Теория информации

Лекция 4

Сжатие данных (продолжение)

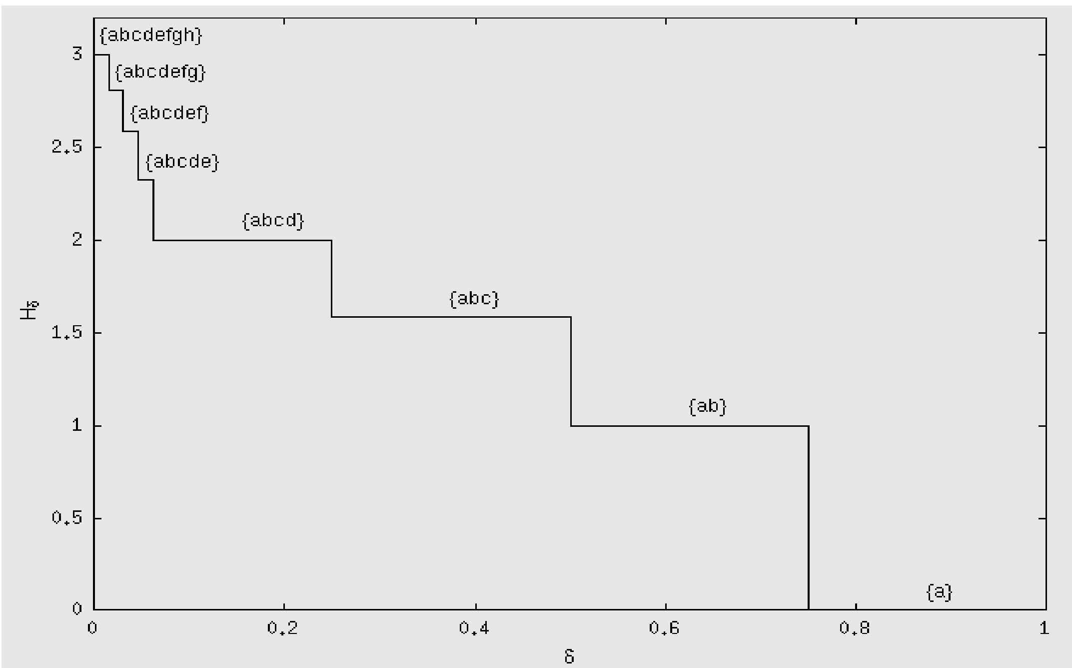
Итак, чтобы осуществить стратегию сжатия данных с риском δ , нужно выбрать наименьшее подмножество $S_\delta \subset A_x$, такое что вероятность непопадания в него x будет меньше или равна δ , т.е.

$$P(x \notin S_\delta) \leq \delta \quad \text{или} \quad P(x \in S_\delta) \geq 1-\delta \ .$$

Подмножество S_δ может быть сконструировано путем ранжирования элементов A_x в порядке убывания соответствующих вероятностей и включения первых наиболее вероятных элементов пока вероятность встретить элемент из S_δ не станет $\geq (1-\delta)$. Таким образом, для каждого δ можно задать свою меру информации:

$$H_\delta(X) = \log_2 |S_\delta| \ . \tag{4.1}$$

Введенная нами ранее величина $H_0(X)$ это случай $\delta=0$. Ниже приведен график $H_\delta(X)$ как функции δ для ансамбля из примера в предыдущей лекции.



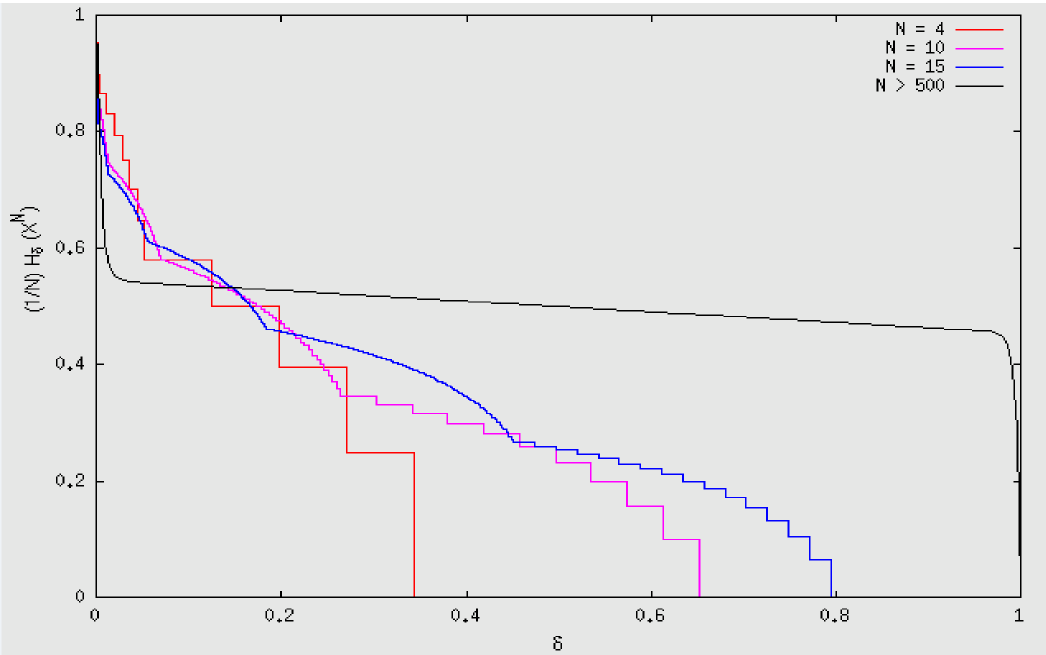
Рассмотрим возможность применить к блокам символов из источника. Рассмотрим случайную величину $\boldsymbol{x}=(x_1, x_2, \ldots, x_N)$, которая представляет собой строку из N независимых, идентично распределенных случайных величин из одного и того же ансамбля X (т.е. те же A_x и P_x). Обозначим такой ансамбль как X^N. Поскольку энтропия есть величина аддитивная для независимых случайных переменных, то

