

1 Введение

Конспект лекций и практики по Теории информации.

В формулах матрицы обозначаются прописной буквой. Строчной буквой обозначаются скаляры. Если строчная буква выделена жирным, то это вектор. Бывают исключения.

2 Измерение информации. Энтропия

2.1 Общая схема системы связи

Общая схема системы связи:

source -> source -> channel -> channel -> channel -> source -> user
 encoder encoder decoder decoder

Источник информации порождает сообщение. Задача кодировщика представить сообщение в наиболее компактной форме. Кодер канала отвечает за защиту сообщения от помех в канале связи. На схеме нет модулятора, который переводит цифровой сигнал в форму, соответствующую физической среде передачи. В данном курсе будет рассматриваться кодирование дискретных источников (source encoder).

Под каналом может пониматься не только передача (кабель), но и хранение (жесткий диск).

2.2 Практическое применение

Практическое применение: архивирование данных (ZIP, RAR, 7-Zip), сжатие речи, сжатие звука, сжатие изображений, сжатие видео.

Архивация данных подразумевает сжатие без потерь. Остальные виды сжатия допускают потери (уменьшается качество).

Пропускная способность сетей растет медленней, чем объем данных, который нужно передвигать. Поэтому актуальность сжатия данных сохраняется.

2.3 Дискретные ансамбли

Дискретное множество содержащее конечное число элементарных событий:

$$X = \{x\}.$$

Множество чисел, которое задает **распределение вероятностей**:

$$\{p(x)\}, p(x) > 0, \sum p(x) = 1,$$

где $p(x)$ — вероятность исхода x .

Дискретный ансамбль:

$$X = \{x, p(x)\}.$$

Множество всевозможных подмножеств X :

$$\Omega = \{A\}.$$

Вероятность сложного события A :

$$P(A) = \sum_{x \in A} p(x), \quad A \in \Omega.$$

Элементарные события несовместны, поэтому вероятности просто складываются.

Вероятность совместного наступления зависимых событий:

$$P(AB) = P(B) \cdot P(A | B).$$

В общем случае:

$$P(A_1 \cdots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1A_2) \cdots .$$

Сама зависимость одного события от другого выражается **формулой условной вероятности**: $P(A | B) = \frac{P(AB)}{P(B)}$. Если $P(B) = 0$, то $P(A | B) = 0$.

Пример зависимого события: достать черный шар из корзины, если из нее уже достали белый шар.

События $A, B \in \Omega$ независимы, если их пересечение:

$$P(AB) = P(A)P(B).$$

Вероятности таких событий:

$$P(A | B) = P(A), \quad P(B | A) = P(B).$$

Вероятность объединения событий (произоидет хотя бы одно событие, учитывается совместный исход):

$$P(A \cup B) = P(A) + P(B) - P(AB),$$

$$P\left(\bigcup_{m=1}^M A_m\right) \leq \sum_{m=1}^M P(A_m).$$

Рассмотрим графы с вероятностями переходов:

$$\begin{array}{ccccc} & p & & p & \\ A & \rightarrow & B & \rightarrow & C \end{array}$$

Нужно 2 перехода, поэтому вероятность перейти из A в C : p^2 .

$$\begin{array}{ccccc} & p & & p & \\ A & \rightarrow & B & \rightarrow & C \\ | & & & & \wedge \\ \text{'-----'} & & & & \\ & p & & & \end{array}$$

Пути 2, но переход в C совершится только по одному: $P(AC) = p^2 + p - p^3$. Для

проверки можно взять $p = 1$.

```

      p      p
A -> B -> C
|           ^
|      p    |
|-----'
|           ^
|--> D ---'
      p      p

```

Такой случай сложный, но его можно легко посчитать по формуле включений-исключений:

$$\begin{aligned}
 P(AC) &= P(ABC) \cup P(AC) \cup P(ADC) = \\
 &= P(ABC) + P(AC) + P(ADC) - \\
 &- P(ABC) \cap P(AC) - P(ABC) \cap P(ADC) - P(AC) \cap P(ADC) + \\
 &+ P(ABC) \cap P(AC) \cap P(ADC) = 2p^2 + p - 2p^3 - p^4 + p^5.
 \end{aligned}$$

Пусть даны M несовместных событий H_1, \dots, H_M , таких что $P(\bigcup_{m=1}^M H_m) = 1$. Тогда вероятность произвольного события A считается по формуле **полной вероятности**:

$$P(A) = \sum_{m=1}^M P(A | H_m)P(H_m).$$

Формула **апостериорной вероятности (Байеса)**:

$$P(H_j | A) = \frac{P(AH_j)}{P(A)} = \frac{P(A | H_j)P(H_j)}{\sum_{m=1}^M P(A | H_m)P(H_m)}.$$

Вероятность того, что предпосылка произошла, если произошло следствие.

Пусть у первого грузчика 200 коробок, у второго — 300, у третьего — 700. Вероятность, что первый разобьет коробку 0,1, второй — 0,2, третий — 0,9. Найти вероятность, что коробка разбита вторым грузчиком. Пусть A — коробка разбита, H — коробку нес грузчик:

$$P(H_2 | A) = \frac{0,2 \cdot \frac{300}{1200}}{0,1 \cdot \frac{200}{1200} + 0,2 \cdot \frac{300}{1200} + 0,9 \cdot \frac{700}{1200}} = \frac{0,05}{0,592} \approx 0,084.$$

Произведение ансамблей $X = \{x, p_X(x)\}$ и $Y = \{y, p_Y(y)\}$:

$$XY = \{(x, y), p_{XY}(x, y)\}.$$

Условное распределение вероятностей:

$$p(x | y) = \begin{cases} \frac{p(x, y)}{p(y)} & \text{if } p(y) \neq 0, \\ 0 & \text{иначе.} \end{cases}$$

Ансамбли X и Y независимы, если

$$p(x, y) = p(x)p(y), \quad x \in X, \quad y \in Y.$$

2.4 Измерение информации

Интуитивный подход: количество информации равняется затратам, необходимым для передачи или хранения данных (буква текста равняется 8 битам).

Информацию логично измерять через вероятность. Если одно событие происходит регулярно, то оно перестает быть информативным. Редкие события несут много информации.

Пусть $X = \{x, p(x)\}$ — ансамбль, $\mu(x)$ — мера (количество) информации в x .

- $\mu(x) \geq 0$.
- $\mu(x)$ — функция от $p(x)$.
- Монотонность: если $x, y \in X$, $p(x) \geq p(y)$, тогда $\mu(x) \leq \mu(y)$.
- Аддитивность: если x и y независимы, тогда $\mu(x, y) = \mu(x) + \mu(y)$.
- $\mu(p(x)^k) = k\mu(p(x))$: если происходит серия одного события, то информация каждого складывается.

Из этого следует формула **собственной информации** в событии x :

$$I(x) = -\log p(x), \quad x \in X.$$

Основание логарифма может быть любым. Если взять 2, то информация будет измеряться в битах.

