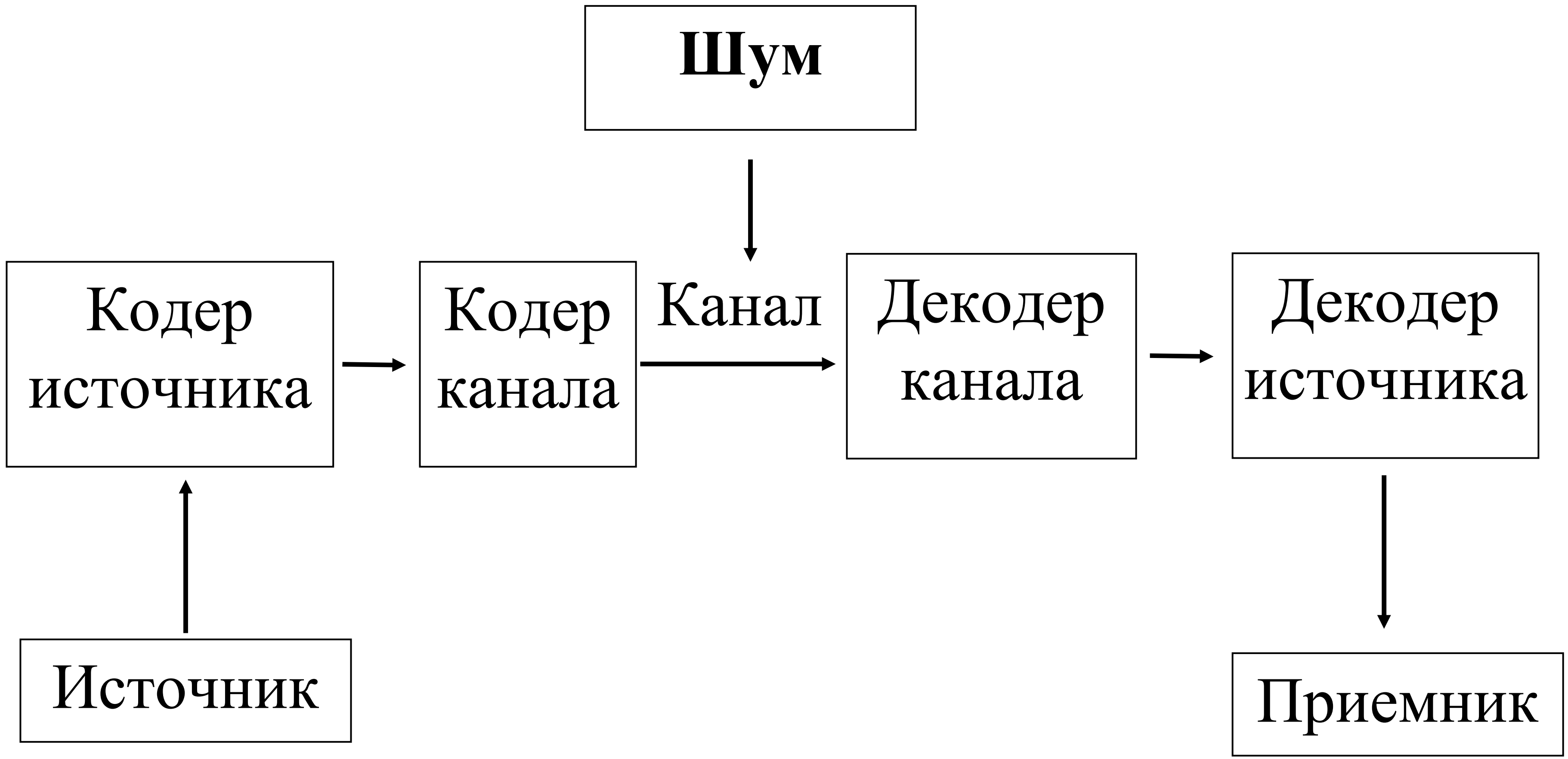
- **Основной моделью, которую изучает теория информации, является *модель системы передачи сигналов:***

Модель системы передачи сигналов



2 Понятие количества информации

2.1 Виды информации

Информация может быть двух видов:

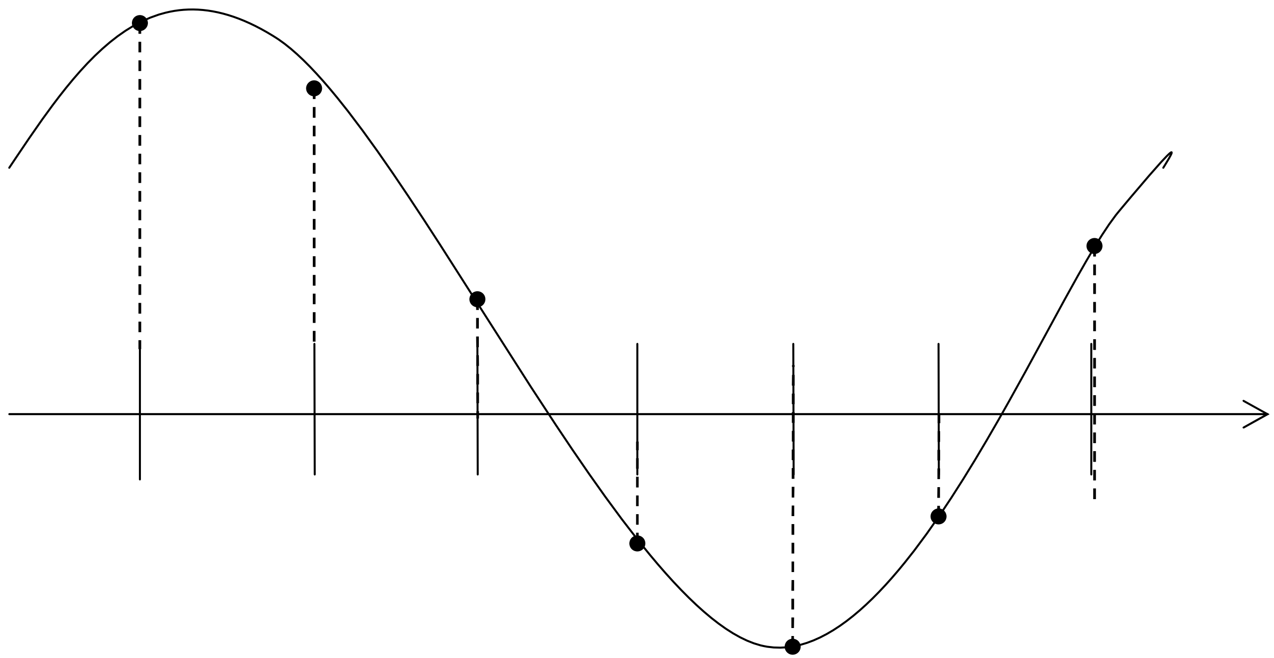
- **дискретная (цифровая)**
- **непрерывная (аналоговая)**

Непрерывная информация характеризуется непрерывным процессом изменения некоторой случайной величины (сигнала)

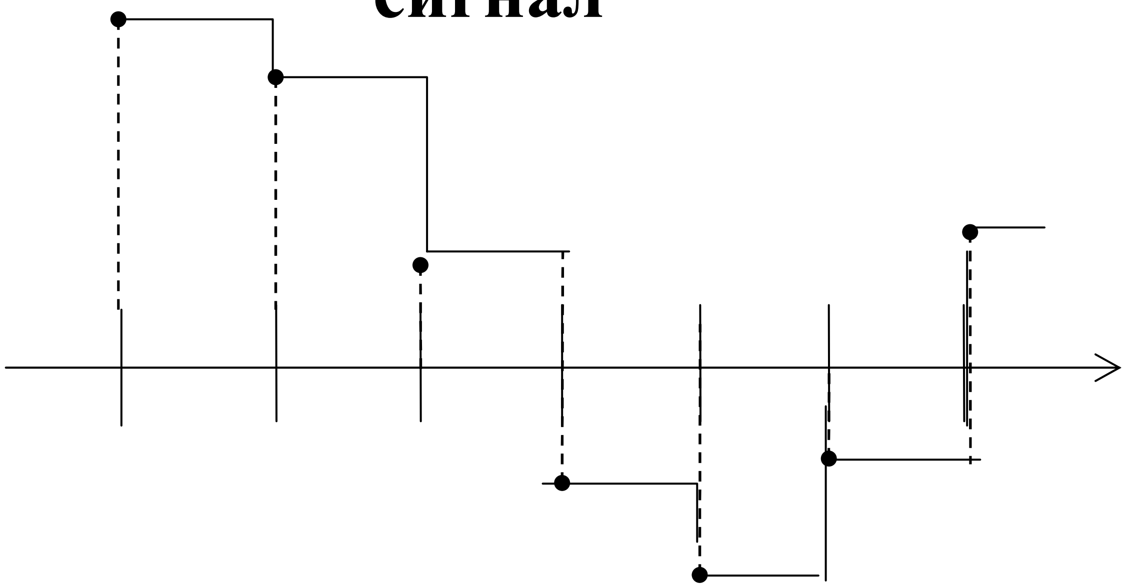
- **Дискретная информация характеризуется последовательными точными значениями некоторой случайной величины**