$$H(X^N) = N\,H(X) \quad . \tag{4.2}$$

Рассмотрим пример случайной величины $\boldsymbol{x}=(x_1, x_2, \ldots, x_N)$, которая представляет собой N последовательных бросков согнутой монеты, для которой $x_i \in \{0,1\}$ с вероятностями $p_0=0.9$ и $p_1=0.1$. Очевидно, что для такой монеты наиболее вероятный вектор $\boldsymbol{x}$ состоит из N нулей. Если $r(\boldsymbol{x})$ это число единиц в $\boldsymbol{x}$, то

$$P(\boldsymbol{x}) = p_0^{N-r(\boldsymbol{x})} \cdot p_1^{r(\boldsymbol{x})} \quad . \tag{4.3}$$

Чтобы определить $H_\delta(X^N)$ нужно определить наименьшее достаточное подмножество S_δ. Такое подмножество будет содержать $\boldsymbol{x}$ с $r(\boldsymbol{x})=1,2,\ldots,r_{max}(\boldsymbol{x})-1$ и некоторые $\boldsymbol{x}$ с $r(\boldsymbol{x})=r_{max}(\boldsymbol{x})$. Если посмотреть на то, что происходит с функцией $H_\delta(X^N)$ с ростом N, то для достаточно больших N эта функция становится все более плоской и близкой к $N\,H(X)$, за исключением областей близких к $\delta=0$ и $\delta=1$. Ниже приведены графики функции $\frac{1}{N} H_\delta(X^N)$ для нескольких N (кривая для $N>500$, в отличие от остальных на графике, не вычислена точно, а лишь качественно отражает функцию).



Все это означает, что для достаточно малого δ можно подобрать достаточно большое N, так что будет возможно сжатие до $N H(X)$ бит на символ. Это и составляет сущность *теоремы кодирования Шеннона*:

Теорема. Пусть X - ансамбль с энтропией $H(X)=H$ бит. Для каждого $\varepsilon>0$ и $0<\delta<1$ существует положительное целое число N_0, такое что для любого целого $N>N_0$ выполняется неравенство

$$\left|\frac{1}{N} H_\delta(X^N) - H\right| < \varepsilon \quad . \tag{4.5}$$

Типичность и типичное множество

Чем же может помочь увеличение N ? Вероятность появления такой строки $\boldsymbol{x}$, в которой r единиц и $N-r$ нулей

$$P(\boldsymbol{x}) = p_1^r (1-p_1)^{N-r} \quad . \tag{4.6}$$

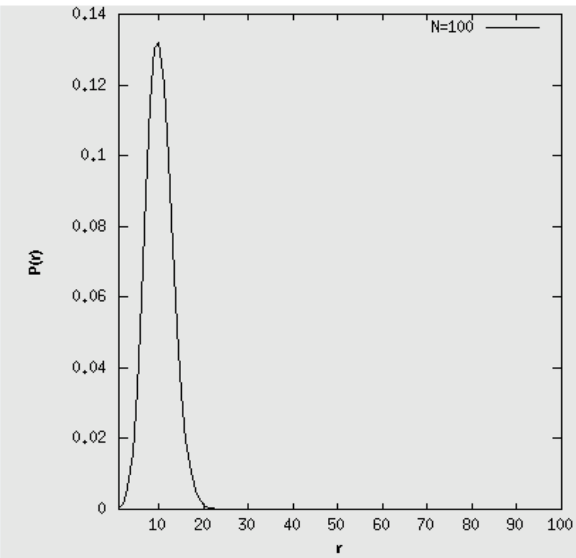
Число строк в ансамбле X^N, которые содержат r единиц определяется биномиальным коэффициентом:

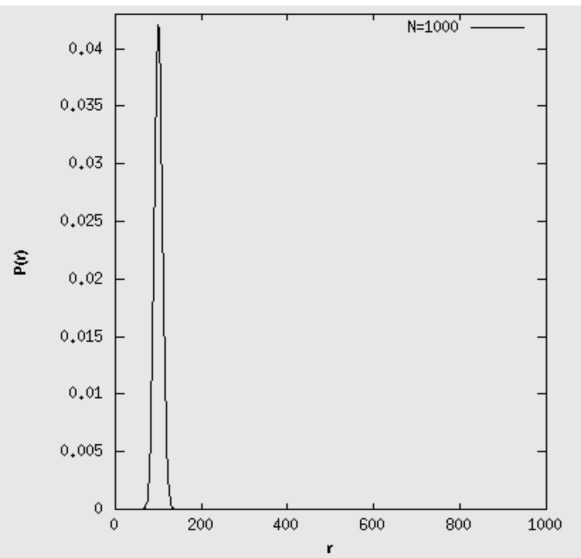
$$n(r) = C_N^r = \frac{N!}{(N-r)!\,r!} \quad . \tag{4.7}$$

Таким образом число единиц в строке имеет биномиальное распределение

$$P(r) = C_N^r p_1^r (1-p_1)^{N-r} \quad . \tag{4.8}$$

Ниже приведены примеры этого распределения для $N=100$ и $N=1000$.





Среднее значение для этого распределения $\langle r\rangle = Np_1$, стандартное отклонение $\sigma_r = \sqrt{Np_1(1-p_1)}$. Поскольку среднее значение пропорционально N, а стандартное отклонение только $\sqrt{N}$, то это означает, что с ростом N распределение (4.8) становится все более концентрированным. Для $N=100$ $r \sim 10 \pm 3$, а для $N=1000$ $r \sim 100 \pm 10$. Таким образом можно предположить, что если r попадет в небольшой интервал вокруг $\langle r\rangle$, то и соответствующая строка $\boldsymbol{x}$ попадет в некоторое небольшое подмножество X^N, которое мы назовем *типичным множеством*.

Распространим теперь понятие типичности множества на произвольный ансамбль X с алфавитом $A_x = \{a_1, a_2, \ldots, a_I\}$ и $P_x = \{p_1, p_2, \ldots, p_I\}$. В длинной строке из N символов обычно первая буква встречается $p_1 N$ раз, вторая — $p_2 N$ раз, и т.д. Следовательно грубо вероятность появления этой строки можно оценить как

$$P(\boldsymbol{x})_{typ} = P(x_1)P(x_2)\ldots P(x_N) \simeq p_1^{p_1 N} p_2^{p_2 N} \ldots p_I^{p_I N} \quad , \tag{4.9}$$

так что количество информации в типичной строке

$$\log_2 \frac{1}{P(\boldsymbol{x})} \simeq N \sum_i p_i \log_2 \frac{1}{p_i} = NH \quad . \tag{4.10}$$

Таким образом случайная величина $\log_2 1/P(\boldsymbol{x})$ весьма вероятно будет близка к величине NH.

Определим в качестве *типичных элементов* множества A_x^N те элементы, вероятность появления которых близка к 2^{-NH}. Здесь следует отметить, что в отличие от достаточного наименьшего подмножества S_δ, типичное множество не включает в себя наиболее вероятный элемент A_x^N, поскольку он вносит незначительную вероятность. Введем параметр β, характеризующий близость $P(\boldsymbol{x})$ к 2^{-NH} для того, чтобы строка $\boldsymbol{x}$ являлась типичной. Множество типичных элементов или типичное множество $T_{N\beta}$ определяется как

$$T_{N\beta} \equiv \left\{ x \in A_x^N : \left| \frac{1}{N} \log_2 \frac{1}{P(\boldsymbol{x})} - H \right| < \beta \right\} \quad . \tag{4.11}$$

Можно показать, что какую бы малая величина β ни была выбрана, типичное множество содержит почти всю вероятность по мере увеличения N. Это утверждение иногда называется как *принцип асимптотического равнораспределения*.

Принцип асимптотического равнораспределения. Для ансамбля N независимых одинаково распределенных случайных величин $X^N \equiv (X_1, X_2, \ldots, X_N)$ при достаточно большом N, реализация $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ почти наверняка принадлежит подмножеству A_x^N имеющему $2^{NH(X)}$ элементов, каждый из которых имеет вероятность близкую к $2^{-NH(X)}$.

Здесь следует отметить, что если $H(X)<H_0(X)$, то число $2^{NH(X)}$ представляет собой лишь небольшую часть от числа элементов множества A_x^N , т.е. $|A_x^N| = |A_x|^N = 2^{NH_0(X)}$.

Принцип асимптотического равнораспределения эквивалентен следующей теореме:

Теорема кодирования источника Шеннона. N независимых одинаково распределенных случайных величин каждая с энтропией $H(X)$ могут быть сжаты в более чем $NH(X)$ бит с пренебрежимо малым риском потери информации при $N\to\infty$. И наоборот, если они сжаты в менее чем $NH(X)$ бит, то почти наверняка информация при этом потеряна.