6. ВВЕДЕНИЕ В ТЕОРИЮ ИНФОРМАЦИИ

6.4.1. ЭНТРОПИЯ И ВЗАИМНАЯ ИНФОРМАЦИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Определение для взаимной информации между непрерывными случайными величинами легко получить путём обобщения случая дискретных случайных величин:

$$\boxed{\text{Rem: } I(X;Y)=\sum_{x\in \mathrm{X}}\sum_{x\in \mathrm{Y}}\mathrm{P}[x,y]\log\frac{\mathrm{P}[x\mid y]}{\mathrm{P}[x]}}\quad \mathrm{P}[x,y]=p(x,y)dxdy;\quad \frac{\mathrm{P}[x\mid y]}{\mathrm{P}[x]}=\frac{p(x\mid y)dx}{p(x)dx}=\frac{p(x\mid y)}{p(x)};$$

$$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(y)p(x\mid y)\log\frac{p(x\mid y)}{p(x)}dydx=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x)p(y\mid x)\log\frac{p(y\mid x)}{p(y)}dxdy$$

Для энтропии это оказывается невозможным:

$$\boxed{H(X)=-\sum_{x\in \mathrm{X}}\mathrm{P}[x]\log \mathrm{P}[x]}\quad \mathrm{P}[x]=p(x)dx$$

$$H(X)=-\int_{-\infty}^{\infty}p(x)\log(p(x)dx)dx$$

$$=-\int_{-\infty}^{\infty}p(x)\log p(x)dx-\log dx\int_{-\infty}^{\infty}p(x)dx$$

$$=-\int_{-\infty}^{\infty}p(x)\log p(x)dx+\infty$$

Практически бесконечность энтропии объясняется необходимостью бесконечного числа бит для представления непрерывного диапазона значений. Другими словами говоря, неопределенность реализации одного состояния из бесконечного и несчетного множества состояний бесконечно велика.

6.4.1. ЭНТРОПИЯ И ВЗАИМНАЯ ИНФОРМАЦИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx + \infty$$

Для непрерывных случайных величин вводят понятие дифференциальной энтропии (differential entropy) (иногда обозначают $h(X)$):

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx$$

Получается, что не возможно определить абсолютную меру информации, но можно сравнивать разные источники. Например, если рассмотреть равномерное на интервале шириной 1 распределение, то для него

$$H(X) = -\int_{0}^{1} 1\log 1 dx = 0$$

Следовательно, можно считать, что дифференциальная энтропия – относительная величина, показывающая насколько неопределенность появления рассматриваемого сообщения больше или меньше неопределенности появления сообщения, значения которого равновероятны на единичном интервале.

Важно: энтропия непрерывного источника не имеет смысла среднего количества информации на отсчёт, более того, она может быть отрицательной!

6.4.1. ЭНТРОПИЯ И ВЗАИМНАЯ ИНФОРМАЦИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

$$\boxed{I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(y)p(x\mid y)\log\frac{p(x\mid y)}{p(x)}dydx}$$

$$\boxed{H(X)=-\int_{-\infty}^{\infty}p(x)\log p(x)dx}$$

Путём обобщения определим условную и совместную энтропию:

$$\boxed{H(Y\mid X)=-\sum_{(x,y)\in \mathsf{X}\times\mathsf{Y}}\mathrm{P}[X=x,Y=y]\log\mathrm{P}[Y=y\mid X=x]}$$

$$H(Y\mid X)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(y\mid x)dydx$$

$$\boxed{H(X,Y)=-\sum_{(x,y)\in \mathsf{X}\times\mathsf{Y}}\mathrm{P}[X=x,Y=y]\log\mathrm{P}[X=x,Y=y]}$$

$$H(X,Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x,y)dydx$$

Аналогично случаю дискретных случайных величин, очевидно, что

$$I(X;Y)=H(X)-H(X\mid Y)=H(Y)-H(Y\mid X)$$

6.4.1. ЭНТРОПИЯ И ВЗАИМНАЯ ИНФОРМАЦИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Практический интерес также представляет случай, когда X – дискретная случайная величина, а Y – непрерывная. Очевидно, что если они зависимы, то

$$p(y) = \sum_i p(y \mid x_i) \mathrm{P}[x_i]$$

Взаимная информация, т.е. количество информации о событии $X = x_i$, которое нам даёт наблюдение $Y = y$:

$$\boxed{\text{Rem: } I(x;y) = \log \frac{\mathrm{P}[y \mid x]}{\mathrm{P}[y]}}$$

$$I(x_i;y) = \log \frac{p(y \mid x_i)}{p(y)}$$

Средняя взаимная информация между двумя источниками:

$$I(X;Y) = \sum_i \int_{-\infty}^{\infty} p(y \mid x_i) \mathrm{P}[x_i] \log \frac{p(y \mid x_i)}{p(y)} dy$$

Пример

X – случайная величина с равновероятными значениями $x_1 = A$ и $x_2 = -A$. Условные плотности вероятности Y задаются так:

$$p(y\,|\,x=A)=\frac{1}{\sqrt{2\pi}\sigma}\exp-\frac{(y-A)^2}{2\sigma^2};\quad p(y\,|\,x=-A)=\frac{1}{\sqrt{2\pi}\sigma}\exp-\frac{(y+A)^2}{2\sigma^2}$$