- **Дискретную информацию можно получить, измеряя непрерывную информацию через определенные интервалы времени.**

Исходный сигнал



Дискретизированный сигнал



- Дискретная информация удобнее для обработки, хранения и передачи, поскольку она описывается последовательностью чисел.
- Если представить каждое число в двоичной системе счисления, то дискретная информация предстанет в виде последовательности нулей и единиц.

2.2 Оценка количества информации сообщений

- **Дискретные сообщения, передаваемые по каналам связи, можно представить как набор из n символов, которые последовательно порождает некоторый дискретный источник информации.**
- **Элементы сообщения принимают значения из конечного алфавита мощности m дискретного источника.**

Общее количество различных сообщений равно

$$L = m^n$$

- **Чем больше L, тем значительней отличается каждое данное сообщение от остальных, т. е. величина L может служить мерой количества информации для равновероятных сообщений.**

Мера *I(L)* количества информации обладает следующими свойствами

1. $I(1)=0$
2. $I(L_1) > I(L_2)$ при $L_1 > L_2$
3. $I(L_1 \cdot L_2) = I(L_1) + I(L_2)$

Формула Хартли

$$I(L) = \log L = \log m^{n} = n \log m$$

Рассмотрим теперь сообщения длины n, символы которого порождаются независимо некоторым источником информации в соответствии с распределением дискретной случайной величины

элементы сообщения могут

принимать значения $a_1, a_2, ..., a_m$

с вероятностями $p_1, p_2, ..., p_m$ $\sum_{i=1}^{m} p_i = 1$

Вычислим среднюю вероятность такого сообщения

- **Обозначим** $\boldsymbol{n_k}$ **количество элементов сообщения, принимающих значение** $\boldsymbol{a_k}$**,** $\boldsymbol{k=1\ldots m}$
- **Тогда** $$\sum_{k=1}^{m} n_k = n$$

- **Поскольку элементы сообщения принимают значения независимо, то вероятность сообщения равна**

$$\prod_{k=1}^{m} p_k^{n_k}$$

- **При достаточно больших *n* по теореме Бернулли допустимо приближение**
- $$\frac{n_k}{n} = p_k \quad \text{или} \quad n_k = np_k$$

и тогда $$\prod_{k=1}^{m} p_k^{n_k} \approx \prod_{k=1}^{m} p_k^{np_k}$$

- **Среднее число всех возможных сообщений**

$$L = \frac{1}{\prod_{k=1}^{m} p_k^{np_k}}$$

Формула Шеннона

- **Среднее количество информации *I(L)*, содержащееся в одном сообщении**

$$I(L) = \log L = -\log \prod_{k=1}^{m} p_k^{np_k} = -n \sum_{k=1}^{m} p_k \log p_k$$

$$H(p_1, p_2, \ldots, p_m) = -\sum_{k=1}^{m} p_k \log p_k$$

При равновероятных состояниях элементов сообщений, т.е. $p_k = \dfrac{1}{m}$ формула Шеннона превращается в уже знакомую формулу

$$I(L) = n \log m$$

Пример

- **Источник информации порождает сообщения, состоящие из символов *a,b,c,d* с вероятностями p_1=0.1, p_2=0.1, p_3=0.1, p_4=0.7. Определить количество информации, приходящуюся на букву источника информации.**

- **При одинаковых вероятностях появления любой из всех *m*=4 букв алфавита количество информации, приходящуюся на одну букву, характеризует величина log4=2 бит.**