Энтропия — средняя информативность ансамбля:

$$H(X) = E[I(x)] = \sum_x p(x) I(x) = - \sum_x p(x) \log p(x).$$

Свойства энтропии:

1. $H(X) \geq 0$.
2. $H(X) \leq \log |X|$. Логарифм от количества сообщений. Равенство, если все элементы X равновероятны.

Доказательство

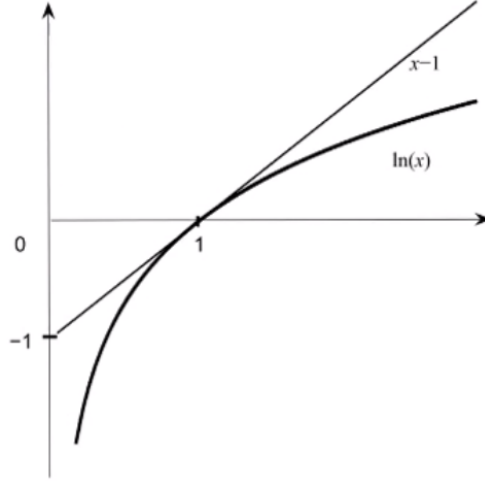
$$\begin{aligned} H(X) - \log |X| &= - \sum_{x \in X} p(x) \log p(x) - \sum_{x \in X} p(x) \log |X| = \\ &= \sum_{x \in X} p(x) \log \frac{1}{p(x)|X|} \leq \\ &\leq \log e \left[\sum_{x \in X} p(x) \left(\frac{1}{p(x)|X|} - 1 \right) \right] = \\ &= \log e \left(\sum_{x \in X} \frac{1}{|X|} - \sum_{x \in X} p(x) \right) = 0 \end{aligned}$$

.

На первом шаге пользуемся тем, что сумма вероятностей равна 1. На втором шаге собираем логарифм и переворачиваем его число, чтобы избавиться от

минуса. На третьем шаге используем свойство логарифма:

$$\begin{aligned}\ln x &\leq x - 1 \\ e^{\ln x} &= e^{x-1} \\ x &= e^{x-1} \\ \log x &= \log e^{x-1} = (x-1) \log e\end{aligned}$$



чтобы преобразовать логарифм, затем вынесем общий множитель $\log e$ за сумму. На четвертом шаге раскроем круглые скобки и заменим одну сумму двумя. Первая сумма равна 0, так как суммируем $|X|$ раз.

3. $X = \{x, p(x)\}$ и $Y = \{y = f(x), p(y)\}$, тогда $H(Y) \leq H(X)$, с равенством, если f обратима.
4. Если X и Y независимы, тогда $H(XY) = H(X) + H(Y)$.
5. $H(X)$ — выпуклая вверх функция распределения вероятностей на элементах ансамбля X .
6. Пусть $X = \{x, p(x)\}$ и $A \subseteq X$. Введем ансамбль $X' = \{x, p'(x)\}$ и $p'(x)$ как:

$$p'(x) = \begin{cases} \frac{P(A)}{|A|}, & x \in A, \\ p(x), & x \notin A \end{cases}$$

тогда $H(X') \geq H(X)$. В первом кейсе усреднение части ансамбля.

7. Если для двух ансамблей X и Y распределения вероятностей отличаются только порядком следования элементов, то $H(X) = H(Y)$.

3 Выпуклые функции. Условная энтропия. Стационарные источники

3.1 Выпуклые функции

Множество вещественных векторов $\mathbb{R} = \{x\}$ выпукло, если $\forall x, x' \in \mathbb{R}$ и $\forall \alpha \in [0, 1]$ вектор $y = \alpha x + (1 - \alpha)x'$ принадлежит \mathbb{R} . Можно представить вектор в виде

точки. Тогда множество называется выпуклым, если любые две точки множества задают отрезок, все точки которого принадлежат этому множеству.

Теорема Множество вероятностных векторов длины M выпукло.

Доказательство Рассмотрим два распределения вероятностей $p = \{p_1, \dots, p_M\}$ и $p' = \{p'_1, \dots, p'_M\}$. Рассмотрим векторы вида

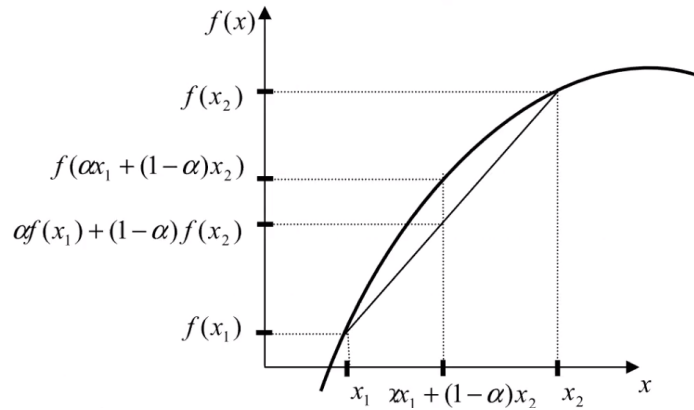
$$q = \alpha p + (1 - \alpha)p',$$

где $q_i \geq 0$, а сумма компонент:

$$\sum_{i=1}^M q_i = \alpha \sum_{i=1}^M p_i + (1 - \alpha) \sum_{i=1}^M p'_i = \alpha + 1 - \alpha = 1.$$

Функция $f(x)$ выпуклая, если $\forall x, x' \in R$, где R — выпуклая область, и $\forall \alpha \in [0, 1]$ выполняется неравенство: $f(\alpha x + (1 - \alpha)x') \geq \alpha f(x) + (1 - \alpha)f(x')$.

Доказательство



Значение функции на прямой выводится через уравнение прямой по двум точкам.

Для случая функции одной переменной:

$$f(\alpha x_1 + (1 - \alpha)x_2) \geq \alpha f(x_1) + (1 - \alpha)f(x_2).$$

Если функция выпукла вниз, то неравенство меняется в обратную сторону.

Теорема Пусть $f(x)$ — выпуклая функция вектора x , определенная на выпуклой области R , и пусть константы $a_1, \dots, a_M \in [0, 1]$ такие, что $\sum_{m=1}^M a_m = 1$. Тогда $\forall x_1, \dots, x_M \in R$ справедливо неравенство:

$$f\left(\sum_{m=1}^M a_m x_m\right) \geq \sum_{m=1}^M a_m f(x_m).$$

Если a_m означает вероятность x_m , то получим **неравенство Йенсена** (функция от мат. ожидания больше либо равна мат. ожидания от функции):

$$f(E\{x\}) \geq E\{f(x)\}.$$

Свойства выпуклых функций:

- сумма выпуклых функций выпукла,
- произведение вып. функции и положительной константы является вып. функцией,
- линейная комбинация (сумма с коэффициентами) с неотрицательными коэффициентами — выпуклая функция (обобщение первого свойства).

Доказательство 5 свойства энтропии.

$$H(p) = - \sum_{m=1}^M p_m \log p_m$$

Рассмотрим слагаемые $f_m(p) = -p_m \log p_m$. Если $f_m(p)$ выпуклая, то их сумма будет выпуклой функцией.

Вторая производная $f_m''(p) = \frac{-(\log e)}{p_m}$. $f_m''(p) < 0 \forall p_m \in (0, 1)$.

