Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

Санкт-Петербургский национальный исследовательский университет

информационных технологий, механики и оптики

Факультет Компьютерных технологий и управления
Кафедра ПБКС

Конспект лекций по дисциплине

"Теория информации"

XeTeX

Санкт-Петербург, 2013

Содержание

Теория вероятности

Основные сведения из теории вероятности

Введение

Термин **информация** в курсе будет пониматься в узком научном смысле.

Теория информации – специальная математическая дисциплина. Её содержанием является абстрактно формулируемые теоремы и модели. ТИ имеет обширное применение к теории передачи сообщений, записывающих устройств, матлингвистике, компьютерной технике.

В самом общем виде теория информации понимается как теория передачи сигналов по линиям связи. Наиболее важное понятие ТИ –- сама информация. В нашей жизни большую роль играет информация и связанные с ней операции: передача, получение, обработка, хранение.

Информация имеет две стороны: количественную и качественную. Иногда важно получение общего количества информации (количественная сторона), иногда важно конкретное содержание самой ИИ. Отметим, что переработка ИИ является технически сложной процедурой, которая усложняет разработку общей теории информации.

Важнейшим этапом в открытии основных закономерностей ТИ были работы американского инженера-связиста, математика Клода Шеннона (1947-49гг).

Для вычисления количества информации была предложена т.н. **логарифмическая мера**. Понятие **количества информации** тесно связано с понятием энтропии как меры степени неопределённости. Приобретение информации сопровождается уменьшением неопределённости, следовательно, количество информации можно измерять количеством ''исчезнувшей неопределённости'' (энтропии).

Теория информации является математической теорией, использующей понятия и методы теории вероятности.

Вероятность. Случайные события и величины

Пусть производится серия из N опытов, причём некоторое событие A происходит в $N_a < N + 1$. Тогда $h_n(A) = N_a/N$ называется частотой появления события A в серии из N опытов. Известный факт: с ростом $N \quad h_n(A) \to p$ (постоянная p - вероятность появления случайного события A).

Наука, изучающая свойства вероятности и применение этого понятия называется **теория вероятности**.

Событие, которое при выполнении некоторого комплекса условий обязательно выполняется называется **достоверным событием**.

Событие, которое при выполнении некоторого комплекса условий не выполняется называется **невозможным событием**.

Пример: Выпадение определённого числа очков на грани игральной кости - достоверное событие.

Выпадение семи очков на грани игральной кости - невозможное событие.

Случайное событие – событие, которое может произойти, а может и не произойти.

Задача: В урне 10 шаров : 5 белых, 3 чёрных и 2 красных. Найти вероятность выпадения шара определённого цвета (шары одинаковы).

Решение: Выписать случайные события:

A –- {вынутый шар белый} $P(A) = 5/10 = 1/2$;
B –- {вынутый шар чёрный} $P(B) = 3/10$;
C –- {вынутый шар красный} $P(C) = 2/10 = 1/5$.

Задача: Какова вероятность, что при бросании кости выпадет число очков, кратное 3?

Решение:

Кратны 3 $\{3, 6\}$.
N исходов $= 6$.
$P(A) = 2/6 = 1/3$.

Общий принцип решения задач сводится к понятию равновероятности или равновозможности. (например, все грани кости одинаковы, и вероятность выпадения той или иной грани равна $1/6$).

Классическое определение вероятности

Пусть из N возможных исходов опыта случайное событие A появляется M раз. Тогда вероятность случайного события A в модели с равновероятными исходами вычисляется по формуле $P(A) = M/N$

Каждому опыту отвечает своя таблица вероятности. К примеру, в задаче с урнами и шарами таблица вероятности имеет вид:

События	A	B	C
Вероятности	$P(A) = 1/2$	$P(B) = 3/10$	$P(C) = 1/5$

Можно сказать, что в опыте с бросанием кости число очков, выпадающих на грани является случайной величиной, которая может принимать одно из возможных 6 числовых значений в зависимости от случая.

Итак, случайная величина –- числовая функция, принимающая то или иное числовое значение в зависимости от случая.

Например, количество рождений в городе за год - случайная величина.

Свойства вероятности. Сложение и умножение случайных событий.

Несовместные и независимые случайные события.

Из определения вероятности вытекают основные свойства вероятности случайного события A:

- $0 \leqslant P(A) \leqslant 1$.

  $P(A) = 1$ – достоверное событие;

  $P(A) = 0$ – невозможное событие.

- Пусть опыт приводит к двум взаимоисключающим событиям или исходам A или B. В этом случае B называют противоположным A событием $(B = \bar{A})$

  Пусть $P(A) = \frac{m}{n}$;

  Тогда $P(\bar{A}) = \frac{(n-m)}{n} = 1 - \frac{m}{n} = 1 - P(A) \Rightarrow P(A) = 1 - P(\bar{A})$.

  Пусть случайное событие $A_1 \subset A$ влечёт появление события $A \Rightarrow P(A1) < P(A)$

- **Правило сложения вероятностей для двух событий:**

  - Пусть A и B – несовместны.

    Тогда $A \cap B = \varnothing$;

    $P(A) = \frac{m_1}{n}$;

    $P(B) = \frac{m_2}{n}$;

    $P(A + B) = \frac{(m_1+m_2)}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$.

    Таким образом, $P(A + B) = P(A) + P(B)$.

    В примере с урной вероятность извлечь чёрный или белый шар равна

    $P(A + B) = P(A) + P(B) = \frac{1}{2} + \frac{3}{10} = \frac{4}{5}$;

    Замечание: Пусть некоторый опыт проиводит к появлению K различных (взаимоисключающих) исходов:

Исходы	A1	A2	…	An
Вероятности	P1	P2	…	Pn

    Заметим, что бывают случаи, когда

$$\sum_{i=1}^{k} P(A_i) = P(A_1 + A_2 + \ \ldots \ + A_n) = 1$$

    В этом случае говорят, что события $A_1, A_2, \ldots, A_n$ составляют **полную группу** случайных событий, то есть $A_1, A_2, \ldots, A_n$ попарно несовместны.

    $A_1, A_2, \ldots, A_n : A_i \cap A_j = \varnothing \, \forall \, i, j : i \neq j$; если $A_1 + A_2 + \ \ldots \ + A_n$ - достоверное событие.

  - Пусть A и B совместны. $P(A+B) = P(A)+P(B)-P(AB)$, где $P(AB)$ – вероятность одновременного происхождения двух случайных событий A и B.

Теорема сложения вероятности для совместных случайных событий.

(диаграмма Венна: $A = m_1; B = m_2, A \cap B = l$).

$P(AB) = \frac{l}{n}$

$P(A+B) = \frac{(m_1+m_2-l)}{n} =$ (в m_1 и m_2 входит l) $= \frac{m_1}{n} + \frac{m_2}{n} - \frac{l}{n}$

События A и B называются **независимыми**, если результат выполнения события A не связан с результатом события B. (извлечение двух чёрных шаров из разных урн – независимые события)

- **Теорема умножения вероятности для двух независимых событий:**

  Если A и B независимы, то $P(AB) = P(A) * P(B)$

Пример 1: Какова вероятность при двух бросках монеты оба раза выпадет орёл?

$P(AB) = ?$
A\{орёл\} $P(A) = \frac{1}{2}$;
B\{решка\} $P(B) = \frac{1}{2}$;
$P(AB) = \frac{1}{4}$.

Пример 2: В колоде 52 карты, 4 масти, 2 козыря. Какова вероятность того, что взятая наугад карта 2 является тузом или козырем?

A\{туз\} $P(A) = 1/13$;
B\{козырь\} $P(B) = 1/4$;
$P(AB) = 1/52$;
A и B совместны, независимы.
$P(A+B) = P(A) + P(B) - P(AB) = 1/13 + 1/4 - 1/52 = 4/13$.

Условная вероятность.

Рассмотрим пример: В урне M чёрных шаров и $N - M$ белых. Случайное событие
A {извлечение чёрного шара} и
B {извлечение чёрного шара из той-же урны после того, как из неё уже вынут один шар}

$$P(B|A) = \frac{(m-1)}{(n-1)}$$

Поскольку, если событие A имело место, то в урне осталось $M - 1$ чёрных шаров.

$$P(B|\bar{A}) = \frac{m}{(n-1)}$$

$\bar{A}$: \{первый вынутый шар - белый\}

Вероятность события B здесь разная. Вероятность, которую имеет событие B в том, случае, когда известно, что событие A имело место называется **условной вероятностью** события B при условии выполнения события A.

$$P(B/A) = P(B|A) = P_A(B)$$

Условные вероятности можно вычислять аналогично вычислению безусловных вероятностей.

В случае если A и B независимы, $P(A|B) = P(A) * P(B)$.

В случае зависимости $P(AB) = P(A) * P(B|A) = P(B) * P(A|B)$.

В обоих случаях мы имеем правило умножения вероятностей. В одном случае для независимых событий, в другом для зависимых. Последнее соотношение часто кладут в определение условной вероятности.

$$P(B|A) = \frac{P(AB)}{P(A)} \qquad P(A|B) = \frac{P(AB)}{P(B)}$$

Из предыдущей формулы можем составить пропорцию:

$$\frac{P(B|A)}{P(B)} = \frac{P(A|B)}{P(A)}$$

Из определения условной вероятности вытекают ее основные свойства:

1. $0 \leqslant P(B|A) \leqslant 1$, причём $P(B|A) = 1$ когда $A \subset B$; B - достоверное случайное событие.

   $P(B|A) = 0 \Longleftrightarrow A, B$ несовместны, или известно, что B – невозможное событие.

2. Пусть $B_1 \subset B$ (появление B_1 вызывает событие B). $P(B1|A) \leqslant P(B|A)$.