Найдем количество информации источника

$$H(A) = -\sum_{k=1}^{4} p_k \log p_k =$$

$$= 0.1 \cdot \log 10 + 0.1 \cdot \log 10 + 0.1 \cdot \log 10 + 0.7 \cdot \log \frac{10}{7} =$$

$$= \log 10 - 0.7 \log 7 = 1.37$$

- Таким образом, неравномерность распределения вероятностей использования букв снижает количество информации источника

  с 2 до 1.37 бит

2.3 Энтропия. Свойства энтропии

- **Любое сообщение представляет собой совокупность сведений о некоторой физической системе.**
- **Если бы состояние физической системы было известно заранее, не было бы смысла передавать сообщение.**
- **Сообщение содержит информацию только тогда, когда состояние системы заранее неизвестно, т.е. неопределено.**

- **В качестве меры априорной неопределенности дискретного источника в теории информации применяется специальная характеристика, называемая энтропией.**

Энтропия Шеннона

$$H(p_1,\ldots,p_m) = \frac{I(L)}{n} =$$

$$= -\sum_{k=1}^{m} p_k \log p_k = \sum_{k=1}^{m} p_k \log \frac{1}{p_k}$$

Основные свойства энтропии

- 1. Значение $H(p_1,...,p_m)$

  не зависит от любой перестановки

  $p_1,...,p_m$

Основные свойства энтропии

- 2. $H(p_1,\ldots,p_m)$ **непрерывна**

  Если $p_k \to 0$, то $p_k \cdot \log \dfrac{1}{p_k} \to 0$

- 3. $0 \le H(p_1,\ldots,p_m) \le \log m$

Основные свойства энтропии

4. Если m=2 и p_1=p, p_2=1-p то выражение для энтропии принимает вид

$$H(p) = -(p\log p + (1-p)\log(1-p))$$

Энтропия двоичного сообщения изменяется от 0 до 1 и достигает максимума при равных вероятностях (p_1=p_2=0.5), т.е. когда ситуация является наиболее неопределенной.


Зависимость энтропии от вероятности
1,1
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
H(p)
0 0,1 0,2 0,3 0,4 0,5 0,6 0,7 0,8 0,9 1
p

Основные свойства энтропии

5. Энтропия сообщения, состоящего из некоторых частных независимых сообщений, равна сумме энтропий составляющих его частей.

- **Действительно, пусть имеются два независимых сообщения *A* и *B* с энтропиями *H(A)* и *H(B)* соответственно.**
- **Вероятность совместного события *AB* равна произведению вероятностей событий *A* и *B***

$$p(AB) = p(A)p(B)$$

- Тогда

$$H(AB) = -\sum_{A,B} p(A)p(B)\log(p(A)p(B)) =$$

$$= -\sum_{A,B} p(A)p(B)(\log p(A) + \log p(B)) =$$

$$= -\sum_{A}(p(A)\log p(A)\sum_{B} p(B)) - \sum_{B}(p(B)\log p(B)\sum_{A} p(A)) =$$

$$= H(A) + H(B)$$

Пример

- **В таблицах заданы распределения вероятностей двух случайных дискретных величин *X* и *Y*.**

Случайные величины *X*	**0.5**	**0.7**	**0.9**	**0.3**
Вероятности их появления	**0.25**	**0.25**	**0.25**	**0.25**
Случайные величины *Y*	**5**	**10**	**15**	**8**
Вероятности их появления	**0.25**	**0.25**	**0.25**	**0.25**

Сравнить энтропии данных случайных величин

- **Энтропия не зависит от конкретных значений случайной величины. Так как вероятности появления символов в обоих случаях одинаковы, то**

$$H(X) = H(Y) = -\sum_{k=1}^{4} p_k \log p_k =$$

$$= -4 \cdot (0.25 \cdot \log 0.25) = 2$$

2.4 Энтропия объединения Условная энтропия

Пусть имеются два источника информации

- **Первый источник А порождает символы** $a_1, a_2, ..., a_m$ **с вероятностями** $p(a_1), p(a_2), ..., p(a_m)$

- **Второй источник В порождает символы** $b_1, b_2, ..., b_s$ **с вероятностями** $p(b_1), p(b_2), ..., p(b_s)$

- **Необходимо определить энтропию сложного сообщения**

ab

- **отдельные части сообщения порождаются источниками *A* и *B***

Энтропия объединения двух источников сообщений *A* и *B* определяется суммой по всем возможным состояниям объединения

$$H(A,B) = -\sum_{k=1}^{m}\sum_{j=1}^{s} p(a_k, b_j)\log p(a_k, b_j)$$

где $p(a_k, b_j)$ – вероятность совместного появления двух статистически зависимых состояний a_k и b_j.