Пример $X = \{0, 1\}$, $p(1) = p$, $p(0) = 1 - p = q$. Энтропия двоичного ансамбля:

$$H(X) = -p \log p - q \log q \triangleq \eta(p)$$

Последнее равенство означает равенство по определению. Просто ввели новое обозначение для двоичной энтропии.

$$\begin{aligned}\eta'(p) &= -\log p + \log(1 - p) \\ \eta''(p) &= -\log \frac{e}{p} - \log \frac{e}{1 - p} < 0\end{aligned}$$

Следовательно, $H(X)$ выпуклая. Максимум энтропии равен 1 (при основании логарифма 2).

Рассмотрим случай, когда $p = 0$:

$$H(X) = -0 \times -\infty - 1 \times 0.$$

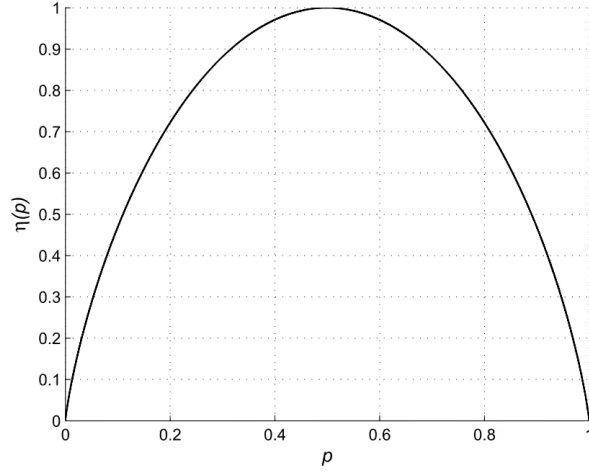
Неопределенность первого слагаемого решим с помощью правила Лопиталя:

$$\lim_{p \rightarrow 0} -\frac{\log p}{p^{-1}} = \lim_{p \rightarrow 0} \frac{p^2}{p \ln a} = \lim_{p \rightarrow 0} \frac{p}{\ln a} = 0$$

Получаем, что $H(X) = 0$.

Если $p = 1$, то аналогичными рассуждениями доказывается, что $H(X) = 0$.

Не сложно получить график энтропии двоичного ансамбля:



В Теории информации битами измеряется среднее количество информации в источнике. Может получиться, что в среднем в одном символе ансамбля или одном сообщении источника меньше одного бита информации. В классическом, инженерном смысле бит является единицей информации, но в данном случае это не так.

Видно, что чем выше неопределенность событий (чем меньше разность между их вероятностями), тем больше их энтропия, т.е. средняя информативность.

Доказательство свойства 6 энтропии.

Обозначим $\tilde{p} = (\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}, p_3, \dots, p_M)$

Докажем, что $H(\tilde{p}) \geq H(p)$

$$p' = p = (p_1, p_2, p_3, \dots, p_M)$$

$$p'' = (p_2, p_1, p_3, \dots, p_M)$$

Заметим, что: $H(p') = H(p'') = H(p)$ $\tilde{p} = \frac{p' + p''}{2}$

Из выпуклости энтропии следует, что

$$H(\tilde{p}) = H\left(\frac{p' + p''}{2}\right) \geq \frac{1}{2}H(p') + \frac{1}{2}H(p'') = H(p)$$

Равенство будет, если события равновероятны.

Условная собственная информация сообщения x при фиксированном y :

$$I(x | y) = -\log p(x | y)$$

Условная энтропия X при заданном $y \in Y$:

$$H(X | y) = - \sum_{x \in X} p(x | y) \log p(x | y)$$

Условная энтропия X при фиксированном ансамбле Y :

$$H(X | Y) = - \sum_{y \in Y} \left(p(y) \sum_{x \in X} p(x | y) \log p(x | y) \right) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x | y)$$

Свойства условной энтропии:

1. $H(X | Y) \geq 0$
2. $H(X | Y) \leq H(X)$, равенство, если X и Y независимы

Доказательство:

$$\begin{aligned} H(X | Y) - H(X) &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x | y) + \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x) = \\ &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x)}{p(x | y)} \leq \\ &\leq \sum_{x \in X} \sum_{y \in Y} p(x, y) \left(\frac{p(x)}{p(x | y)} - 1 \right) \log e = \\ &= \left(\sum_{x \in X} \sum_{y \in Y} p(y) p(x) - \sum_{x \in X} \sum_{y \in Y} p(x, y) \right) \log e = 0 \end{aligned}$$

На первом шаге $H(X)$ заменили ее формулой, затем расписали $p(x)$ по формуле полной вероятности как $\sum_{y \in Y} p(y) p(x | y) = \sum_{y \in Y} p(x, y)$.

На третьем шаге нужно вспомнить доказательство 2 свойства энтропии.

3. $H(XY) = H(X) + H(Y | X) = H(Y) + H(X | Y)$
4. $H(X | YZ) \leq H(X | Y)$, равенство, если X и Z условно независимы для всех $y \in Y$. Чем больше условий, тем меньше информации.
5. $H(X_1 \cdots X_n) = H(X_1) + H(X_2 | X_1) + H(X_3 | X_1 X_2) + \cdots + H(X_n | X_1, \cdots, X_{n-1})$.
В качестве $X_1 \cdots X_n$ может быть текст длинны n , где каждую букву определяет ансамбль.
6. $H(X_1 \cdots X_n) \leq \sum_{i=1}^n H(X_i)$ равенство, если X_1, \cdots, X_n являются совместно независимыми.

Пусть загадано слово. Если угадывать первую букву в этом слове, то она может быть почти любой. Если нам известно начало слова, например, МОСКВ, то становится очевидней, какая буква следующая. Поэтому энтропия первых букв больше, чем последних. Энтропию первой буквы можно обозначить $H(X)$, второй — $H(Y | X)$ и т.д.

Доказательство 3 свойства энтропии:

Используя свойство 3 условной энтропии:

$$H(XY) = \underbrace{H(X | Y)}_{\geq 0} + H(Y) = \underbrace{H(Y | X)}_{=0} + H(X)$$

Поскольку $f(x)$ определена для каждого x , получим, что $H(Y | X) = 0$, $H(X | Y) \geq 0$

Так как знаем функцию, неопределенность Y исчезает. Но функция может возвращать одно значение для разных x , поэтому для X имеется неопределенность.

3.2 Дискретные источники

Дискретный источник — это устройство, которое в каждый момент времени выбирает одно сообщение из дискретного множества. Например, букву из алфавита.

Если множество значений времени также дискретно, то источник называется **дискретным по времени**.

Источник считается заданным, если известна его вероятностная модель. Должна быть определена вероятностная модель случайного процесса генерирования случайных сообщений на выходе источника.

Случайный процесс (также называемый стохастическим) — набор случайных величин, проиндексированных множеством, которое часто обозначает разные моменты времени. Множество индексов может быть или множеством натуральных чисел (случайный процесс с дискретным временем), или множеством вещественных чисел (случайный процесс с непрерывным временем). Случайные величины могут иметь непрерывное или дискретное **пространство состояний** (пространство возможных результатов в каждый момент времени).

