есть $P\left(x_i\right) = 1/N$. Тогда собственная информация события xi в соответствии с формулой

$I\left(x_i\right) = -\log P\left(x_i\right)$ равна $I\ x_i\ = \log N$. Но правая часть этой формулы равна правой части выражения $C = \log N$. 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.

21. Свойства энтропии 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={x1, x2, … xm} и Y = {y1, y2, … yn} *Объединение ансамблей* описывается в общем случае прямоугольной матрицей P(X,Y) вероятностей P(xi, yj) всех возможных комбинаций состояний xi, 1<= i <=m ансамбля X и состояний yj 1 <= j <= n ансамбля Y.

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

Энтропия одной дискретной случайной величины (или события), ξ равная:

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

Где m - число сотояний ξ

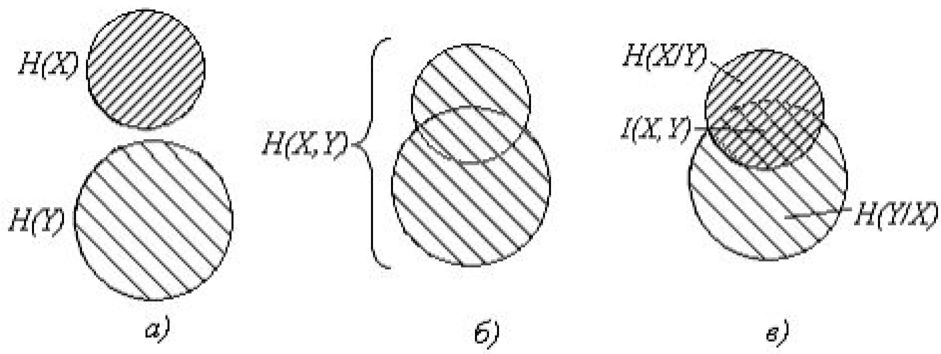

Рис. 4. - Графическое изображение соотношений между энтропиями

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) Действительно, при статистически независимых величинах условная вероятность равна безусловной вероятности: P(yj/xi)=P(yj), а совместная веро- ятность равна произведению вероятностей: P(xi, yj)=P(xi)P(yj). Тогда условная энтропия:

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

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 энтро- пия равна безусловной.

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(yj/xi) и P(xi/yj) равны нулю или единице. При этом имеют место равенства P(yj/xi)log(P(yj/xi))=0 и P(xi/yj)log(P(xi/yj))=0. Условные энтропии H(Y/X)=H(X/Y)=0 и H(Y/X)=H(X)=H(Y) Применительно к Рис.4 это означает, что такая ситуация возможна только в том случае, когда круги H(X) H(Y) имеют одинаковую величину и совмещены друг с другом.

3 Условная энтропия может изменяться в пределах

0 <= H(Y/X)<=H(Y)

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

4 Для совместной энтропии всегда справедливо соотношение H(X/Y)<= H(X)+H(Y)

Это свойство совместной энтропии является непосредственным следствием предыдущего соотношения.

30. Вычисление взаимной информации гауссовских случайных величины x и h=x+n , где n - не коррелированный с x шум.