- **Вероятность совместного появления двух состояний может быть выражена через вероятность появления одного из состояний и условную вероятность появления другого состояния равенством**

$$p(a_k, b_j) = p(a_k)\, p(b_j / a_k)$$

$$H(A,B) = -\sum_{k=1}^{m}\sum_{j=1}^{s} p(a_k)p(b_j/a_k)\log(p(a_k)p(b_j/a_k)) =$$

$$= -\sum_{k=1}^{m}\sum_{j=1}^{s} p(a_k)p(b_j/a_k)\log p(a_k) - \sum_{k=1}^{m}\sum_{j=1}^{s} p(a_k)p(b_j/a_k)\log p(b_j/a_k)$$

- **Первая сумма в соотношении представляет собой энтропию сообщений *H(A)*, а вторая сумма – условную энтропию *H*(*B/A*).**

- **Таким образом, окончательно имеем**

$$H(A,B) = H(A) + H(B/A)$$

- **Аналогично можно получить симметричное соотношение**

$$H(A,B) = H(B) + H(A/B)$$

- **Для фиксированного символа a_k совокупность условных вероятностей определяет частную условную энтропию**

$$H(B/a_k) = -\sum_{j=1}^{s} p(b_j/a_k)\log p(b_j/a_k)$$

которая характеризует информативность сообщений B после того, как стал известен символ a_k

- **Если частную условную энтропию усреднить по всем значениям a_k , то найдем общую условную энтропию сообщений *B* относительно сообщений *A***

$$H(B/A) = -\sum_{k=1}^{m} p(a_k) H(B/a_k)$$

$$H(B/A) = -\sum_{k=1}^{m} \sum_{j=1}^{s} p(a_k) p(b_j/a_k) \log p(b_j/a_k)$$

Поскольку

$$p(a_k, b_j) = p(a_k)\, p(b_j / a_k)$$

то

$$H(B/A) = -\sum_{k=1}^{m} \sum_{j=1}^{s} p(a_k, b_j) \log p(b_j / a_k)$$

Основные свойства энтропии объединения

- 1. $$H(A,B)=H(B,A)$$

Основные свойства энтропии объединения

2. **Если источники сообщений *A* и *B* статистически независимы, то энтропия объединения равна сумме энтропий источников сообщений *A* и *B***

$$H(A,B)=H(A)+H(B)$$

Основные свойства энтропии объединения

3. **Если источник В полностью статистически зависит от источника А, то** $H(A,B) = H(A)$

Основные свойства условной энтропии

1. **Если сообщения X и Y статистически независимы, то условная энтропия сообщений Y относительно сообщений X равна безусловной энтропии сообщений , т.е.**

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

Основные свойства условной энтропии

2. Если сообщения X и Y статистически жестко зависимые, т.е. появление одного из них непременно влечет появление другого, то условная энтропия сообщений Y относительно сообщений X равна нулю, т.е.

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

Основные свойства условной энтропии

3 $$0 \le H(Y / X) \le H(Y)$$

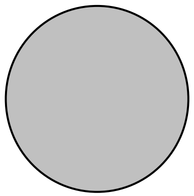
Пример

Известны энтропии двух зависимых источников информации X и Y

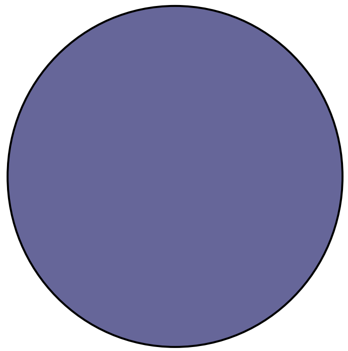
$$H(X)=5 \text{ бит}, \; H(Y)=10 \text{ бит}$$

Определить, в каких пределах могут изменяться величины условных энтропий $H(X/Y)$ и $H(Y/X)$