3. Если B и C несовместны $P(B + C|A) = P(B|A) + P(C|A)$ (теорема сложения вероятностей для несовместных событий)

4. $P(\bar{B}|A) = 1 - P(B|A)$

Замечание: Пусть имеется K (и только K) попарно несовместных исходов некоторого опыта $A_1, A_2, \ldots, A_k$, называемых гипотезами. Пусть некоторое случайное событие B может произойти при выполнении одной из гипотез. Тогда очевидно, что $B = A_1B + A_2B + \ldots + A_kB$ (все события A_iB несовместны, поэтому можно воспользоваться теоремой сложения вероятностей)

$$P(B) = P\left(\sum_{i=1}^{k} A_iB\right) = \sum_{i=1}^{k} P(A_iB) = \sum_{i=1}^{k} \big(P(A_i) * P(B|A_i)\big)$$

Формула носит название формулы полной вероятности

$$P(B) = \sum_{i=1}^{k} P(A_i) * P(B|A_i)$$

Задача: Имеется 5 урн : в двух по одному белому и пять чёрных шаров; в одной – 2 белых, 5 чёрных; в двух – 3 белых, 5 чёрных шаров. Наудачу выбирается одна урна. Из неё извлекается один шар. Какова вероятность того, что шар белый?

Решение: Выберем в качестве гипотез 3 способа

A_1 : {Выбрана урна с 1 б.ш} $P(A_1) = 2/5$ $P(B|A_1) = 1/6$
A_2 : {Выбрана урна с 2 б.ш} $P(A_2) = 1/5$ $P(B|A_2) = 2/7$
A_3 : {Выбрана урна с 3 б.ш} $P(A_3) = 2/5$ $P(B|A_3) = 3/8$
B = извлечён белый шар
$P(B) = \frac{1}{6} * \frac{2}{5} + \frac{2}{7} * \frac{1}{5} + \frac{3}{8} * \frac{2}{5} = \frac{23}{84}$

Математическое ожидание случайной величины и его основные свойства

Введение.

Важнейшей числовой характеристикой ξ является её математическое ожидание или среднее значение, вычисляемое по правилу $M\xi = \sum_{i=1}^{n} x_i p_i)$, где x_i – принимаемые ξ значения, p_i – вероятности их выпадения.

С помощью математического ожидания мы можем сравнивать между собой две случайные величины (например, из двух стрелков лучший тот, кто выбивает в среднем наибольшее число очков), однако встречаются задачи, в которых знание одного лишь $M\xi$ недостаточно.

Пример: Пушка ведёт прицельный огонь по мишени, удалённой от пушки на расстояние a. Обозначим дальность полёта снаряда через ξ километров; $M\xi = a$

Отклонение $M\xi$ от a свидетельствует о наличии систематической ошибки (производственный дефект, неправильный угол наклона). Ликвидация систематической ошибки достигается изменением угла наклона орудия.

Вместе с тем, отсутствие систематической ошибки ещё не гарантирует высокую точность стрельбы. Чтобы оценить точность надо знать, насколько близко ложатся снаряды к цели.

Как определить точность стрельбы и сравнить между собой качество стрельбы двух орудий?

Отклонение снаряда от цели - $\xi - a$

$M(\xi - a) = M\xi - a = a - a = 0$

В среднем, положительные и отрицательные значения $M\xi$ сокращаются. Поэтому принято характеризовать разброс значений случайной величины математическим ожиданием квадрата её отклонения от своего математического ожидания. Полученное таким образом число называется дисперсией случайной величины ξ.

$D\xi = M(\xi - a)^2 = M[\xi - M\xi]^2$

Ясно, что в случае орудий, ведущих стрельбу, лучшим следует считать орудие, у которого $D\xi$ будет наименьшей.

Пусть ξ характеризуется таблицей вероятностей

$x_i:$	x_1	x_2	$\dots$	x_n
$p_i:$	p_1	p_2	$\dots$	p_n

$$M\xi = \sum_{i=1}^{n} x_i p_i; \qquad D\xi = M(\xi - M\xi)^2 = \sum_{i=1}^{n} (x_i - M\xi)^2 * p_i$$

Определение математического ожидания

Пусть есть некоторое пространство, в котором имеется некоторое $\xi = \xi(\omega_i)$.

ω_i – неразделимое событие (пример: исходы броска монеты); $\omega_i : (i = 1, \bar{n})$.

Совокупность ω_i образует пространство элементарных событий $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$

Математическим ожиданием случайной величины ξ называется число, обозначаемое $M\xi$ и равное

$$M\xi = \sum_{\omega_i \in \Omega} \{(\omega_i) * P(\omega_i)\} = \sum_{i=1}^{n} \xi(\omega_i) * p(\omega_i),$$ где p_i - элементарные вероятности.

Из определения математического ожидания вытекают следующие свойства:

1. Аддитивность. $M(\xi + \eta) = M\xi + M\eta$.

   Следствие $M\left(\sum_{k=1}^{n} \xi_k\right) = \sum_{k=1}^{n} (M\xi_k)$.

2. $\forall C = const : M(C * \xi) = C * M\xi$. Совокупность свойств 1 и 2 даёт нам свойство линейности математического ожидания:

   $$M(C_1\xi_1 + C_2\xi_2 + \dots + C_n\xi_n) = C_1M(\xi_1) + C_2M(\xi_2) + \dots + C_nM(\xi_n)$$

3. Математическое ожидание индикатора случайного события равно вероятности этого случайного события.

   Индикатор $[\chi]$: $M\chi_A(\omega) = P(A)$ - случайная величина, принимающая 2 значения: $\chi_A(\omega) = \{1, \omega \in A \,|\, 0, \omega \notin A\}$

   $$\sum_{\omega \in A} P(\omega) = P(A)$$

   $$M\chi_A(\omega) = \sum_{\omega \in A} 1 * p(\omega) + \sum_{\omega \notin A} 0 * p(\omega) = \sum_{\omega \in A} 1 * p(\omega) = P(A).$$

4. Свойство монотонности $\xi \geqslant \eta \Rightarrow M\xi \geqslant M\eta$.

Докажем вначале, что имеет место следующее свойство $\xi \geqslant 0 \Rightarrow M\xi \geqslant 0$ (при разложении по определению неотрицательны).

$$M\xi = \sum_{\omega} \xi(\omega)p(\omega) \geqslant 0.$$

Применим полученное свойство:

$$\xi - \eta \geqslant 0 \Rightarrow M(\xi - \eta) \geqslant 0 \Rightarrow M\xi - M\eta \geqslant 0 \Rightarrow M\xi \geqslant M\eta.$$

Формулы вычисления математического ожидания

Пусть $x_1, x_2, \ldots, x_n$ –- значения случайной величины ξ, принимаемые с вероятностями $p_1, \ldots, p_i$. Тогда имеет место следующая формула для вычисления математического ожидания :

$$M\xi = \sum_{i=1}^{n} x_i * P(\xi = x_i)$$

Чтобы доказать формулу будем исходить из того, что ξ может быть представлена в виде линейной комбинации индикаторов случайных событий

$$\xi = \sum_{i=1}^{n} x_i * \chi_{A_i}(\omega)$$
$$A_i\{\omega_i : \xi = x_i\}$$

Левые и правые части соотношения совпадают. Применим к написанному равенству операцию математического ожидания:

$$M\left(\sum_{i=1}^{n} x_i\chi_{A_i}(\omega)\right) = \sum_{i=1}^{n} M\left(x_i\chi_{A_i}(\omega)\right) = \sum_{i=1}^{n} x_i M\left(\chi_{A_i}(\omega)\right) = \sum_{i=1}^{n} x_i P(\xi = x_i)$$

Рассуждая аналогично, нетрудно получить формулы вычисления математического ожидания от величин, представляющих собой функции случайных величин.

Пусть заданы $f(\xi), g(\xi, \eta)$.

В этом случае

$$M(f(\xi)) = \sum_{i=1}^{n} (f(x_i) * P(\xi = x_i))$$

$$M(g(\xi, \eta)) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} g(x_i, y_j) * P\left(\xi = x_i,\, \eta = y_i\right)\right)$$

Здесь $P(\xi, \eta)$ – совместная вероятность.

5 Мультипликативное свойство математического ожидания

Пусть ξ, η - независимые случайные величины, то $M(\xi, \eta) = M\xi * M\eta$

Доказательство:

$$M(\xi, \eta) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} x_i * y_j * P\left(\xi = x_i, \eta = y_j\right) \right)$$

Если ξ, η независимы, то для них применима теорема умножения вероятности.

$$P(\xi = x_i, \eta = y_j) = (\xi, \eta \text{ независимы}) = P(\xi = x_i) * P(\eta = y_j)$$

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{m} x_i * y_j * P(\xi = x_i) * P(\eta = y_j) \right) =$$

$$= \sum_{i=1}^{n} x_i * P(\xi = x_i) \ * \sum_{j=1}^{m} y_i * P(\eta = y_j) = M\xi * M\eta$$

Замечание: Все написанные формулы имеют место, если вероятностное пространство конечно, т.е. число элементарных событий конечно $\omega_i = (1, \bar{n})$.

В случае, если вероятностное пространство счётно, количество элементарных сообщений бесконечно, тогда для случайной величины $\xi(\omega), \omega \in$ (счетное вероятностное пространство) имеют место следующие формулы:

$$\omega_i, i = \overline{[1, \infty]}$$

$$M\xi = \sum_{i=1}^{\infty} (x_i * P(\xi = x_i))$$

$$Mf(\xi) = \sum_{i=1}^{\infty} (f(xi) * P(\xi = x_i))$$

В формулах справа стоят ряды. Чтобы математические ожидания существовали надо, чтобы эти ряды сходились. Ряд сходится, если он имеет конечную сумму.

Задача: Вычислить $M\xi$, распределённой по закону Пуассона. $P(\xi = k) = (a^k/k!)e^{-a}$, где $k = \{0, 1, 2, 3, 4, \ldots, \infty\};\ a > 0$ – заданный заранее характер распределения.

Решение:

$$M\xi = \sum_{k=0}^{\infty} k * \frac{(a^k * e^{-a})}{k!} = e^{-a} \sum_{k=0}^{\infty} k * \frac{(ka^k)}{k!} =$$

$$= e^{-a}\sum_{k=0}^{\infty}\frac{(k*a^{k-1}a)}{(k-1)!} = e^{-a}a\sum_{s=0}^{\infty}\frac{a^s}{s!} \quad \text{(формула Маклорена)} = e^{-a}ae^a = a$$

Математическое ожидание случайной величины, распределённой по закону Пуассона с параметром распределения a равно этому параметру распределения.