Дискретный источник задан, если для $n = 1, 2, \dots$ (длина угадываемого слова) и $i = 0, 1, 2, \dots$ (номер угадываемого слова) известна вероятность $p(x_{i+1}, x_{i+2}, \dots, x_{i+n})$ случайной последовательности из $\{X_{i+1}X_{i+2} \dots X_{i+n}\}$, которая начинается с индекса $i + 1$ и имеет длину n , где $x_j \in X_j$, $j = i + 1, \dots, i + n$. Обычно рассматривается случай, когда $X_j = X$ для всех j , то есть алфавит для каждого символа в одном сообщении фиксированный.

Пусть $x = (x_1, x_2, \dots, x_n, \dots)$ — дискретный случайный процесс (генерируемая последовательность букв). Рассмотрим случайный вектор $x_{j+1}^{j+n} = (x_{j+1}, x_{j+2}, \dots, x_{j+n})$ (окно на генерируемой последовательности с длиной n и смещением j от начала, слово).

Стационарность процесса означает, что для любого n и j , $p(x_{j+1}^{j+n})$ не зависит от сдвига во времени j , т.е., $p(x_{j+1}^{j+n}) = p(x_1^n)$.

Источник называют **дискретным без памяти**, если (следующая буква не зависит от предыдущей)

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i)$$

3.2.1 Цепь Маркова

Дискретный случайный процесс называется **цепью Маркова** порядка s , если для любого n и $x = (x_1, \dots, x_n) \in X^n$ выполняется следующее равенство:

$$p(x) = p(x_1, \dots, x_s)p(x_{s+1} | x_1, \dots, x_s)p(x_{s+2} | x_2, \dots, x_{s+1}) \dots p(x_n | x_{n-s}, \dots, x_{n-1})$$

Другими словами, первые s букв независимы, затем каждая следующая буква зависит от s предыдущих.

Для цепи Маркова справедливо равенство:

$$p(x_n | x_1, \dots, x_{n-1}) = p(x_n | x_{n-s}, \dots, x_{n-1})$$

x_t — состояние марковской цепи в момент времени t . Марковская цепь порядка s задается начальным распределением вероятностей первых s значений (состояний) и условными вероятностями вида $p(x_n | x_{n-s}, \dots, x_{n-1})$ для всевозможных последовательностей x_{n-s}, \dots, x_{n-1} . Если эти условные вероятности не изменяются при сдвиге во времени, то марковская цепь называется **однородной**.

Чтобы определить условную вероятность буквы, нужно знать s предыдущих букв. Таким образом порядок марковской цепи определяет память источника.

Марковская цепь порядка $s = 1$ с состояниями $X = \{0, 1, \dots, M-1\}$ определяется начальным распределением $\{p(x_1), x_1 \in X\}$ и условными вероятностями

$$\pi_{ij} = P(x_t = j | x_{t-1} = i), \quad i, j = 0, 1, \dots, M-1$$

Матрица переходных вероятностей (вероятность, что текущее состояние j , а предыдущее было i):

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} & \dots & \pi_{0,M-1} \\ \pi_{10} & \pi_{11} & \dots & \pi_{1,M-1} \\ \dots & \dots & \dots & \dots \\ \pi_{M-1,0} & \pi_{M-1,1} & \dots & \pi_{M-1,M-1} \end{pmatrix}$$

Обозначим через $p_t = (p_t(0), \dots, p_t(M-1))$ стохастический вектор, компоненты которого — вероятности состояний цепи Маркова в момент времени t , где $p_t(i)$, $i = 0, 1, \dots, M-1$ — вероятность состояния i в момент времени t .

Из формулы полной вероятности следует:

$$p_{t+1}(i) = \sum_{j=0}^{M-1} p_t(j) \pi_{ji}$$

В матричном виде:

$$\mathbf{p}_{t+1} = \mathbf{p}_t \Pi$$

Для произвольного числа шагов n :

$$\mathbf{p}_{t+n} = \mathbf{p}_t \Pi^n$$

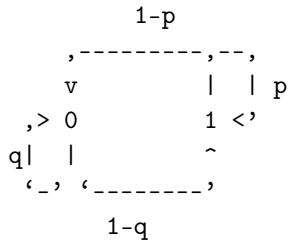
Из формулы следует, что распределение вероятностей в момент времени t зависит от величины t и от начального распределения p_1 . Отсюда следует, что в общем случае рассматриваемый случайный процесс нестационарен. Однако, если существует стохастический вектор p , так что

$$\mathbf{p} = \mathbf{p} \Pi$$

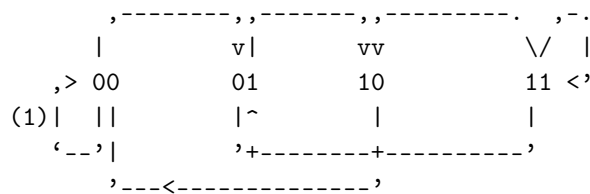
то выбрав $\mathbf{p}_1 = \mathbf{p}$ мы получим стационарный процесс. Вектор \mathbf{p} , удовлетворяю-

ций называется **стационарным распределением вероятностей** для марковской цепи с матрицей переходных вероятностей Π .

Рассмотрим марковскую цепь с 2 состояниями в качестве источника с памятью 1:



Увеличим память до 2:



В кружках написаны последние 2 символа, так как каждый следующий зависит от предыдущих 2

Энтропия символа x_t (из ансамбля $X_t = X$) сгенерированного в момент времени t не зависит от t ($H(X_t) = H(X)$) и называется **одномерной энтропией источника** (процесса). Обозначим ее как $H_1(X)$.

$H_1(X)$ не учитывает зависимость между символами, порожденными источником. Это не относится к марковскому источнику.

3.3 Энтропия на сообщение

Рассмотрим $x = (x_1, x_2, \dots, x_n)$ из $X_1 X_2 \dots X_n = X^n$. Энтропия $H(X_1 X_2 \dots X_n) = H(X^n)$ называется **n-мерной энтропией** процесса.

Энтропия на символ для последовательности длины n определяется как:

$$H_n(X) = \frac{H(X^n)}{n},$$

То есть берем всю последовательность, сгенерированную источником, и делим ее на блоки длины n . Делим энтропию блока на количество символов в нем.

другой способ:

$$H(X_n | X_1, \dots, X_{n-1}) = H(X | X^{n-1}).$$

Энтропия на сообщение:

$$\lim_{n \rightarrow \infty} H_n(X) = \lim_{n \rightarrow \infty} H(X | X^{n-1})$$

Проблема блочного кодирования в том, что первые символы в блоке будут плохо кодироваться, так как мало информации о предыдущих символах. Разумно не использовать маленький размер блока.

Для дискретного стационарного процесса (источника)

1. $H(X | X^n)$ не возрастает с увеличением n .

Следует из невозрастания энтропии с увеличением числа условий.