- При отсутствии взаимосвязи между источниками информации



Энтропия источника H(X)

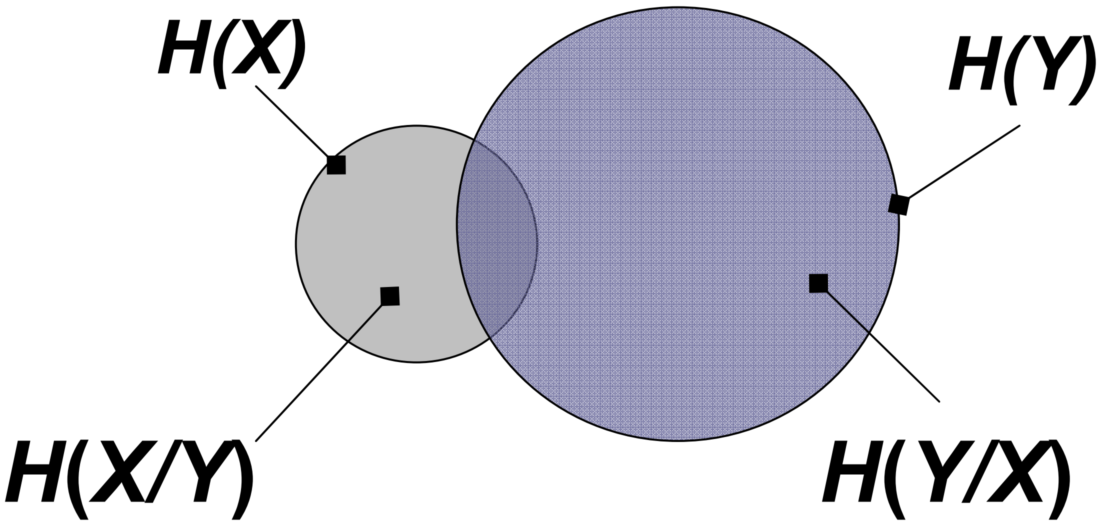


Энтропия источника H(Y)

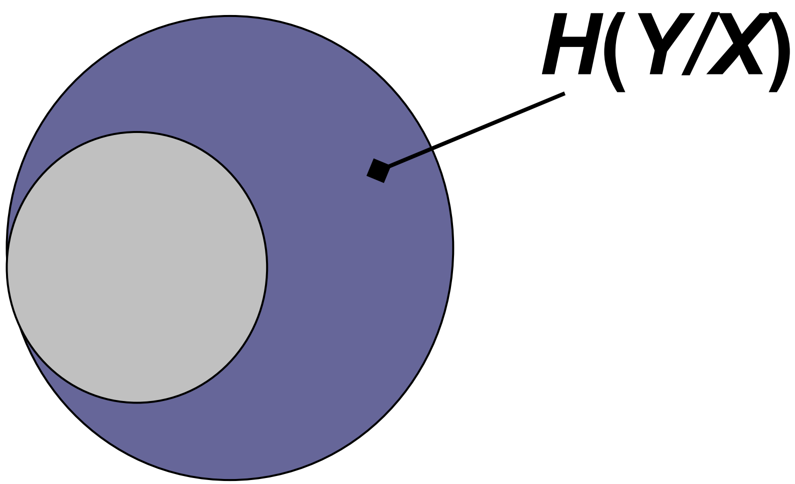
- **Если источники информации независимы, то**
  $H(Y/X) = H(Y)$=10 бит,
  $H(X/Y) = H(X)$=5 бит.

- **Эти значения условных энтропий являются максимальными.**

- **По мере увеличения взаимосвязи источников условные энтропии $H(X/Y)$ и $H(Y/X)$ будут уменьшаться**



- **При полной статистической зависимости двух источников один из них не вносит никакой информации**



- **При этом**
  $H(Y/X)=H(Y)-H(X)=$ 10-5=5 бит,

  Поэтому $H(Y/X)$ будет изменяться от 10 бит до 5 бит при максимально возможном изменении $H(X/Y)$ от 5 бит до 0 бит.