Если ξ непрерывна, её закон распределения определяется плотностью распределения $f_\xi(x) \geqslant 0 \Rightarrow M\xi = \int\limits_{-\infty}^{\infty} xf_\xi(x)dx$. Если имеется функция $g(\xi),\ g(\xi,\eta)$, то математическое ожидание вычисляется по формулам:

$$M_{g_\xi} = \int\limits_{-\infty}^{\infty} g(x)f_\xi(x)dx$$

$$M(\xi,\eta) = \iint\limits_{-\infty}^{\infty} g(x,y)f_{\xi\eta}(x,y)dx\,dy$$

где $f(\xi,\eta)$ - плотность совместных случайных величин.

Эти математические ожидания существуют, если все написанные несобственные интегралы сходятся.

Пример: вычислить математическое ожидание ξ, равномерно распределённое по закону Пуассона

Решение:

$$f_\xi(x) = \frac{1}{b-a}, a \leqslant x \leqslant b \mid 0, x \in \text{в остальных случаях}$$

$$M_\xi = \int\limits_{-\infty}^{\infty} xf_\xi(x)dx = \frac{1}{b-a}\int\limits_a^b xdx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}$$

Пример 2: вычислить математическое ожидание случайной величины ξ, распределённой нормально (по закону распределения Гаусса)

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}(x-a)e^{-\frac{(x-a)^2}{2\sigma^2}}dx +$$

$$+\frac{a}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}(x-a)e^{-\frac{(x-a)^2}{2\sigma^2}}dx \quad \text{(интеграл Лапласа)} = 0 + \frac{a}{\sqrt{2\pi}\sigma}\sqrt{2\pi}\sigma = a$$

Вывод: Распределение случайной величины, распределённой нормально, равно параметру распределения.

Дисперсия случайной величины и её основные свойства.

Дисперсия $D\xi$ - число, определяемое формулой $D\xi = M(\xi - M\xi)^2$ (1), т.е. дисперсия представляет собой квадрат разности случайной величины и её математического ожидания. Другое название - квадрат среднеквадратического отклонения.

Часто в прикладных задачах вместо D рассматривают величину $\sqrt{D}$, называемую среднеквадратическим отклонением

Формулу (1) можно продолжить, тогда мы получим

$$D_\xi = M\left(\xi^2 - 2\xi M\xi + (M\xi)^2\right) = M^2\xi - 2M\xi + 2M\xi + (M\xi)^2 = M^2\xi - (M_\xi)^2,$$

откуда $(2) \quad D\xi = M^2\xi + (M\xi)^2$

1. Пусть ξ - дискретная величина, принимающая значения $x_1, \ldots, x_n$ с вероятностями $p_1, \ldots, p_n$

$$D\xi = \sum_{k=1}^{n}(x_k M\xi)^2 * p_k = (2) = \sum_{k=1}^{n}(x_k^2 * p_k)^2 - (M\xi)^2$$

2. Пусть ξ - непрерывная случайная величина, значит может быть определена функция $f_\xi(x)$.

$$D\xi = (1) = \int_{-\infty}^{\infty} x^2 f_\xi(x)dx - (M\xi)^2$$

Дадим механическую интерпретацию математического ожидания и дисперсии случайной величины. Будем представлять закон распределения вероятностей $p_k = P(\xi = x_k), \sum_{k=1}^{n} p_k = 1$ случайной величины ξ, как закон распределения единичной массы на прямой: в точках x_k сосредоточены массы p_k :

$$---\frac{x_1}{p_1}---\frac{x_2}{p_2}-\cdots-\frac{x_n}{x_n}--->x$$

Тогда

$M\xi = \sum_{k=1}^{n} x_k P(\xi = x_k)$ - центр тяжести СМАТ

$D\xi = \sum_{k=1}^{n}(x_k - M_\xi)^2 * p_k$ - момент инерции относительно начала координат

Пример: $D\xi = ?,\ f_\xi(x) = e^{-\frac{(x-a)^2}{2\sigma^2}}$

Решение:

$$D\xi = \int\limits_{-\infty}^{\infty} (x - M\xi)^2 f_\xi(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$

Произведём замену переменной по формуле $y = \frac{x-a}{\sigma},\ x - a = \sigma y,\ dx = \sigma dy$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} \sigma^2 e^{-\frac{y^2}{2}} \sigma dy = \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (-y) de^{-\frac{y^2}{2}} = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ye^{-\frac{y^2}{2}} \Big|_{-\infty}^{+\infty} + \int\limits_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) = \sigma^2$$

Вывод: дисперсия нормального распределения случайной величины равна второму параметру распределения (σ^2): $M_\xi = a;\ D_\xi = \sigma^2$

Свойства дисперсии

1. Дисперсия неотрицательна: $D_\xi \geqslant 0$. $D_\xi = 0 \Longleftrightarrow \xi = const$.

   Доказательство: $D\xi = M(\xi - M\xi)^2 \geqslant 0-$ по свойству монотонности.

   Пусть $\xi = c = const$.

   Тогда $D_c = M(c - M_c)^2 = (c-c)^2 = 0$.

2. Если $a = const$, то дисперсия $D(a\xi) = a^2 D\xi$.

   Доказательство:

   $D(a\xi) = M(a\xi - Ma\xi)^2 = M(a\xi - aM\xi)^2 = M[a^2(\xi - M\xi)^2] = a^2 M(\xi - M\xi) = a^2 D\xi$

3. Если ξ, η независимы, то

$$D(\xi + \eta) = M\left(\xi + \eta - M(\xi + \eta)\right)^2 = M\left((\xi - M\xi) + (\eta - M\eta)\right)^2 =$$
$$= M(\xi - M\xi)^2 + 2(\xi - M\xi)(\eta - M\eta) + M(\eta - M\eta)^2 =$$
$$= M(\xi - M\xi)^2 + 2M\left((\xi - M\xi) * (\eta - M\eta)\right) + M(\eta - M\eta)^2 = D\xi + D_\eta$$

Энтропия и информация

Энтропия как мера неопределённости

Для практики важно уметь численно оценивать степень неопределённости самых разнообразных опытов, чтобы иметь возможность их сравнивать.

Начнём с рассмотрения опытов имеющих K равновероятных исходов. Степень неопределённости каждого такого опыта определяется числом K. При $K = 1$ исход опыта не является случайным. При большом значении K предсказание результата опыта становится затруднительным.

Таким образом, искомая численная характеристика степени неопределённости должна зависть от K, т.е быть функцией $f(k);\ f(1) = 0$; при возрастании аргумента, функция должна возрастать. Для более полного определения функции $f(k)$ необходимо предъявить к ней дополнительные требования.

Рассмотрим сложный опыт $\alpha\beta$, состоящий в одновременном выполнении опытов α и β. Неопределённость выполнения сложного опыта больше неопределённости опыта α, т.к. к его неопределённости надо добавить неопределённость опыта β. Поэтому естественно считать, что **степень неопределённости** опыта $\alpha\beta$ равна сумме неопределённостей, характеризующих α и β.

Пусть $\alpha\beta$ имеет $k * l$ равновероятных исходов, $k\alpha,\ l\beta$. Приходим к следующему условию, которму должна удовлетворять функция $f(kl) = f(k) + f(l)$. Последнее условие наталкивает на мысль принять за меру неопределённости опыта, имеющего К равновероятных исходов число $\log k$: $\log(kl) = \log k + \log l$. Такое определение меры неопределённости согласуется с первоначальными условиями, что $f(1) = \log 1 = 0; f(k)$ - возрастающая функция. Можно доказать, что логарифмическая функция является единственной, удовлетворяющей этим условиям.

Замечание: отметим, что выбор основания логарифма большой роли не играет, поскольку в силу известной формулы перехода можем написать $\log_b a = \log_c a / \log_c b \Rightarrow \log_b k = \log_b a * log_a k$ сводится к домножению на константу, т.е. равносилен простому изменению **единицы измерения** степени неопределённости. Обычно за меру степени неопределённости берут логарифмы при основании 2: $log_2 k = log k$, причём основание 2 не фиксируют. Т.е. за единицу измерения степени неопределённости принимают неопределённость опыта, имеющего 2 равновероятных исхода: $\log_2 2 = 1$ бит. Везде далее будем пользоваться двоичными единицами измерения.

Таблица вероятности для опыта, имеющего K равновероятных исходов:

α				
Исходы	A_1	A_2	$\dots$	A_k
Вероятности	$\frac{1}{k}$	$\frac{1}{k}$	$\dots$	$\frac{1}{k}$

Поскольку при наших допущениях неопределённость равна $f(k) = \log k$. В этом случае каждый отдельный исход вносит неопределённость $\frac{1}{k}$. $\frac{\log k}{k} = \frac{1}{k} \log k = -\frac{1}{k} \log \frac{1}{k}$.

В самом общем случае опыт имеет следующую таблицу вероятности:

α				
Исходы	A_1	A_2	$\dots$	A_k
Вероятности	$P(A_1)$	$P(A_2)$	$\dots$	$P(A_k)$

Для опыта общая мера неопределённости равна $-p(A_1) \log p(A_1) - p(A_2) \log p(A_2) - \dots - p(A_k) \log p(A_k) = H(\alpha)$ - энтропия опыта α

Рассмотрим некоторые свойства энтропии $H(\alpha)$:

1. $H(\alpha) \geqslant 0$

   Доказательство:

   $-p(A) \log p(A) \geqslant 0$ (множители $\in$ промежутку $(0 \leqslant p(A) \leqslant 1)$)

   $-p(A) \log p(A) = 0 \iff \{p = 0; p = 1\}$

В случае, если опыт имеет K попарно несовместных исходов, то $H(\alpha) = 0$ равносильно тому, что один исход - достоверное событие, а все другие - невозможны, так как $\big(p(A_1) + \ldots + p(A_k) = 1\big)$. Это обстоятельство хорошо согласуются с величиной $H(\alpha)$ - только в этом случае опыт вообще не содержит неопределённости.

2. Из всех опытов с K исходами самым неопределённым является опыт опыт с K равновероятными исходами. Можно показать, что имеет место неравенство

   $H(\alpha) = -p(A_1)\log p(A_1) - \ldots - p(A_k)\log p(A_k) \leqslant H(\alpha_0)$

   $H(\alpha_0) = \log k = -\frac{1}{k} - \ldots - \frac{1}{k}$.