2. $H_n(X)$ не возрастает с увеличением n .

$$\begin{aligned} H(X^{n+1}) &= H(X_1 \cdots X_n X_{n+1}) = \\ &= H(X_1 \cdots X_n) + H(X_{n+1} | X_1, \dots, X_n) \leq \\ &\leq nH_n(X) + H_n(X) = \\ &= (n+1)H_n(X). \end{aligned}$$

На третьем шаге для второго слагаемого из предыдущего шага используется свойство 3.

$$H(X^{n+1}) \leq (n+1)H_n(X) \implies \frac{H(X^{n+1})}{n+1} = H_{n+1}(X) \leq H_n(X)$$

3. $H_n(X) \geq H(X | X^{n-1})$

$$H(X^n) = H(X) + H(X | X^1) + \cdots + H(X | X^{n-1}) \geq nH(X | X^{n-1}) \geq nH(X | X^n)$$

4. $\lim_{n \rightarrow \infty} H_n(X) = \lim_{n \rightarrow \infty} H(X | X^n)$

Доказательство

Из 3 свойства следует, что $\lim_{n \rightarrow \infty} H_n(X) \geq \lim_{n \rightarrow \infty} H(X | X^n)$

Для $m < n$:

$$\begin{aligned} H(X^n) &= H(X_1 \cdots X_n) = \\ &= H(X_1 \cdots X_m) + H(X_{m+1} | X_1, \dots, X_m) + \cdots + H(X_n | X_1, \dots, X_{n-1}) \leq \\ &\leq mH_m(X) + (n-m)H(X | X^m) \end{aligned}$$

После деления обеих частей на n :

$$\lim_{n \rightarrow \infty} H_n(X) \leq H(X | X^m) \text{ для любого } m.$$

Справа от неравенства:

$$\lim_{n \rightarrow \infty} \frac{mH_m(X)}{n} + \frac{nH(X | X^m)}{n} - \frac{H(X | X^m)}{m} \frac{m}{n} \xrightarrow{0} 0$$

Устремляем $m \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} H_n(X) \leq \lim_{m \rightarrow \infty} H(X | X^m)$$

4 Энтропия на сообщение дискретного источника. Префиксные коды

4.1 Энтропия на сообщение

Обозначим $H_\infty(X) = \lim_{n \rightarrow \infty} H_n(X)$, $H(X | X^\infty) = \lim_{n \rightarrow \infty} H(X | X^n)$, тогда

$$H_\infty(X) = H(X | X^\infty)$$

Два способа кодирования:

- Расширение алфавита: буквы становятся последовательностями исходных букв длины n .
- Учет зависимости текущей буквы от n предшествующих букв.

Энтропии для дискретного стационарного источника без памяти:

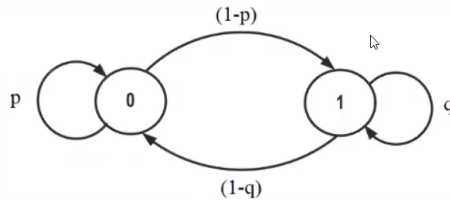
- $H(X_1 \cdots X_n) = H(X_1) + \cdots + H(X_n)$ (нет памяти)
- $H(X^n) = nH(X)$ (один алфавит)
- $H_n(X) = H(X)$ (из определения)
- $H_\infty(X) = H(X)$
- $H(X | X^n) = H(X_{n+1} | X_1, \cdots, X_n) = H(X)$,
- $H(X | X^\infty) = H(X)$

Отсюда не следует, что для такого источника нужно кодировать каждую букву независимо от других.

Энтропии для марковского источника:

- $H(X | X^n) = H(X_{n+1} | X_1, \cdots, X_n) = H(X_{n+1} | X_{n-s+1}, \cdots, X_n) = H(X | X^s)$
- $H(X | X^\infty) = H(X | X^s)$
- $H(X^n) = H(X_1 \cdots X_s X_{s+1} \cdots X_n)$
 $= H(X_1 \cdots X_s) + H(X_{s+1} \cdots X_n | X_1, \cdots, X_s)$
- $H(X_{s+1} \cdots X_n | X_1, \cdots, X_s) = H(X_{s+1} | X_1, \cdots, X_s) + H(X_{s+2} | X_2, \cdots, X_{s+1}) + \cdots + H(X_n | X_{n-s}, \cdots, X_{n-1}) = (n-s)H(X | X^s)$
- $\frac{H(X^n)}{n} = \frac{sH_s(X)}{n} + \frac{(n-s)H(X|X^s)}{n}$

Рассмотрим марковский источник (состояния и вероятности переходов):



Выведем энтропию на символ:

$$H(X) = -\pi_0 \log \pi_0 - \pi_1 \log \pi_1$$

Здесь обозначили вероятность, что находимся в состоянии 0, как π_0 . Найдём π_0 и π_1 :

$$\begin{aligned}\pi_0 &= \pi_0 p + \pi_1 (1 - q) \\ \pi_1 &= 1 - \pi_0 \\ \pi_0 &= \pi_0 p + (1 - \pi_0) - q(1 - \pi_0) \\ \pi_0 - \pi_0 p + \pi_0 - \pi_0 q &= 1 - q \\ \pi_0 &= \frac{1 - q}{2 - p - q} \\ \pi_1 &= \frac{1 - p}{2 - p - q}\end{aligned}$$

Если цепь будет большой, то аналитический способ нахождения вероятностей может оказаться слишком сложным. Найдём стационарное распределение путем моделирования. Произвольно зададим начальное распределение p_1 , например, пусть мы находимся в состоянии 1: $p_1 = (0, 1)$. Матрица переходных вероятностей для рассматриваемой цепи:

$$\Pi = \begin{pmatrix} p & 1 - p \\ 1 - q & q \end{pmatrix}$$

Программа:

```

p = 0.25
q = 0.35
P = [[p, 1 - p], [1 - q, q]]

p1 = [0, 1]

pn = [p1]
for _ in range(10):
    p1 = [p1[0] * P[0][0] + p1[1] * P[1][0], \
          p1[0] * P[0][1] + p1[1] * P[1][1]]
    pn.append(p1)

import matplotlib.pyplot as plot

```

```

plot.plot(range(11), [p[0] for p in pn])
plot.plot(range(11), [p[1] for p in pn])
plot.show()

```

Конечное распределение будет приближено к стационарному (или тому, которое нашли аналитическим путем).

Надем условную энтропию. Для этого воспользуемся формулой, приведенной ранее:

$$\begin{aligned}
 H(X|X) &= -[\pi_0(p(0|0)\log p + p(1|0)\log p) + \pi_1(p(0|1)\log p + p(1|1)\log p)] = \\
 &= -[\pi_0(p\log p + (1-p)\log(1-p)) + \pi_1((1-q)\log(1-q) + q\log q)] = \\
 &= \pi_0\eta(p) + \pi_1\eta(q)
 \end{aligned}$$

Условные вероятности определяются вероятностями переходов. Вероятность нахождения в определенном состоянии определяется из стационарного распределения.