Тогда средняя взаимная информация

$$\boxed{I(X;Y)=\sum_i\int_{-\infty}^{\infty}p(y\,|\,x_i)\mathrm{P}[x_i]\log\frac{p(y\,|\,x_i)}{p(y)}dy}$$

$$\boxed{p(y)=\frac{1}{2}\big(p(y\,|\,A)+p(y\,|\,-A)\big)}$$

$$I(X;Y)=\frac{1}{2}\int_{-\infty}^{\infty}\left(p(y\,|\,A)\log\frac{p(y\,|\,A)}{p(y)}+p(y\,|\,-A)\log\frac{p(y\,|\,-A)}{p(y)}\right)dy$$

6.4.2. ЭПСИЛОН-ЭНТРОПИЯ (ФУНКЦИЯ СКОРОСТЬ-ИСКАЖЕНИЕ)

Ограниченный по полосе случайный процесс $X(t)$ можно дискретизировать с частотой Найквиста. Для полного преобразования в цифровую форму необходимо выполнить квантование по уровню (аппроксимацию) полученных отсчётных значений. Самый простой способ – равномерное квантование. Например, для квантования на L уровней потребуется $R = \log_2 L$ бит на один отсчёт, если L – целая степень 2 либо $R = \lfloor \log_2 L \rfloor + 1$ – в противном случае.

Если распределение значений $X(t)$ неравномерное, то можно использовать дополнительное кодирование, например кодом Хаффмана.

Очевидно, что в любом случае будут иметь место потери, вызванные ошибкой аппроксимации (искажением (distortion) сигнала), и возможно лишь пытаться их минимизировать.

6.4.2. ЭПСИЛОН-ЭНТРОПИЯ (ФУНКЦИЯ СКОРОСТЬ-ИСКАЖЕНИЕ)

Часто используемой метрикой искажения сигнала (ошибки аппроксимации) является квадратичная метрика:

$$d(\hat{x}_k, x_k) = (x_k - \hat{x}_k)^2$$

где x_k – истинное значение, $\hat{x}_k$ – квантованное значение, $d()$ – обозначение для метрики искажения.

Среднее искажение последовательности $\mathbf{x}_n$, состоящей из n значений:

$$d(\hat{\mathbf{x}}_n, \mathbf{x}_n) = \frac{1}{n}\sum_{k=1}^{n} d(\hat{x}_k, x_k)$$

При переходе к случайным величинам $d()$ оказывается случайной величиной. Её среднее значение – искажение D, которое при условии стационарность процесса равно:

$$D = \mathbf{E}\left[d(\mathbf{X}_n, \hat{\mathbf{X}}_n)\right] = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\left[d(X_k, \hat{X}_k)\right] = \mathbf{E}\left[d(X, \hat{X})\right]$$

Пусть имеется источник без памяти, описываемый непрерывной случайной величиной X и известна плотность вероятности $p(x)$. Также определён $\hat{X}$ алфавит квантователя и метрика искажения $d(x,\hat{x})$. Будем называть минимальное число бит на один отсчёт, достаточное для представления значений X алфавитом $\hat{X}$ с искажением, не превосходящим D, **эпсилон-энтропией** (**функцией скорость-искажение - rate distortion function**), $R(D)$, и определять её так:

$$R(D) = \min_{p(\hat{x}|x):\mathbf{E}[d(X,\hat{X})]\le D} I(X;\hat{X})$$

Понятно, что чем больше допустимое искажение D, тем меньше $R(D)$ и наоборот.

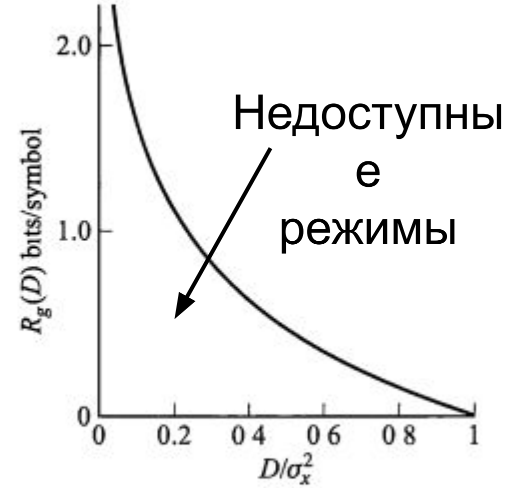
Значение эпсилон-энтропии зависит как от статистики источника $p(x)$, так и от метрики искажения $d(x,\hat{x})$. Возможность получения результата в замкнутой форме – редкость.

Третья теорема Шеннона (кодирование источника с ограничением искажения, 1959)
Для источника без памяти, описываемого случайной величиной X, может быть выполнено кодирование со скоростью R и искажением D, только, если $R > R(D)$. Для любого кода, обеспечивающего скорость $R < R(D)$, искажение превышает D.

Иначе говоря, эпсилон-энтропия является нижней границей скорости кодирования при заданном уровне искажения (ошибки аппроксимации).