   Равенство достигается при равных вероятностях $P(A_i);\ i = \overline{[1,k]}$

Пример: Имеется две урны с 20-ю шарами каждая. Первая - 10 белых, 5 чёрных, 5 красных. Вторая - 8 белых, 8 чёрных, 4 красных.

Из каждой урну вынимают по 1 шару. Исход какого из двух опытов следует считать более неопределённым?

Решение: Обозначим опыты как А1 и А2.

А1

Исходы	Бел	Чёр	Крас
Вероятности	1/2	1/4	1/4

А2

Исходы	Бел	Чёр	Крас
Вероятность	2/5	2/5	1/5

Энтропия опыта А1: $H(\alpha_1) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = -\frac{1}{2} * 1 - \frac{1}{2} * (-2) = -\frac{1}{2} + 1 = 1,5$бита.

Энтропия опыта А2: $H(\alpha_2) = -\frac{2}{5}\log\frac{2}{5} - \frac{2}{5}\log\frac{2}{5} - \frac{1}{5}\log\frac{1}{5} = -\frac{4}{5}(\log 2 - \log 5) - \frac{1}{5}(\log 1 - \log 5) = -0.8 + -\frac{4}{5}\log 5 + \frac{1}{5}\log 5 = -0.8 + \log 5 = 1,52$ бита.

Вывод: Если оценивать степень неопределённости опыта его энтропией, то исход второго опыта более неопределённый, нежели первого.

Историческая справка

Исторически первые шаги к введению понятия энтропии были сделаны в 1928 году американским инженером-связистом Хартли, предложившим характеризовать степень неопределённости опыта с К различными исходами числом $\log k$. Предложенная им мера степени неопределённости иногда бывает удобна в некоторых практических задачах, но часто оказывается малопоказательной, поскольку полностью игнорирует различие между характером имеющихся исходов. Поэтому почти невероятному исходу у Хартли придаётся такое-же значение, как и исходу весьма вероятному. Однако, он считал, что различия между отдельными исходами определяются в первую очередь "психологическими факторами" и должны учитываться лишь психологами, но не инженерами или математиками.

Ошибочность точки зрения Хартли была показана другим американским инженером - математиком К. Шенноном. Он предложил принять в качестве меры неопределённости опыта с K различными исходами $A_1, \ldots, A_k$ величину

$H(\alpha) = -p(A_1) \log p(A_1) - \ldots - p(A_k \log p(A_k)$.

Иначе говоря, исходу A_i следует приписать неопределённость, равную $-\log p(A_i)$. В качестве неопределённости всего опыта $H(\alpha)$ принимается среднее значение случайной величины (математическое ожидание), равное $H(\alpha)\xi$,где ξ принимают значения $-\log p(A_i)$ с вероятностями $p(A_i)$.

Таким образом, загадочные "психологические факторы" учитываются с помощью использования понятия вероятности, имеющего чисто математический, а точнее статистический характер.

Использование величины $H(\alpha)$ в качестве меры неопределённости опыта A оказалось полезным во многих областях, а особенно в теории передачи сообщений по линиям связи.

Энтропия сложных событий. Условная энтропия

Условная энтропия. Пусть имеются два независимых опыта A, B с таблицей вероятностей $A_1, p(A_1); \ldots; A_k, p(A_k); \quad B_1, p(B_1); \ldots; B_l, p(B_l)$.

Рассмотрим сложный опыт $\alpha\beta$, когда осуществляются оба опыта одновременно, имеющий $k * l$ исходов ($A \times B$ - декартово произведение).

$A_1B_1 : \alpha = A_1; \beta = B_1$

Очевидно, что неопределённость опыта $\alpha\beta$ больше неопределённости каждого из опытов, из-за осуществления обоих опытов. Поэтому имеет место соотношение $H(\alpha\beta) = H(\alpha) + H(\beta)$. Написанное равенство называется правилом сложения энтропии для опытов α и β.

Для доказательства этого равенства рассмотрим выражение

$$H(\alpha\beta) = -p(A_1B_1) \log p(A_1B_1) - \ldots - p(A_kB_l) \log p(A_kB_l)$$

α, β - независимы, следовательно $p(A_iB_j) = p(A_i) * p(B_j) \Rightarrow$

$$\log p(A_iB_j) = \log p(A_i)p(B_j) = \log p(A_i) + \log p(B_j).$$

Предположим далее, что α и β - зависимые опыты (пример: α, β - последовательные извлечения двух шаров из одной урны.) Постараемся выяснить, чему равна энтропия сложного опыта $\alpha\beta$ в этом случае.

Здесь уже нельзя заменить $p(A_1B_1), p(A_1B_2), \ldots$ произведением вероятностей, а необходимо использовать условную вероятность $p(A_1B_1) = p(A_1) * p(B_1|A_1)$

В этом случае можно доказать следующую формулу:

$H(\alpha\beta) = H(\alpha) + [p(A_1) * H(\beta|A_1) + p(A_2) * H(\beta|A_2) + \ldots + p(A_k) * H(\beta|A_k)]\,(*)$, где $H(\beta|A_i)$ - условная энтропия опыта β при условии, что значение опыта α равно A_i.

$H(\beta|A_i) = -p(B_1|A_i)\log p(B_1|A_i) - p(B_2|A_i)\log p(B_2|A_i) - \ldots - p(B_e|A_i)\log p(B_e|A_i)(**)$

Это выражение представляет собой энтропию опыта β при условии, что имеет место событие A_i.

$$\begin{cases} H(\alpha,\beta) = H(\alpha) + H(\beta) & (\text{ для независимых } \alpha,\beta) \\ H(\alpha,\beta) = H(\alpha) + [\ldots](*) & (\text{ для зависимых } \alpha,\beta) \end{cases}$$

Первый член последнего выражения $(*)$ - энтропия опыта α. Что же касается второго - он есть математическое ожидание случайной величины, принимающей с вероятностями $p(A_1), \ldots, p(A_k)$ значения $H(\beta|A_1), \ldots, H(\beta|A_k)$, то есть значения, равные условной энтропии опыта β, при условии, что опыт α имеет исходы $\alpha : A_1, \ldots, A_k$. Это среднее значение естественно назвать **условной энтропией** выполнения опыта β при условии выполнения опыта α,

$$H(\beta|\alpha) = [\ldots](*) = p(A_1)H(\beta|A_1) + p(A_2)H(\beta|A_2) + \ldots + p(A_k)H(\beta|A_k)$$

Тогда соотношение $(*)$ переписывается как $H(\alpha\beta) = H(\alpha) + H(\beta|\alpha)(*)$; α, β - зависимы.

Это и есть общее правило для определения энтропии сложного опыта $\alpha\beta$. Его также можно назвать правилом сложения энтропии, для **зависимых** опытов $\alpha\beta$.

Укажем основные свойства условной энтропии:

1. $H(\beta|\alpha) \geqslant 0$.
2. $p(A_1), \ldots, p(A_k) \neq 0$ (опыт имеет к штук исходов).

   Тогда $H(\beta|\alpha) = 0 \Longleftrightarrow H(\beta|A_1) = \ldots = H(\beta|A_k) = 0$, т.е. при любом исходе опыта α результат опыта β полностью определён, и при этом имеем $H(\alpha\beta) = H(\alpha)$.

   Если α и β **независимы**, то тогда $H(\beta|\alpha) = H(\beta)$, и $H(\alpha\beta) = H(\alpha) + H(\beta)$.
3. Во всех случаях условная энтропия $H(\beta|\alpha)$ заключается между 0 и $H(\beta)$:

   $0 \leqslant H(\beta|\alpha) \leqslant H(\beta)$.

   Таким образом случаи, когда исход β полостью предопределяется исходом α и когда опыты α и β независимы, являются в определённом смысле крайними.
4. Условная энтропия.

$$H(\alpha\beta) = H(\beta\alpha) \Rightarrow H(\alpha) + H(\beta|\alpha) = H(\beta) + H(\alpha|\beta) \Rightarrow$$

$$\Rightarrow H(\beta|\alpha) = H(\alpha|\beta)(.) + H(\beta) - H(\alpha)$$

$H(\alpha|\beta) = 0$ (исход опыта β полностью определяет опыта α

$H(\beta|\alpha) = H(\beta) - H(\alpha)$

Задача: Задача о болезненной реакции.

Известно, что некоторой болезнью в среднем болеют 2 человека из 100. Для выявления больных используется определённая реакция, которая всегда оказывается положительной в том случае, когда человек болен. Если же человек здоров, то она столь же часто бывает положительной, как и отрицательной. Пусть опыт β состоит в определении того болен или здоров человек, а опыт α - в определении результата указанной реакции. Спрашивается, какова будет энтропия $H(\beta) =?$ опыта β и условная энтропия $H(\beta|\alpha) =?$.

Решение: Очевидно, что β имеет 2 исхода: $\beta : \{B_1$ - здоров; B_2 - болен$\}$.

$p(B_1) = 0.98; \quad p(B_2) = 0.02.$

$H(\beta) = -0.98 \log 0.98 - 0.02 \log 0.02 \approx 0.14$бит.

$H(\beta) \approx 0.14.$

Рассмотрим опыт α: $\alpha : A_1 -$ положительная реакция; $A_2 -$ отрицательная реакция

$p(A_1) = p(\frac{B_1}{2} + B_2) = p(\frac{B_1}{2}) + p(B_2) = 0.49 + 0.02 = 0.51.$

$p(A_2) = p(\frac{B_1}{2}) = 0.49.$

$\alpha = A_1 : p(B_1|A_1) = \frac{p(B_1A_1)}{p(A_1)} = \frac{0.49}{0.51} = \frac{49}{51}.$

$\alpha = A_2 : p(B_2|A_1) = \frac{p(B_2A_1)}{p(A_1)} = \frac{0.02}{0.51} = \frac{2}{51}.$

Пользуясь этими данными мы можем найти условную энтропию $H(\beta)$ при выполнении события A_1

$$H(\beta|A_1) = -\frac{49}{51} \log \frac{49}{51} - \frac{2}{51} \log \frac{2}{51} \approx 0.24 \text{ бит.}$$

При $\alpha = A_2 \Rightarrow \beta = B_1; \ H(\beta|A_2) = 0$, т.е. мы с уверенностью можем утверждать, что человек здоров, и опыт β имеет исход B_1.