Найдем энтропию на символ из блока. Наш источник имеет память 1, поэтому условная энтропия не может быть глубже 1. Если бы у источника не было памяти, то в качестве энтропии на символ, мы бы просто взяли $H(X)$.

$$H_2(X) = \frac{H(X) + H(X|X)}{2} \quad H_3(X) = \frac{H(X) + 2H(X|X)}{3}$$

Просто $H(X)$ игнорирует модель источника. Это самый плохой случай. $H(X|X)$ — меньшая и лучшая величина, лучшая так как она учитывает модель. Чем больше у нас размер блока, тем меньше вклад худшей величины и тем лучше значение энтропии на символ.

4.2 Неравномерное побуквенное кодирование

Рассмотрим дискретный источник без памяти: $X = \{1, \dots, M\}$ (источник генерирует числа от 1 до M), $\{p_1, \dots, p_M\}$ (соответствующие вероятности). $C = \{c_1, \dots, c_M\}$ — кодовые слова длины l_1, \dots, l_M .

Средняя длина кодового слова:

$$\bar{l} = E[l_i] = \sum_{i=1}^M p_i l_i$$

$H(X)$ — нижняя граница для \bar{l} (доказательство ниже).

Длины кодовых слов могут отличаться, поэтому кодирование неравномерное (variable length coding).

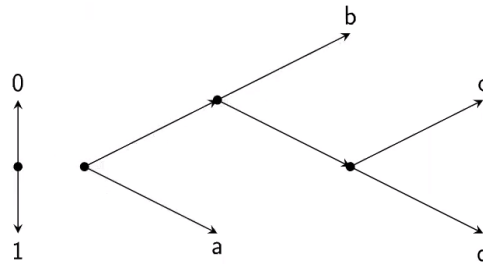
Рассмотрим код Морзе:

Для однозначного декодирования кода необходимы разделители (паузы) между

Буква	Кодовое слово
e	.
a	. -
j	. - - -
q	- - . -

кодовыми словами.

Рассмотрим двоичное кодовое дерево:



Буква	Кодовое слово
a	1
b	00
c	010
d	011

Такой код называется **префиксным**. Здесь нет проблемы, которая была в коде Морзе. Вводить дополнительный символ-разделитель не требуется.

Свойства префиксного кода:

- ни одно кодовое слово не является началом другого кодового слова,
- префиксный код является однозначно декодируемым,
- если только листья двоичного дерева соответствуют кодовым словам, то код является префиксным,
- однозначно декодируемый код не обязательно является префиксным,
- древовидный код является префиксным.

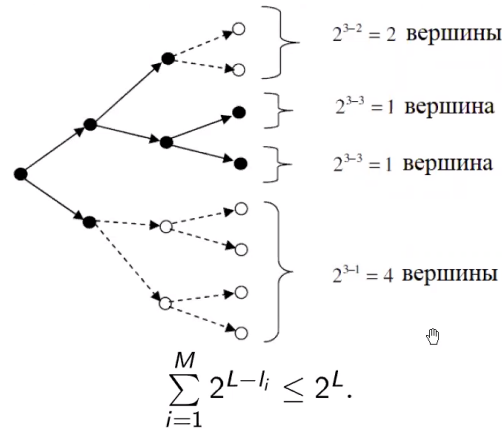
Необходимое и достаточное условие существования префиксного кода (неравенство Крафта):

$$\sum_{i=1}^M 2^{-l_i} \leq 1$$

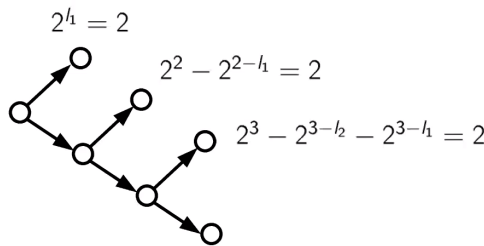
Доказательство необходимости Пусть $L \geq \max l_i$, тогда если достроить двоичное дерево до полного, справедливо неравенство:

$$\sum_{i=1}^M 2^{L-l_i} \leq 2^L.$$

Меньше, если одну букву удалили. Если поделить на правую часть, то получим неравенство Крафта.



Доказательство достаточности Длину кодового слова определяет уровень буквы в двоичном дереве. Если в дереве нет листьев (ни одна буква не кодируется), то для кода длины l есть 2^l вариантов (количество узлов на уровне l). Количество вариантов будет уменьшаться в зависимости от количества листов на верхних уровнях, но для любой длины всегда найдется хотя бы один свободный узел.



Для любого однозначно декодируемого двоичного кода объемом M с длинами кодовых слов l_1, \dots, l_M справедливо неравенство:

$$\sum_{i=1}^M 2^{-l_i} \leq 1.$$

Прямая теорема неравномерного побуквенного кодирования Для ансамбля $X = \{x, p(x)\}$ с энтропией $H(X) = H$ существует побуквенный неравномерный префиксный код со средней длиной кодовых слов $\bar{l} < H + 1$.

Доказательство Пусть $l_i = \lceil -\log p_i \rceil$ (логарифм округлим вверх). Тогда $\sum_{i=1}^M 2^{-l_i} = \sum_{i=1}^M 2^{-\lceil -\log p_i \rceil} \leq \sum_{i=1}^M 2^{\log p_i} = 1 \implies$ такой префиксный код существует.

$$\bar{l} = \sum_{m=1}^M p_m l_m < \sum_{m=1}^M p_m (-\log p_m + 1) < H + 1. \quad \square$$

Обратная теорема неравномерного побуквенного кодирования Для любого однозначно декодируемого кода для дискретного источника $\{X, p(x)\}$ с энтропией H справедливо $\bar{l} \geq H$.

Доказательство

$$\begin{aligned}
H - \bar{l} &= - \sum_{x \in X} p(x) \log p(x) - \sum_{x \in X} p(x) l(x) = \\
&= \sum_{x \in X} p(x) \log \frac{2^{-l(x)}}{p(x)} \leq \log e \sum_{x \in X} p(x) \left(\frac{2^{-l(x)}}{p(x)} - 1 \right) \leq \\
&\leq \log e \left(1 - \sum_{x \in X} p(x) \right) = 0
\end{aligned}$$

На втором шаге преобразовали: $-l(x) = \log 2^{-l(x)}$.

Свойства оптимального побуквенного кода:

- если $p_i < p_j$, то $l_i \geq l_j$ (чем чаще встречается кодовое слово, тем меньше должна быть его длина),
- есть кодовые слова, которые имеют одинаковую длину $l_M = \max_m l_m$,
- среди кодовых слов длиной $l_M = \max_m l_m$ найдутся два слова, различающиеся только в одном последнем символе.
- Пусть $p_1 \leq p_2 \leq \dots \leq p_M$. Для ансамбля $X = \{1, \dots, M\}$ и кода C , удовлетворяющего свойствам 1-3, введем ансамбль $X' = \{1, \dots, M-1\}$, сообщениями которого приписаны вероятности $\{p'_1, \dots, p'_{M-1}\}$ так, что

$$\begin{aligned}
p'_1 &= p_1, \\
p'_2 &= p_2, \\
p'_{M-1} &= p_{M-1} + p_M.
\end{aligned}$$

Из кода C построим код C' для ансамбля X' , приписав сообщениям x'_1, \dots, x'_{M-2} те же кодовые слова, что и в коде, т.е. $c'_i = c_i$, а сообщению x'_{M-1} слово c'_{M-1} , как общую часть слов C_{M-1} и C_M .