Пример

Закон распределения вероятностей системы, объединяющей зависимые источники информации Х и У, задан в таблице совместных вероятностей:

таблица совместных вероятностей

$X \backslash Y$	y_1	y_2	y_3
x_1	0.4	0.1	0
x_2	0	0.2	0.1
x_3	0	0	0.2

- **Определить величины**

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

- Сначала вычислим безусловные вероятности p(x) и p(y)

- **Сложив вероятности по строкам таблицы, получим вероятности появления значений $p(x_i)$**

$$p(x_1) = 0.4 + 0.1 + 0 = 0.5$$

$$p(x_2) = 0 + 0.2 + 0.1 = 0.3$$

$$p(x_3) = 0 + 0 + 0.2 = 0.2$$

- **Сложив вероятности по столбцам таблицы, получим вероятности появления значений** y_j

$$p(y_1) = 0.4 + 0 + 0 = 0.4$$

$$p(y_2) = 0.1 + 0.2 + 0 = 0.3$$

$$p(y_3) = 0 + 0.1 + 0.2 = 0.3$$

- **Вычислим теперь энтропии источников информации *X* и *Y* по формуле Шеннона, используя вычисленные ранее значения безусловных вероятностей.**

$$H(X) = -\sum_{i=1}^{3} p(x_i)\log p(x_i) = -(0.5 \cdot \log 0.5 + 0.3 \cdot \log 0.3 + 0.2 \cdot \log 0.2) = 1.485$$

$$H(Y) = -\sum_{j=1}^{3} p(y_j)\log p(y_j) = -(0.4 \cdot \log 0.4 + 0.3 \cdot \log 0.3 + 0.3 \cdot \log 0.3) = 1.57$$

- **Определим условные вероятности появления** y_j **при условии** x_i **появления**

  по формуле $$p(y_j / x_i) = \frac{p(x_i, y_j)}{p(x_i)}$$

$$p(y_1 \mid x_1) = \frac{p(x_1, y_1)}{p(x_1)} = \frac{0.4}{0.5} = 0.8$$

$$p(y_2 \mid x_1) = \frac{p(x_1, y_2)}{p(x_1)} = \frac{0.1}{0.5} = 0.2$$

$$p(y_3 \mid x_1) = \frac{p(x_1, y_3)}{p(x_1)} = \frac{0}{0.5} = 0$$

$$p(y_1 \mid x_2) = \frac{p(x_2, y_1)}{p(x_2)} = \frac{0}{0.3} = 0$$

$$p(y_2 \mid x_2) = \frac{p(x_2, y_2)}{p(x_2)} = \frac{0.2}{0.3} = 0.67$$

$$p(y_3 \mid x_2) = \frac{p(x_2, y_3)}{p(x_2)} = \frac{0.1}{0.3} = 0.33$$

$$p(y_1 \mid x_3) = \frac{p(x_3, y_1)}{p(x_3)} = \frac{0}{0.2} = 0$$

$$p(y_2 \mid x_3) = \frac{p(x_3, y_2)}{p(x_3)} = \frac{0}{0.2} = 0$$

$$p(y_3 \mid x_3) = \frac{p(x_3, y_3)}{p(x_3)} = \frac{0.2}{0.2} = 1$$

таблица условных вероятностей

$X \backslash Y$	y_1	y_2	y_3
x_1	0.8	0.2	0
x_2	0	0.67	0.33
x_3	0	0	0.1

$$H(Y \mid X) = -\sum_{k=1}^{m}\sum_{j=1}^{s} p(x_k, y_j)\log p(y_j \mid x_k) =$$

$$= -(0.4 \cdot \log 0.8 + 0.1 \cdot \log 0.2 + 0 \cdot \log 0 +$$

$$+ 0 \cdot \log 0 + 0.2 \cdot \log 0.67 + 0.1 \cdot \log 0.33 +$$

$$+ 0 \cdot \log 0 + 0 \cdot \log 0 + 0.2 \cdot \log 1) = 0.635$$

$$H(X,Y) = -\sum_{k=1}^{m}\sum_{j=1}^{s} p(x_k, y_j)\log p(x_k, y_j) =$$

$$= -(0.4 \cdot \log 0.4 + 0.1 \cdot \log 0.1 + 0 \cdot \log 0 +$$

$$+ 0 \cdot \log 0 + 0.2 \cdot \log 0.2 + 0.1 \cdot \log 0.1 +$$

$$+ 0 \cdot \log 0 + 0 \cdot \log 0 + 0.2 \cdot \log 0.2) = 2.12$$

$$H(X,Y) = H(X) + H(Y \mid X) = 1.485 + 0.635 = 2.12$$

2.6 Относительная энтропия и избыточность сообщения

С практической точки зрения оценка количества информации необходима

- **для построения экономичных кодов**
- **для оценки свойств каналов связи и их пропускной способности,**
- **для определения избыточности кодов и повышения их помехоустойчивости**