Таким образом, условная энтропия β при условии осуществления α будет равна

$$H(\beta) \ = 0.14|sysH(\beta|A_1) \approx 0.045 H(\beta|A_2) = 0 \sim\sim$$

$$H(\beta|\alpha) = p(A_1)H(\beta|A_1) + p(A_2)H(\beta|A_2) \approx 0.51 * 0.24 + 0.49 * 0 = 0.12 \text{ бит.}$$

Иначе говоря, выполнение опыта α уменьшает неопределённость опыта β на 0.002 бита.

Понятие об информации.

Вернёмся вновь к величине $H(\beta)$, характеризующей степень неопределённости опыта β. Равенство этой величины 0 означает, что исход опыта β заранее известен. Большее или меньшее значение числа $H(\beta)$ отвечает большей или меньшей проблематичности определения результата опыта β. Какое-либо измерение или наблюдение в виде опыта α, предшествующее β может ограничить количество возможных исходов опыта β, и тем самым уменьшить степень его неопределённости: так, к примеру степень неопределённости опыта, состоящего в нахождении самого тяжёлого из 3 грузов уменьшается после сравнения на весах двух из них.

Для того, чтобы результат измерения(наблюдения) α мог сказаться на последующем опыте β необходимо, чтобы α **не был известен заранее.** Поэтому, α можно рассматривать как вспомогательный опыт, также имеющий несколько допустимых исходов.

Тот факт, что осуществление α уменьшает степень неопределённости β отражается в неравенстве, где условная энтропия $H(\beta|\alpha) \leqslant H(\beta)$ первоначальной энтропии опыта β.

При этом, если опыт β не зависит от α, то осуществление α не уменьшает энтропии β. Это значит, что $H(\beta|\alpha) = H(\beta)$. Если же результат α полностью предопределяет исход опыта β, то энтропия β уменьшается до 0: $H(\beta|\alpha) = 0$. Таким образом, разность $\boxed{I(\beta,\alpha) = H(\beta) - H(\beta|\alpha)(*)}$.

Таким образом написанная разность указывает, насколько осуществление α уменьшает неопределённость β, т.е. как много мы узнаём об исходе опыта β, произведя измерение(наблюдение) в виде опыта α. Эта разность $(*)$ называют количеством информации относительно опыта β, содержащейся в опыте α. Таким образом, мы получаем возможность **численного измерения** информации. К примеру, в условиях задачи о болезненной реакции можно сказать, что используемая реакция в виде опыта α даёт информацию о заболевании в виде опыта β, равное $0.14 - 0.12 = 0.02$ бита. Эта цифра и оценивает пользу реакции.

Соотношение между понятиями энтропии и информации напоминает соотношение между физическими понятиями потенциала и разности потенциалов. Энтропия есть абстрактная мера неопределённости. Ценность этого понятия в значительной мере заключается в том, что оно позволяет оценить влияние на опыт β какого-либо другого опыта α как разность энтропий по формуле $(*)$.

Подчеркнём также, что информация относительно опыта β, содержащаяся в опыте α представляет собой среднее значение (математическое ожидание) случайной величины $H(\beta) - H(\beta|A_i)$, связанной с отдельными исходами A_i опыта α.

Пример: Задача о шарах и предварительной информации.

Пусть опыт β состоит в извлечении одного шара из урны:

β : 1 шар из 5 чёрных и 10 белых.

А опыт α_k состоит в предварительном извлечении (без возвращения обратно) K шаров:

α_k : K шаров извлечено.

$H(\beta) = ?$

$I(\beta, \alpha_1) = ?$

$I(\beta, \alpha_2) = ?$

$I(\beta, \alpha_{13}) = ?$

$I(\beta, \alpha_{14}) = ?$

Чему равна энтропия $H(\beta)$ и информация, содержащаяся в опыте α_1?

Решение:

$H(\beta) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} \approx 0.92$бита.

$$I(\beta, \alpha_1) = H(\beta) - H(\beta|\alpha_1)$$

$$H(\beta|\alpha_1) = \left[p(A_1^{\text{чёр}}) * H(\beta|A_1^{\text{чёр}})\right] + \left[p(A_1^{\text{бел}}) * H(\beta|A_1^{\text{бел}})\right] =$$

$$= -\frac{1}{3}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{чёр}})}_{4/14} * \log\frac{4}{14} + \underbrace{p(B^{\text{бел}}|A_1^{\text{чёр}})}_{5/7} * \log\frac{5}{7}\right] -$$

$$-\frac{2}{3}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{бел}})}_{5/11} * \log\frac{5}{11} + \underbrace{p(B^{\text{бел}}|A_1^{\text{бел}})}_{9/14} * \log\frac{9}{14}\right] \approx 0.004\text{бит}.$$

$$I(\beta|\alpha_2) = H(\beta) - H(\beta|\alpha_2)$$

$$H(\beta|\alpha_2) = p(A_1^{\text{ч}}A_2^{\text{ч}}) * H(\beta|A_1^{\text{ч}}A_2^{\text{ч}}) + p(A_1^{\text{ч}}A_2^{\text{б}}) * H(\beta|A_1^{\text{ч}}A_2^{\text{б}}) + p(A_1^{\text{б}}A_2^{\text{б}}) * H(\beta|A_1^{\text{б}}A_2^{\text{б}}) =$$

$$= -\left\{\frac{C_5^2}{C_{15}^2}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{чёр}}A_2^{\text{чёр}})}_{3/13} * \log\frac{3}{13} + \underbrace{p(B^{\text{бел}}|A_1^{\text{чёр}}A_1^{\text{чёр}})}_{10/13} * \log\frac{10}{13}\right] +\right.$$

$$+\frac{C_{10}^1 C_5^1}{C_{15}^2}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{чёр}}A_2^{\text{бел}})}_{4/13} * \log\frac{4}{13} + \underbrace{p(B^{\text{бел}}|A_1^{\text{чёр}}A_2^{\text{бел}})}_{9/13} * \log\frac{9}{13}\right] +$$

$$\left.+\frac{C_{10}^2}{C_{15}^2}\left[\underbrace{p(B^{\text{чёр}}|A_1^{\text{бел}}A_2^{\text{бел}})}_{5/13} * \log\frac{5}{13} + \underbrace{p(B^{\text{бел}}|A_1^{\text{бел}}A_2^{\text{бел}})}_{8/13} * \log\frac{8}{13}\right]\right\} \approx 0.008\text{бит}.$$

$$I(\beta, \alpha_{13}) = H(\beta) - H(\beta|\alpha_{13})$$

$H(\beta|\alpha_{13})$ - здесь неопределённость только в оставшихся двух шарах: они должны быть двух цветов, значит, мы взяли 4 чёрных и 9 белых шаров.

$$H(\beta|\alpha_{13}) = \frac{C_5^4 C_{10}^9}{C_{15}^{13}}(-)\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2})\right) = -\frac{\frac{5!}{4!(5-4)!}\frac{10!}{9!(10-9)!}}{\frac{15!}{13!(15-13)!*(-1)}} = \frac{2*5*10}{14*15} \approx 0.44$$

$I(\beta, \alpha_{14}) = [H(\beta|\alpha_1 4) = 0] \approx 0.92.$

Вообще, надо сказать, что количество информации об опыте β: $I(\beta\alpha)$, которая заключается в опыте α является объективной характеристикой **ценности прогноза**. $H(\beta|\alpha) = H(\beta)$, если α, β независимы, или если $H(\beta) = 0$ (исход β известен заранее и не нуждается в прогнозе). Во всех остальных случаях имеем $0 < I(\beta|\alpha) \leq H(\beta)$

Рассмотрим ситуацию, когда опыт β имеет **бесконечное** число исходов (непрерывное множество исходов). В этом случае $H(\beta) = \infty$, однако вместо неё часто можно рассматривать **конечную энтропию** $H_\varepsilon\beta$, которая получается при объединении исходов β, отличающихся не более, чем на малое число ε, в один исход. В практических задачах обычно H_ε (ε - энтропия) и имеет смысл, так как мы вообще не можем различить исходы β, отличающиеся меньше, чем на ε. (ε определяется точностью используемых измерительных приборов)

Информация $I(\beta, \alpha) = I(\alpha, \beta)$. Проверим: Очевидно равенство

$H(\alpha\beta) = H(\beta\alpha) \Rightarrow H(\alpha) + H(\beta|\alpha) = H(\beta) + H(\alpha|\beta) \Rightarrow H(\alpha) + H(\alpha|\beta) = H(\beta) + H(\beta|\alpha) \Rightarrow I(\alpha|beta) = I(\beta|\alpha)$

Таким образом, информацию $I(\beta, \alpha)$ можно назвать **взаимной информацией** опытов α и β друг относительно друга.

Пусть α, β, γ - 3 произвольных опыта. В таком случае всегда $I(\beta, \alpha\gamma) \geqslant I(\beta, \alpha)$

Иначе говоря, опыт $\alpha\gamma$ содержит не меньше информации, чем простой опыт β

Про **последовательной** передаче информации об опыте α, осуществляемой посредством цепочки опытов $\beta, \gamma, \delta \ldots$, где только опыт β непосредственно связан с α,а γ всю содержащуюся информацию об α получает из связи с опытом β (так, что $\beta\gamma$ не содержит дополнительной информации по сравнению с опытом β); δ всю информацию получает из связи с опытом $\gamma \ldots$

Информация об опыте α в этом случае может только уменьшаться.

$$H(\alpha) = I(\alpha, \alpha) \geqslant I(\alpha, \beta) \geqslant I(\alpha, \gamma) \geqslant I(\alpha, \delta) \geqslant \ldots$$

Наглядной иллюстрацией этого положения может служить игра "испорченный телефон".

Величина $I_\beta(\alpha, \gamma) = H(\alpha|\beta) - H(\alpha|\beta\gamma)$ называется условной информацией двух опытов α и γ друг от друга при выполнении опыта β.