Тогда если C' оптимален для X' , то код C оптимален для X .

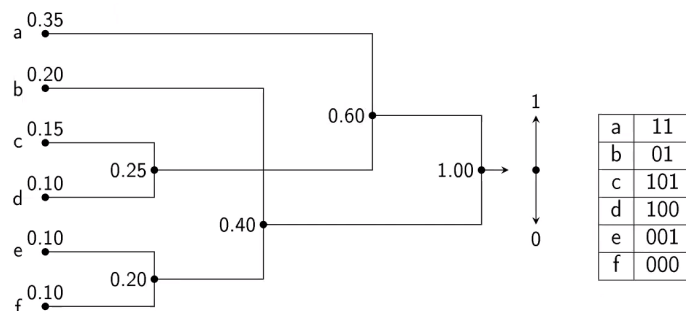
Доказательство

$$l_m = \begin{cases} l'_m & m \leq M-2, \\ l'_{M-1} + 1 & m = M-1, M. \end{cases}$$

$$\begin{aligned}
\bar{l} &= \sum_{m=1}^M p_m l_m = \sum_{m=1}^{M-2} p_m l_m + p_{M-1} l_{M-1} + p_M l_M = \\
&= \sum_{m=1}^{M-2} p_m l_m + (p_{M-1} + p_M)(l'_{M-1} + 1) = \\
&= \sum_{m=1}^{M-2} p'_m l'_m + p'_{M-1} l'_{M-1} + p_{M-1} + p_M = \\
&= \sum_{m=1}^{M-1} p'_m l'_m + p_{M-1} + p_M = \bar{l}' + p_{M-1} + p_M
\end{aligned}$$

Во второй строке смотрим на кейсы l_m , в третьей строке раскрываем скобки.

Код Хаффмана позволяет построить оптимальный побуквенный код:



Имеются исходные символы, которым сопоставлены вероятности, с которыми их генерирует источник.

Находим 2 сообщения с минимальными вероятностями и соединяем их в новую вершину, вероятность которой является суммой. Повторяем это действие, пока не получим вероятность вершины 1.

Для данного примера:

$$H = - \sum_x p(x) \log p(x) = 2.4016$$

$$\bar{l} = \sum_x l(x)p(x) = 2.4500$$

5 Подготовка к первой контрольной работе

Мат. ожидание суммы зависимых (и независимых) случайных величин:

$$\begin{aligned}
 E(x + y) &= \sum_i^{|x|} \sum_j^{|y|} (x_i + y_j) p(x_i, y_j) = \\
 &= \sum_i^{|x|} \sum_j^{|y|} (x_i p(x_i, y_j) + y_j p(x_i, y_j)) = \\
 &= \sum_i^{|x|} \sum_j^{|y|} x_i p(x_i) p(y_j | x_i) + \sum_j^{|y|} \sum_i^{|x|} y_j p(x_i) p(y_j | x_i) = \\
 &= \sum_i^{|x|} x_i p(x_i) \sum_j^{|y|} p(y_j | x_i) + \sum_j^{|y|} \sum_i^{|x|} y_j p(y_j) = \\
 &= E(x) + E(y)
 \end{aligned}$$

Мат. ожидание случ. величины и константы:

$$E(c \cdot x) = \sum c \cdot x \cdot p(x) = c \sum x p(x) = cE(x)$$

Мат. ожидание произведения независимых случ. величин:

$$\begin{aligned}
 E(xy) &= \sum_i^{|x|} \sum_j^{|y|} x_i y_j p(x_i) p(y_j) = \\
 &= \sum_i^{|x|} x_i p(x_i) \sum_j^{|y|} y_j p(y_j) = \\
 &= E(x)E(y)
 \end{aligned}$$

Дисперсия двух случайных величин, которые **не коррелируют** ($E[(x - E(x))(y - E(y))] = 0$):

$$\begin{aligned}
 D(x + y) &= \sum_i^{|x|} \sum_j^{|y|} p(x_i) p(y_j) (x_i + y_j - E(x + y))^2 = \\
 &= \sum_i^{|x|} \sum_j^{|y|} p(x_i) p(y_j) (x_i - E(x) + y_j - E(y))^2 = \\
 &= \sum_i^{|x|} \sum_j^{|y|} p(x_i) p(y_j) [(x_i - E(x))^2 + (y_j - E(y))^2 + 2(x_i - E(x))(y_j - E(y))] = \\
 &= \sum_i^{|x|} \sum_j^{|y|} p(x_i) p(y_j) (x_i - E(x))^2 + \sum_j^{|y|} \sum_i^{|x|} p(x_i) p(y_j) (y_j - E(y))^2 + \\
 &+ 2 \sum_i^{|x|} \sum_j^{|y|} p(x_i) p(y_j) (x_i - E(x))(y_j - E(y)) = \\
 &= \sum_i^{|x|} p(x_i) (x_i - E(x))^2 + \sum_j^{|y|} p(y_j) (y_j - E(y))^2 + 0 = D(x) + D(y)
 \end{aligned}$$

В более компактной форме:

$$\begin{aligned}
 D(x + y) &= E(x + y - E(x + y))^2 = E(x - E(x) + y - E(y))^2 = \\
 &= E((x - E(x))^2 + (y - E(y))^2 + 2(x - E(x))(y - E(y))) = \\
 &= E(x - E(x))^2 + E(y - E(y))^2 + 2E[(x - E(x))(y - E(y))] = \\
 &= D(x) + D(y)
 \end{aligned}$$

Дисперсия случ. величины и константы:

$$D(cx) = E(cx - E_{cx})^2 = E(cx - cE_x)^2 = E(c^2(x - E_x)^2) = c^2 D(x)$$

Удобная формула дисперсии:

$$D(x) = E(x - E(x))^2 = E(x^2 + E^2(x) - 2xE(x)) = E(x)^2 + E^2(x) - 2E(x)E(x) = E(x)^2 - E^2(x)$$

Дисперсия суммы случ. величины и константы:

$$D(x + c) = E(x + c - E(x + c))^2 = E(x + c - E(x) - c)^2 = D(x)$$

По свойству мат. ожидания: $E(c) = c$.

Если $|X| = n$, то $|Y|Y \subseteq X| = 2^n$.

Доказательство через двоичный вектор Пусть подмножество определяется двоичным вектором, каждая компонента которого указывает на принадлежность соответствующего элемента подмножеству. Тогда мощность подмножеств равняется количеству размещений с возвращением: $|\{0, 1\}^n| = 2^n$.

Доказательство через биномиальные коэффициенты Число подмножеств размера k : C_n^k . Тогда для всех размеров: $|Y| = \sum_{i=0}^n C_n^i = C_n^0 + C_n^1 + C_n^2 + \dots$. Такая сумма равна $(1 + 1)^n = 2^n$.