- **Максимальное количество информации на элемент сообщения может быть получено только в случае равновероятных и независимых сообщений.**

- **Сообщения, энтропия которых равна максимальному значению являются оптимальными сообщениями в смысле наибольшего количества передаваемой информации.**
- **Реальные сообщения редко полностью удовлетворяют этому условию, поэтому информационная нагрузка на каждый элемент обычно меньше той, которую они могли бы передавать.**

- **Энтропия таких сообщений меньше максимальной и сообщение обладает информационной избыточностью.**
- **В теории информации избыточность показывает количество «лишней информации», которая определяется структурой множества состояний элементов и обычно заранее известна из статистических данных.**

- *По рзелульаттам илссеовадний одонго анлигйсокго унвиертисета, не иеемт занчнеия, в кокам пряокде рсапожолены бкувы в солве. Галвоне, чотбы преавя и пслоендяя бквуы блыи на мсете. Осатьлыне бкувы мгоут селдовтаь в плоонм бсепордяке, все-рвано ткест чтаитсея без побрелм. Пичрионй эгото являтеся то, что мы чиатем не кдаужю бкуву по отдльенотси, а все солво цликеом.*

- **Мерой количественной оценки того, насколько данное реальное сообщение отличается от соответствующего ему оптимального сообщения, служит**

относительная энтропия

$$\mu = H / H_{\max}$$

- **Наряду с коэффициентом сжатия используется и величина *избыточности сообщения***

$$r = 1 - \mu$$

- **Для уменьшения избыточности сообщения необходимо увеличить энтропию сообщения, т.е. стремиться к тому, что элементы сообщения были максимально информативны.**
- **Для этой цели используются методы кодирования информации.**

- **Величина** $H_L = \frac{1}{L} \cdot \sum_{x \in A^L} P(x) \log P(x)$

  называется *энтропией L-той степени,*
  где A^L **- множество всех сообщений источника длины *L* в алфавите *A* .**

- **Обозначим** $H_\infty = \lim\limits_{L\to\infty} H_L$

- **Эту величину называют *предельной энтропией источника*.**
- **Показано, что для стационарного бернуллиевского источника**

$$H_\infty = H(p_1, \ldots, p_n)$$

Пример

- Для русского языка, состоящего из 32 букв (буквы «е» и «ё», «ь» и «ъ» не различаются, добавлен символ пробела « »), максимальное значение энтропии при условии равновероятности букв составляет

$$H_{\max} = \log m = \log 32 = 5 \text{ бит}$$

Частоты появления отдельных букв русского языка

пробел	**о**	**Е, Ё**	**а**	**и**	**т**	**н**	**с**
0.175	0.090	0.072	0.062	0.062	0.053	0.045	0.045
р	**в**	**л**	**к**	**м**	**д**	**п**	**у**
0.040	0.038	0.035	0.028	0.026	0.025	0.023	0.021
я	**ы**	**з**	**Ь,Ъ**	**б**	**г**	**ч**	**й**
0.018	0.016	0.016	0.014	0.014	0.013	0.012	0.010
х	**ж**	**ю**	**ш**	**ц**	**щ**	**э**	**ф**
0.009	0.007	0.006	0.006	0.004	0.003	0.003	0.002

Используя эти частоты в качестве вероятностей появления букв, можно получить приближенное значение энтропии одной буквы русского языка $H_1 = 4.35$ **бит,**
что меньше максимального значения энтропии.

Если учитывать статистику появления буквенных сочетаний и словесных сочетаний, то исследования показали, что энтропия на букву русского языка не превышает 2 бит

- **Таким образом, коэффициент сжатия или относительная энтропия для русского языка составляет**

$$\mu = \frac{H}{H_{\max}} \le \frac{2}{\log 32} = 0.4$$

- **а величина избыточности**

$$r = 1 - \mu \ge 1 - 0.4 = 0.6$$