Свойства условной информации:

1. $I_\beta(\alpha, \gamma) \geqslant 0$
2. $I_\beta(\alpha, \gamma) = I_\beta(\gamma, \alpha)$ (симметричность).
3. $I(\alpha, \beta\gamma) = I(\alpha, \beta) + I_\beta(\alpha, \gamma)$

   Получается из формул

   $I_\beta(\alpha, \gamma) = H(\alpha|\beta) - H(\alpha|\beta\gamma)$

   $I(\alpha, \beta\gamma) = H(\alpha) - H(\alpha|\beta\gamma)$

   $I(\alpha, \beta) = H(\alpha) - H(\alpha|\beta)$

Определение энтропии перечислением её свойств

Понятие энтропии с необходимостью вытекает из простейших требований, которые естественно наложить на величину, служащую количественной характеристикой степени неопределённости. Энтропия (мера степени неопределённости опыта α) определяется с помощью таблицы вероятностей.

α

Исходы опыта	A_1	...	A_k
Вероятности	$p(A_1)$	...	$p(A_k)$

$H(\alpha) = H(p_1, \ldots, p_k)$

$p(A_1) = p_1, \ldots, p_k$

Сформулируем те условия. выполнение которых надо требовать от функции энтропии:

1. $H(p_1, \ldots, p_k)$ не меняется при любой перестановке чисел $p_1 \ldots p_k$
2. $H(p_1, \ldots, p_k)$ непрерывна, т.е она мало изменяется при малых изменениях вероятностей p_k. В самом деле, при малых изменениях вероятностей и степень неопределённости опыта должна мало изменяться
3. Функция $H(p_1, ...p_k)$ удовлетворяет соотношению $H(\alpha) = H(p_1, p_2, ..., p_k) = \overbrace{H(p_1 + p_2, p_3, ..., p_k)}^{H(\beta)} + (p_1 + p_2) * H(\frac{p_1}{p_1+p_2} \frac{p_2}{p_1+p_2})$ Это значит, что неопределённость опыта β с таблицей вероятностей,

Исходы опыта	B	...	A_k
Вероятности	$p_1 + p_2$	...	$p(A_k)$

   получаемая отождествлением первых двух исходов опыта α меньше неопределённости последнего опыта, умноженную на меру неопределённости опыта $p_1 + p_2$, состоящую в выяснении того, какой именно из первых исходов опыта α имел место.

Можно доказать, что условия $1, 2, 3$ полностью определяют вид функции $H(p_1, p_2, \ldots, p_k)$: единственная функция, удовлетворяющая этим условиям имеет вид $H(p_1, p_2, \ldots, p_k) = -p_1 \log p_1 - ... - p_k \log p_k$ Следуя Шеннону, условия $1, 2, 3$ дополняют ещё одним условием : вводят в рассмотрение функцию $H(\frac{1}{k}, \ldots, \frac{1}{k}) = H(\alpha_0)$ (опыт α_0 имеет K равновероятных исходов).

Очевидно, что в силу равновероятности исходов, функция зависит лишь от числа K. Также ясно, что степень неопределённости опыта α_0 должна быть тем больше, чем больше число K его исходов. Тогда можно утверждать, что

4 $H(1, ...1) = f(k) \nearrow$ (растёт)

Процессы кодирования

Различные виды кодов и их характерные особенности

Основные понятия

Рассмотрим прежде всего **общую схему** передачи сообщений по линиям связи. Для определённости будем говорить о телеграфии. На одном конце линии отправитель подаёт некоторое сообщение, записанное при помощи 33 букв русского алфавита (ъ = ь, е = ё), но включая сюда и нулевую букву - промежуток между словами, или с помощью 27 букв латинского алфавита, или при помощи 10 цифр(цифровое сообщение)

Для передачи этого сообщения в случае обычного проводного телеграфа используется постоянный ток, некоторые характеристики которого телеграфист может менять по своему усмотрению. При этом он создаёт определённую **последовательность** сигналов, воспринимаемых вторым телеграфистом на другом конце линии. Простейшими различимыми элементарными сигналами, широко используемыми на практике являются посылка тока и пауза (отсутствие посылки). При помощи только двух этих сигналов можно передать любое сообщение, если условиться заменять каждую букву или цифру определённой комбинацией посылок тока и пауз. В технике связи правила, сопоставляющие каждому передаваемому сообщению некоторую комбинацию сигналов, называется **кодом** (телеграфным), а сама операция перевода сообщения в последовательность различимых сигналов называется **кодированием** сообщения. При этом коды, использующие только 2 различных элементарных сигнала называются **двоичными** кодами. Коды, использующие 3 различных сигнала, называются троичными и.т.д.

В телеграфии применяется целых ряд различных кодов, важнейшими из которых являются код Морзе, код Бодо.

В коде Морзе каждой букве/цифре сообщения сопоставляется некоторая последовательность кратковременных посылок тока (точек) и втрижды более длинных посылок тока (тире), разделяемых кратковременными паузами той-же длительности, что и точки. Пробел отмечается специальным разделительным знаком - длинной паузой, длинной с тире, а пробел между словами - ещё вдважды более длинной паузой.

В настоящее время код Морзе используется лишь при повреждении основных телеграфных линий, а также в КВ-радиотелеграфии.

Код Бодо: В обычных буквопечатающих телеграфных аппаратах, стоящих на всех больших телеграфных линиях чаще всего применяется двоичный код Бодо, сопоставляющий каждой букве некоторую последовательность из **пяти** простейших сигналов (посылок ток и пауз) одинаковой длительности. Так, как при этом все буквы передаются комбинациями сигналов одной и той же длительности (**равномерными кодами**), и поэтому в коде Бодо не требуется спецзнака, отделяющего одну букву от другой - и без того известно, что через каждые 5 элементарных сигналов кончается одна буква и начинается другая. (в приёмных аппаратах такое разделение на комбинации из 5 сигналов производится автоматически) Поскольку, комбинируя две возможности для первого сигнала с двумя возможностями для второго и.т.д. мы можем в результате составить всего $2^5 = 32$ различных комбинации, поэтому код Бодо в его простейшей форме позволяет передавать 32 различных буквы.

В некоторых телеграфных аппаратах кроме простого включения и выключения тока можно также изменять его направление на обратное (3 элементарных сигнала) Возможны

также ещё более сложные телеграфные аппараты, в которых посылки тока различаются не только по направлению, но и по силе тока, тем самым увеличивая число элементарных сигналов.

Увеличение числа элементарных сигналов позволяет сделать код более сжатым (т.е. уменьшить число сигналов, требующихся для передачи данного сообщения. Однако, вместе с тем, оно усложняет и удорожает систему передачи. Поэтому в технике связи всё-же предпочтительно применяются коды с малым числом элементарных сигналов.

Экономность кода

Сформулируем основную математическую задачу, с которой приходится иметь дело в технике связи. Пусть имеется сообщение, записанное при помощи некоторого алфавита, содержащего n символов. Требуется **закодировать** это сообщение, т.е. указать правило, сопоставляющее каждому такому сообщению определённую последовательность из m различных элементарных сигналов ($m = 2$ - двоичный, $m = 3$ - троичный)

Как выгоднее всего это сделать? Прежде всего надо объяснить, в каком смысле понимается слово выгоднее. Будем считать кодирование тем более выгодным, чем меньше элементарных сигналов требуется затратить на передачу сообщения.

Если считать, что каждый из элементарных сигналов продолжается одно и то же время, то наиболее выгодный код позволит затратить на передачу сообщение меньше всего времени. Поэтому переход к более выгодному коду, позволяющему увеличивать эффективность использования данной линии связи имеет несомненное практическое значение. Постараемся подробнее разобраться в том, какие вообще бывают коды. Для определённости будем считать, что код двоичный. В этом случае кодирование состоит в том, что каждой из m букв нашего алфавита сопоставляется какая-то последовательность двух элементарных сигналов - **кодовое обозначение** этой буквы. Элементарные сигналы заменим цифрами 0 и 1 (например, посылка тока - 1, отсутствие тока - 0). Требуется ещё, чтобы закодированное сообщение можно было **однозначно декодировать**, т.е, чтобы в длинной последовательности из нулей и единиц, сопоставляемой многобуквенному сообщению всегда можно было понять, где кончается кодовое обозначение одной буквы и начинается другой.

Нетрудно понять, какой код будет наиболее выгодным: будем измерять **выгодность** (экономность) данного двоичного кода при помощи максимального числа элементарных сигналов (цифр) из 0 и 1, требующегося для передачи одной буквы. Можно показать, что наибольшее число k элементарных сигналов, приходящихся на 1 букву удовлетворяет неравенству $k \geqslant \log n$ (n- число букв алфавита). Этот факт объясняется соображениями теории информации: одна буква N - буквенного алфавита может содержать информацию, равную $1\text{б} = \log n$, а каждый передаваемый элементарный сигнал, принимающий одно из двух значений (посылка тока или пауза) может содержать информацию не более, чем 1 бит. Поэтому, для передачи одной буквы надо, чтобы число элементарных сигналов было не менее $\log n$.

Метод двоичной системы счисления

Этот метод используется для построения наиболее экономного двоичного кода. В десятичной системе счисления каждое число представляется в виде суммы степеней числа 10: $k = a_k \cdot 10^k + a_{k-1} \cdot 10^{k-1} + \ldots + a_0 \cdot 10^0$, где числа $a_k, \ldots, a_1, a_0$ - цифры числа,

принимающие значения от 0 до 9. Число N при этом обозначается последовательностью своих цифр $a_k a_{k-1} \ldots a_0$. Аналогично этому число n можно представить в виде степеней числа 2: $n = b_l \cdot 10^l + \ldots + b_0 \cdot 10^0$, где числа $b_l, \ldots, b_0$ - цифры числа, принимающие значения 0 и 1. Число n обозначается последовательностью соответствующих цифр ($6 = 1 * 2^2 + 1 * 2^1 + 0 * 2^0$). Аналогично можно представить число в виде суммы степеней числа m - т.н. m-ичная система счисления.