Можем воспользоваться свойством: $C_{n+1}^k = C_n^k + C_n^{k-1}$. Докажем его:

$$C_n^k + C_n^{k-1} = \frac{n!}{(n-k)!k!} + \frac{n!}{(n-k+1)!(k-1)!} = \frac{n!(n-k+1) + n!k}{(n+1-k)!k!} = \frac{(n+1)!}{(n+1-k)!k!} = C_{n+1}^k.$$

Разложим биномиальные коэффициенты, кроме первого и последнего (у них просто уменьшим n): $C_{n-1}^0 + C_{n-1}^1 + C_{n-1}^0 + C_{n-1}^2 + C_{n-1}^1 + \dots = 2C_{n-1}^0 + 2C_{n-1}^1 + \dots = 2(C_{n-1}^0 + C_{n-1}^1 + \dots) = \dots = 2^n$.

Пусть $X = \{a, b, c\}$,
 $p_1(a) = p_1(b) = p_1(c) = \frac{1}{3}$
 $p_2(a) = p_2(b) = \frac{1}{4}, p_2(c) = \frac{1}{2}$.

Найти: $I(x)$, $H(X)$, $I(abaac)$.

$$\begin{aligned} I(a_1) &= I(b_1) = I(c_1) = -\log \frac{1}{3} = \log 3 \\ H_1(x) &= E(I_1) = \log 3 \\ I_1(abaac) &= -\log p(abaac) = -\log \left(\frac{1}{3}\right)^5 = 5 \log 3 \end{aligned}$$

$$\begin{aligned} I(a_2) &= I(b_2) = \log 4, \quad I(c_2) = \log 2 \\ H_2(x) &= \frac{1}{4} \log 4 + \frac{1}{4} \log 4 + \frac{1}{2} \log 2 = \log 2 + \frac{1}{2} \log 2 \\ I_2(abaac) &= -\log \frac{1}{4 \cdot 3 \cdot 4 \cdot 2} = 9 \end{aligned}$$

Пусть $X = \{a, b, c, d, e, f\}$,
 $P = \{0.4, 0.3, 0.15, 0.05, 0.05, 0.05\}$.

Найти: $H(X)$, \bar{l} .

$$\begin{aligned} H(X) &= -(0.4 \log 0.4 + 0.3 \log 0.3 + 0.15 \log 0.0075) \approx 2,11 \\ \bar{l} &= 0,4 \cdot 1 + 0,3 \cdot 2 + 0,15 \cdot 3 + 0,05 \cdot 4 + 0,05 \cdot 5 + 0,05 \cdot 5 = 0,4 + 0,6 + 0,45 + 0,2 + 0,25 + 0,25 = 2,15 \end{aligned}$$

Избыточность кодирования:

$$R = \bar{l} - H$$

Рассмотрим геометрический источник. Дано множество целых чисел: $I = \{0, 1, 2, 3, \dots\}$.

Вроятность, что источник сгенерирует число из множества: $p(y = i) = (1 - \alpha)\alpha^i$.

Проверим условие нормировки (сумму вероятностей):

$$\sum_{i=0}^{\infty} (1 - \alpha)\alpha^i = (1 - \alpha) \sum_{i=0}^{\infty} \alpha^i = 1 - \alpha \frac{1}{1 - \alpha} = 1$$

Сумму бесконечно убывающей геометрической прогрессии записали по формуле.

$$E(I) = (1 - \alpha) \sum_{i=0}^{\infty} i\alpha^i.$$

Общий член ряда похож на производную степени:

$$\left(\sum_{i=0}^{\infty} \alpha^i\right)' = \sum_{i=0}^{\infty} i\alpha^{i-1} = \sum_{i=0}^{\infty} \frac{i\alpha^i}{\alpha}.$$

Выразим ряд через производную:

$$E(I) = (1 - \alpha) \left(\sum_{i=0}^{\infty} \alpha^i\right)' \alpha = (1 - \alpha) \left(\frac{1}{1 - \alpha}\right)' \alpha = \frac{\alpha}{1 - \alpha}$$

Найдем дисперсию:

$$D(X) = E(X^2) - E^2(X).$$

Найдем первую часть:

$$E(X^2) = (1 - \alpha) \sum_{i=0}^{\infty} i^2 \alpha^i$$

Выразим ряд через производную:

$$\left(\sum i\alpha^i\right)' = \sum i^2 \alpha^{i-1} = \frac{\sum i^2 \alpha^i}{\alpha}$$

$$E(X^2) = (1 - \alpha) \left(\sum i\alpha^i\right) \alpha = (1 - \alpha) \left(\left(\sum i\alpha^i\right)' \alpha\right)' \alpha = (1 - \alpha) \left(\frac{\alpha}{(1 - \alpha)^2}\right)' \alpha = \frac{\alpha(1 + \alpha)}{(1 - \alpha)^2}$$

Таким образом, дисперсия:

$$D(X) = \frac{\alpha(1 + \alpha)}{(1 - \alpha)^2} - \frac{\alpha^2}{(1 - \alpha)^2} = \frac{\alpha}{(1 - \alpha)^2}$$

Рассмотрим кодирование длин серий:

Пусть $X = \{0, 1\}$. Источник определяется последовательностью символов: 0011000010100. Перейдем к новому источнику, для этого перед каждой единицей запишем количество нулей: $0^2 1 0^0 1 0^4 1 0^1 1 0^2 \rightarrow 2, 0, 4, 1, 2$. Будем кодировать не символы, а полученные длины серий, то есть $X \rightarrow I = 0, 1, 2, 3 \dots$.

Если $p(0) = \alpha$, то вероятность, что длина серии равна i : $p(y = i) = (1 - \alpha)\alpha^i$

Найдем энтропию источника (ансамбля $\{p(i), i \mid i \in I\}$):

$$\begin{aligned} H &= - \sum_i (1 - \alpha)\alpha^i \log[(1 - \alpha)\alpha^i] = \log(1 - \alpha) + i \log \alpha \\ &= -(1 - \alpha) \log(1 - \alpha) \sum_i \alpha^i - (1 - \alpha) \log \alpha \sum_i i \alpha^i = \\ &= -\log(1 - \alpha) - \frac{\alpha \log \alpha}{1 - \alpha} = \frac{-(1 - \alpha) \log(1 - \alpha) - \alpha \log \alpha}{1 - \alpha} = \frac{\eta(\alpha)}{1 - \alpha} \end{aligned}$$

$$X = \{1, 2, 3, \dots, M\}, p(1) = p(2) = \dots = p(M) = \frac{1}{M}$$

Если будем кодировать кодом Хаффмана, а $M = 2^n$, то длина кода любого символа $\log_2 M = n$, значит $H(X) = \log M$, $\bar{l} = \log M$, значит получаем оптимальный код. Если количество символов не является степенью двойки, то код будет не оптимальным.

Пусть $\lfloor \log M \rfloor = S$ — длина кодового слова в полном дереве, $M - 2^S = F$ — количество веток, которые добавили к полному дереву. Тогда средняя длина кодового слова при равных вероятностях

$$\frac{S(2^S - F) + (S + 1)2F}{M}$$

Рассмотрим марковскую цепь:

$$A = \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix}$$

Пусть начальное распределение: $p = \{3/4, 1/4\}$.

$$I(110) = -\log 1/4 \cdot 2/3 \cdot 1/3$$