Число k цифр в десятичной системе для записи числа n удовлетворяет неравенству $10^{k-1} \leqslant n < 10^k$. В промежутке между 10^1 и $10^2 - 1$ все цифры будут двузначными, а в промежутке $10^2..10^3 - 1$ - трёхзначными и.т.д. В двоичной системе счисления число k удовлетворяет неравенству $2^{k-1} \leqslant n < 2^k$, значит, число 6 - трёхзначное, а число 9 - четырёхзначное. При добавлении к двоичной записи ведущих нулей, мы придём к **равномерному** двоичному коду для n-буквенного алфавита с минимальной длиной кодового обозначения k.

Код Шеннона - Фано

Основной результат предыдущего параграфа : если число букв в алфавит равно n, а число используемых элементарных сигналов равно m, то при любом методе кодирования, среднее число k элементарных сигналов, приходящихся на одну букву алфавита должно быть $k > \frac{\log n}{\log m}$. (из неравенства $2^{k-1} \leqslant n < 2^k \Rightarrow \log n < k \cdot \log m$).

На самом деле, результаты предыдущих параграфов могут быть значительно улучшены, если воспользоваться введённым понятием энтропии и учесть статистические свойства реальных сообщений. Рассмотрим далее простейший случай сообщений, записанных при помощи некоторых n букв нашего алфавита, частоты появления которых на любом месте сообщения характеризуются вероятностями $p_1, \ldots, p_n$, причём $\sum\limits_n p_n = 1$. Вероятность p_i появления i-й буквы на **любом** месте сообщения принимается одной и той же вне зависимости от того, какие буквы стояли на всех предыдущих местах. Последовательные буквы сообщения предполагаются **независимыми** друг от друга.

Будем далее рассматривать только двоичные коды с элементарными сигналами 0 и 1. Среднее число k двоичных элементарных сигналов, приходящихся в закодированном сообщении на 1 букву исходного сообщения должно быть больше числа энтропии H : ($k > H = -p_1 \log p_1 - p_2 \log p_2 - \ldots - p_n \log p_n$), где H - энтропия опыта, состоящего в распознавании одной буквы текста (энтропии одной буквы). Отсюда сразу следует, что при любом методе кодирования, для записи сообщения из M букв не меньше MH элементарных сигналов.

Если вероятности $p1, \ldots, p_n$ - не все равны между собой, тогда $H < \log n = H(\alpha_0)$, поэтому учёт статистических закономерностей сообщения может позволить построить код более экономный, чем наилучший равномерный код, требующий $M \cdot \log n$ двоичных знаков для записи текста из M букв.

Код Шеннона - Фано

Для получения наиболее экономного кода удобно начать с того, чтобы расположить все n букв алфавита в один столбец в порядке убывания вероятностей. Затем все эти буквы следует разбить на 2 группы: **верхнюю** и **нижнюю**, так, чтобы суммарные вероятности для буквы сообщения принадлежать каждой из этих групп были возможно **более близки** одна к другой. Для букв первой группы в качестве первой цифры кодового обозначения используется 1, для букв второй группы - 0.

Далее, каждую из 2 полученных групп снова надо разделить на 2 части с возможно более близкой суммарной вероятностью. В качестве второй цифры кодового обозначения используют 1 или 0 в зависимости, принадлежит ли наша буква к первой или ко второй подгруппе. Процесс повторяется до тех пор, пока мы не придём к группам, каждая из которых не содержит по одной букве. Такой метод кодирования сообщений был впервые предложен в 1948-49 гг. независимо друг от друга Р.Фано, К.Шенноном.

Пример Пусть $n = 6$ (алфавит из 6 букв), вероятности которых равны соответственно $0.4; 0.2; 0.2; 0.1; 0.05; 0.05$.

На первом этапе деления на группы отщепим одну первую букву, оставив во второй группе все остальные. Далее, вторая буква составит первую подгруппу второй группы …

N буквы	Вероятность	Разбиение на подгруппы	Кодовое обозначение
1	0.4	I	1
2	0.2	II I	01
3	0.2	II II I	001
4	0.1	II II II I	0001
5	0.05	II II II II I	00001
6	0.05	II II II II II	00000

Выводы: Основной принцип, положенный в основу кодирования по методу Шеннона - Фано заключается в том, что при выборе каждой цифры кодового обозначения мы стараемся, чтобы содержащееся в ней количество информации было наибольшим, т.е. чтобы независимо от значения всех предыдущих цифр эта цифра принимала оба возможных для неё значений 0 и 1 о возможности с одинаковой вероятностью. Существенно, что буква, имеющая большую вероятность в коде Шеннона - Фано соответствуют более короткие кодовые обозначения (эти буквы быстрее оказываются выделенными в отдельную группу. В результате среднее значение k длины такого кодового обозначения оказывается только немногим большим минимального значения H, допускаемого соображениями сохранения количества информации при кодировании. Так, для рассмотренного выше примера шестибуквенного алфавита наилучший равномерный код состоит из трёхзначных кодовых обозначений (т.к. $2^2 \leqslant 6 < 2^3$), поэтому в нём на каждую букву приходится ровно 3 элементарных сигнала. При использовании кода Шеннона - Фано среднее число k элементарных сигналов, приходящихся на одну буку равно $M_\xi = \sum L \cdot$ принимаемые значения вероятности $= 1 * 0.4 + 2 * 0.2 + 3 * 0.2 + 4 * 0.1 + 5 * (0.05 + 0.05) = 2.3$. Это значение заметно меньше, чем 3, но не очень сильно отличается от энтропии $H = -0.4 \log 0.4 - 2 * 0.2 \log 0.2 - 0.1 \log 0.1 - 2 * 0.05 \log 0.05 \approx 2.22$.

Код Хаффмана

Близким к коду Шеннона-Фано, но ещё более выгодным является код Хаффмана. Построение этого кода опирается на простое преобразование того алфавита, на котором записываются передаваемые по линиям связи сообщения. Это преобразование называется **сжатием** алфавита.

Пусть имеется алфавит A, содержащий буквы $A_1, A_2, \ldots, A_n$, вероятности появления которых в сообщении соответственно равны $p_1, p_2, \ldots, p_n$. При этом мы считаем буквы расположенными в порядке убывания их вероятности ($p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$). Условимся не различать между собой две **наименее вероятные** буквы алфавита (a_{n-1}, a_n) - новая буква b алфавита $A_1 : a_1, a_2, \ldots, a_{n-2}, b$ и вероятностью появления $p_1, p_2, \ldots, p_{n-2}, (p_{n-1} + p_n)$.

Алфавит A_1 называется алфавитом, подученным из алфавита A с помощью одной операции сжатия. Расположим буквы нового алфавита A_1 в порядке убывания их вероятностей и подвергнем сжатию алфавит A_1. При этом мы придём к алфавиту A_2, получаемому из алфавита A двукратным сжатием. Продолжая эту процедуру мы будем приходить ко всё более коротким алфавитам. После $N-2$-кратного сжатия мы придём к алфавиту A_{n-2}, содержащему всего 2 буквы.

Пример Пусть $n = 6$ (алфавит из 6 букв), вероятности которых равны соответственно $0.4; 0.2; 0.2; 0.1; 0.05; 0.05$.

Рассмотрим наш алфавит из 6 букв с соответствующими вероятностями:

N буквы	Исх алфавит A	A_1	A_2	A_3	A_4
1	0.4 0	0.4 0	0.4 0	0.4 0	0.6 1
2	0.2 10	0.2 10	0.2 10	0.4 11	0.4 0
3	0.2 111	0.2 111	0.2 111	0.2 10	
4	0.1 1101	0.1 1101	0.2 110		
5	0.05 11001	0.1 1100			
6	0.05 11000				

Условимся приписывать двум буквам последнего алфавита кодовые обозначения 1 и 0. Если кодовые обозначения уже приписаны всем буквам алфавита A_j, то буквам предыдущего алфавита A_{j-1}, сохранившимся и в алфавите A_j мы припишем те-же кодовые обозначения, которые они имели в алфавите A_j. Двум буквам алфавита A_{j-1} : A', A", слившимся в одну букву b алфавита A_j мы припишем обозначения, получающиеся из кодового обозначения буквы b добавлением цифр 1 и 0 в конце.

Легко увидеть, что из самого построения получаемого таким образом кода Хаффмана вытекает, что он удовлетворяет **условию**: никакое кодовое обозначение не является здесь начало другого, более длинного, кодового обозначения. Заметим также, что кодирование некоторого алфавита по методу Хаффмана также, как и по методу Шеннона - Фано не является однозначно определённой процедурой, однако отметим, что среднее число k элем сигналов, приходящихся на одну букву алфавита для всех построенных кодов всегда остаётся одинаковым. Можно показать, что код Хаффмана является **самым экономным** из всех возможных в том смысле,что ни для какого другого метода кодирования букв некоторого алфавита среднее число k элементарных сигналов, приходящихся на одну букву не может быть меньше того, какое получается при кодировании по методу Хаффмана.

Основная теорема о кодировании

Достигнутая степень близости среднего числа k двоичных знаков, приходящихся на одну букву сообщения к H может быть ещё сколь угодно увеличена при помощи перехода к кодированию всё более и более длинных блоков. Это вытекает из следующего общего утверждения, которое называется основной теоремой о кодировании.

Теорема: При кодировании сообщения , разбитого на N-буквенные блоки можно, выбрав N достаточно большим добиться того, чтобы среднее число k элементарных двоичных сигналов, приходящихся на одну букву исходного сообщения было сколь угодно близко к H. Замечание: Очень длинное сообщение из M букв может быть закодировано

при помощи сколь угодно близкого к числу MH (но большего) числа элементарных сигналов, если только предварительно разбить это сообщение на достаточно длинные блоки из N букв и сопоставлять отдельные кодовые обозначения сразу целым блокам. Методы кодирования блоков могут быть самыми различными (например, можно использовать методы Шеннона - Фано, Хаффмана)

m-ичные коды

Содержание предыдущих параграфов легко переносится и на случай m-ичных кодов, использующих m элементарных сигналов. Так, например для построения m-ичных кодов Шеннона - Фано надо лишь разбивать группы символов не на 2, а на m частей, по возможности близкой суммарной вероятности, а для построения m-ичного кода Хаффмана надо использовать операцию сжатия алфавита, при которой каждый раз сливаются не две, а m букв исходного алфавита, имеющих наименьшие вероятности.

Ввиду важности кодов Хаффмана остановимся на этом вопросе подробнее. Сжатие алфавита, при котором m букв заменяются на одну приводит к уменьшению числа букв на $m-1$. Так, как для построения m-ичного кода, очевидно, требуются, чтобы последовательность сжатий привела нас к алфавиту из m букв (сопоставляемых m сигналам кода), то необходимо, чтобы число n букв первоначального алфавита было представимо в виде $n = m + s(m-1)$, где s - целое число сжатий.

Этого всегда можно добиться, добавив, если нужно, к первоначальному алфавиту ещё несколько "фиктивных букв", вероятности появления которых считаются равными 0. После этого, построение m-ичного кода Хаффмана производится точно так-же, как и в случае двоичного кода.

Пример: В случае алфавита из 6 букв, имеющих вероятности $0.4; 0.2; 0.2; 0.1; 0.05; 0.05$. Для построения **троичного** кода Хаффмана надо присоединить к нашему алфавиту ещё одну фиктивную букву с нулевой вероятностью и поступать, как указано в таблице:

Номер буквы	Вероятности и кодовые обозначения		
	Исходный алфавит A	Сжатый алфавит A_1	Сжатый алфавит A_2
1	0.4 - 0	0.4 - 0	0.4 - 0
2	0.2 - 2	0.2 - 2	0.4 - 1
3	0.2 - 10	0.2 - 10	0.2 - 2
4	0.1 - 11	0.2 - 11	
5	0.05 - 120		
6	0.05 - 121		
7	0 - ---		

Теорема: Любые n чисел $k_1, k_2, \ldots, k_n$, удовлетворяющие неравенству $\frac{1}{m^{k_1}} + \frac{1}{m^{k_2}} + \ldots + \frac{1}{m^{k_n}} \leqslant 1 \quad (*)$. Любые из k_n чисел являются длинами сообщений некоторого m-ичного кода, сопоставляющего n буквам алфавита n последовательностей элементарных сигналов, принимающих m возможных значений.

Это утверждение $(*)$ впервые было доказано в 1949 году американским учёным Л.Крафтом и позднее было обобщено Б. Макмилланом, поэтому неравенство $(*)$ часто называют неравенством Крафта - Макмиллана. Используя неравенство $(*)$ можно получить следующий результат:

Теорема: основная теорема о кодировании для m-ичных кодов. При любом методе кодирования, использующем m-мчнный код среднее число k элементарных сигналов, приходящихся на одну букву сообщения никогда не может быть меньше отношения $\frac{H}{\log m}$, где H - энтропия одной букв сообщения. Однако, оно всегда может быть сделано сколь угодно близким к этой величине, если кодировать сразу достаточно длинные блоки из N букв.

Следствие: Если по линии связи за единицу времени можно передать L элементарных сигналов, принимающих m различных значений, то скорость **передачи сообщений** по такой линии не может быть большей, чем $v = \frac{L \cdot \log m}{H} \left[\frac{\text{букв}}{\text{ед.времени}}\right]$. Однако, передача со скоростью, сколь угодно близкой к v (но меньшей v) является возможной. Величина $C = L \cdot \log m$ зависит лишь от характеристик самой линии связи, в то время, как знаменатель H характеризует передаваемое сообщение. Величина C указывает наибольшее количество единиц информации, которое можно передать по линии связи за единицу времени. Она называется пропускной способностью линии связи.

Энтропия и информация конкретных типов сообщений. Письменная речь.

Основной результат, полученный ранее состоял в том, что для передачи M-буквенного сообщения по линии связи, допускающей m различных элементарных сигналов требуется затратить не меньше, чем $\frac{M \cdot \log n}{\log m}$, где n - число букв алфавита, с помощью которого записано исходное сообщение. При этом существуют методы кодирования, позволяющие сколь угодно близко подойти к границе этой величины.

Так, как русский телеграфный алфавит содержит 32 буквы (е=ё, ь=ъ, пробел), то согласно этому результату на передачу M-буквенного сообщения надо затратить $\frac{M \cdot \log 32}{\log m} = \frac{M H_0}{\log m}$, где $H_0 = \log 32 = 5$ бит - энтропия опыта, состоящего в приёме одной буквы русского текста (информация, содержащаяся в одной букве), при условии, что все буквы считаются **одинаково вероятными**.

На самом деле, появление в сообщении разных букв совсем не одинаково вероятны. Буквы О, Е встречаются много чаще, чем буквы Ф или Щ. Для более точного вычисления информации, содержащейся в одной букве русского текста надо знать вероятности (частоты) появления различных букв. Ориентировочные значения частот отдельных букв русского языка задаются следующей таблицей:

Буква	Пробел	О	Е,Ё	А	И	Т	Н	С
Вероятность	0.175	0.09	0.072	0.062	0.062	0.053	0.053	0.045
Буква	Р	В	Л	К	М	Д	П	У
Вероятность	0.04	0.038	0.035	0.028	0.026	0.025	0.023	0.021
Буква	Я	Ы	З	Ь,Ъ	Б	Г	Ч	Й
Вероятность	0.018	0.016	0.016	0.014	0.014	0.013	0.012	0.001
Буква	Х	Ж	Ю	Ш	Ц	Щ	Э	Ф
Вероятность	0.009	0.007	0.006	0.006	0.004	0.003	0.003	0.002

Приравняв эти частоты вероятностям появления соответствующих букв получим для энтропии одной буквы русского текста следующее значение: $H_1 = H(\alpha_1) = -0.175 \cdot \log 0.175 - 0.09 \cdot \log 0.09 - \ \ldots \ - 0.002 \cdot \log 0.002 \approx 4.35$бит.

Из сравнения этого значения с величиной $H_0 = 5$ бит видно, что неравномерность появления различных букв алфавита приводит к уменьшению информации, содержащейся в одной букве русского текста на ≈ 0.65 бита.

Воспользовавшись этим обстоятельством можно уменьшить число элементарных сигналов, необходимых для передачи M-буквенного сообщения до значения $\frac{M \cdot H_1}{\log m}$, т.е. в случае двоичного кода до значения $M \cdot H_1$. Сокращение числа требующихся элементарных сигналов может быть достигнута, например, кодированием отдельных букв русского алфавита по методу Шеннона - Фано.

Передача непрерывно изменяющихся сообщений. Телевизионные сообщения

В устной речи, музыке, возможными сообщениями являются уже не последовательности символов или букв, а совокупности звуковых колебаний, которые меняются **непрерывным** образом. Мы вполне можем объединить большое число схожих между собой звуков (фонем), если только замена одного из них другим не изменяет смысла прозвучавшего (сказанного). На самом деле, различимо лишь конечное число градаций громкости и высоты тона. Отождествив все звуки, громкость и высоту тона, которые находятся в пределах одной градации, мы снова придём к привычному для нас случаю последовательности сигналов, которые могут принимать только **конечное** число различных значений.

Опыт β вполне можно заменить новым опытом $\beta \to \beta_\varepsilon$, получающимся из β при помощи отождествления всех его исходов, отличающихся друг от друга меньше, чем на некоторое малое число ε.

Энтропию H_ε нового опыта β_ε называют ε - энтропией опыта β. При передаче непрерывно меняющихся сообщений совокупность возможных значений сигнал всегда разбивается на конечное число градаций (ячеек), и все значения в пределах одной градации отождествляются между собой. Эта операция замены непрерывного сообщения новым сообщением, принимающим конечное число возможных значений называется в технике связи **квантованием** сообщения.

Важным классом таких непрерывно меняющихся сообщений являются изображения, передаваемые по телевидению или фототелеграфным линиям связи. Любое изображение можно передавать по точкам, каждая из которых является сигналом, принимающим конечное число значений. Учитывается градация яркости. На самом деле, изображения являются неподвижными, на экране ежесекундно сменяется 25 кадров, создавая впечатление движения.

Общее число элементов(точек), на которые следует разлагать изображение определяется в первую очередь т.н. "разрешающей способностью" глаза, т.е. его способностью различать близкие фрагменты изображения. В современном телевидении это число имеет значение от 200 - 300 тыс до 500 тыс. - 1 млн. Нетрудно понять, что по этой причине энтропия телевизионного изображения имеет огромную величину.

Вопросы к зачёту по курсу ТИ.

1. 1.1. Вероятность. Случайные события и случайные величины.
   1.2. Свойства вероятности. Сложение и умножение вероятностей. Несовместные и независимые случайные события
   1.3. Условные вероятности и их свойства.
   1.4. Математическое ожидание случайной величины. Свойства.
   1.5. Дисперсия случайной величины. Свойства.
2. 2.1. Энтропия как мера степени неопределённости.
   2.2. Энтропия сложных событий. Условная энтропия.
   2.3. Свойства условной энтропии
   2.4. Задача о болезненной реакции.
   2.5. Понятие об информации. Количество информации.
   2.6. Задача о шарах и предварительной информации.
   2.7. ε -энтропия опыта. Иллюстрация игры в испорченный телефон.
   2.8. Условная информация. Свойства условной информации.
   2.9. Определение энтропии перечислением её свойств.
3. 3.1. Кодирование. Основные понятия. Определение кода. Двоичный код. Код Морзе.
   3.2. Код Бодо. Равномерный код. Экономность кода
   3.3. Метод двоичной системы счисления.
   3.4. Код Шеннона - Фано
   3.5. Пример на использование кода Шеннона - Фано
   3.6. Код Хаффмана. Сжатие кода. Пример на использование кода Хаффмана
   3.7. Задача о кодировании букв русского алфавита по методу Шеннона - Фано

Литература по кодированию и декодированию

- Питерсон У. Коды, исправляющие ошибки. М.:Мир, 1964.
- Добрушин Р.Л. Теория оптимального кодирования информации (сборник "Кибернетика на службе коммунизма" под ред. ак. Берга А.И. Т.3, с. 13-46., 1966г.)
- Элайс П. Кодирование и декодирование. ("Лекции по теории связи" под ред. Е.Багдади.:М, Мир, 1964, с 289 - 317.)