Эпсилон-энтропия гауссовского источника для квадратичной метрики искажения

$$R_g(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \le D \le \sigma^2, \\ 0, & D > \sigma^2. \end{cases}$$



1. Эпсилон-энтропия не зависит от значения среднего.
2. Если $D \ge \sigma^2$, вовсе не нужно передавать информацию и при этом значение $D = \sigma^2$, может быть получено путём установки $\hat{X} = \mathbf{E}[X]$.

Выразим ошибку аппроксимации через скорость кодирования и получим **функцию искажение-скорость (distortion rate function)** для гауссовского источника:

$$D_g(R) = 2^{-2R}\sigma^2$$

Полученную ошибку аппроксимации можно выразить в дБ:

$$10\lg D_g(R) = -6R + 10\lg\sigma^2$$

т.е. ошибка аппроксимации возрастает на 6дБ на каждый бит/символ.

6.4.2. ЭПСИЛОН-ЭНТРОПИЯ (ФУНКЦИЯ СКОРОСТЬ-ИСКАЖЕНИЕ)

Теорема о верхней границе для эпсилон-энтропии

Эпсилон-энтропия непрерывного источника без памяти с нулевым средним и конечной дисперсией σ^2 при рассмотрении квадратичной метрики искажения ограничена сверху эпсилон-энтропией гауссовского источника для такой же метрики искажения:

$$R(D) \le R_g(D) = \tfrac{1}{2}\log\frac{\sigma^2}{D}, \quad 0 \le D \le \sigma^2$$

Аналогично для функции искажение-скорость:

$$D(R) \le D_g(R) = 2^{-2R}\sigma^2$$

Также для эпсилон-энтропии существует нижняя граница Шеннона при рассмотрении квадратичной метрики искажения:

$$R^*(D) = H(X) - \tfrac{1}{2}\log 2\pi e D$$

Заметим: $H_g(X) = \tfrac{1}{2}\log 2\pi e\sigma^2$

где $H(X)$ – дифференциальная энтропия непрерывного источника без памяти. Аналогично для функции искажение-скорость:

$$D^*(R) = \frac{1}{2\pi e}2^{-2(R-H(X))}$$

В итоге имеем:

$$R^*(D) \le R(D) \le R_g(D) \qquad R_g^*(D) = R_g(D)$$

$$D^*(R) \le D(R) \le D_g(R) \qquad D_g^*(R) = D_g(R)$$

6.4.2. ЭПСИЛОН-ЭНТРОПИЯ (ФУНКЦИЯ СКОРОСТЬ-ИСКАЖЕНИЕ)

Верхняя граница ошибки аппроксимации – ошибка аппроксимации гауссовского источника, значит, эта граница возрастает на 6дБ на каждый бит/символ.

Рассмотрим нижнюю границу ошибки аппроксимации:

$$\boxed{\text{Rem: } D^*(R) = \frac{1}{2\pi e} 2^{-2(R-H(X))}} \qquad D^*(R) = \frac{\sigma^2}{2\pi e \sigma^2} 2^{-2(R-H(X))}$$

$$\boxed{\text{Rem: } H_g(X) = \tfrac{1}{2}\log 2\pi e\sigma^2} \quad 10\lg D^*(R) = 10\lg\left(2^{2\frac{1}{2}\log_2 \frac{1}{2\pi e\sigma^2}} \sigma^2 2^{-2(R-H(X))} \right)$$

$$10\lg D^*(R) = -6R - 6[H_g - H(X)] + 10\lg\sigma^2$$

Нижняя граница также возрастает на 6дБ на каждый бит/символ

Разница между верхней и нижней границами ошибки аппроксимации:

$$\boxed{\text{Rem: } 10\lg D_g(R) = -6R + 10\lg\sigma^2} \qquad 10\lg\frac{D_g(R)}{D^*(R)} = 6[H_g(X) - H(X)]$$

$$\boxed{\text{Rem: } R^*(D) = H(X) - \tfrac{1}{2}\log 2\pi e D} \qquad = 6[R_g(X) - R(X)]$$

Очевидно, что дифференциальная энтропия ограничена сверху дифференциальной энтропией гауссовского источника!

6.4.2. ЭПСИЛОН-ЭНТРОПИЯ (ФУНКЦИЯ СКОРОСТЬ-ИСКАЖЕНИЕ)

Эпсилон-энтропия двоичного источника для хемминговой метрики искажения

Важный случай, для которого возможно получение результатов в замкнутой форме.
Описание источника: $p = \mathrm{P}[X = 1] = 1 - \mathrm{P}[X = 0]$.

Из теоремы о кодировании без потери информации следует, что скорость кодирования должна удовлетворять неравенству:

$$R > H(X) = H_b(p) = -p \log p - (1-p) \log(1-p)$$

Если $R < H(X)$, то неизбежны потери.

Определим **хеммингову метрику ошибки аппроксимации (Hamming distortion)**:

$$d(\hat{x}, x) = \begin{cases} 1, & x \neq \hat{x}, \\ 0, & x = \hat{x}. \end{cases}$$

Особенностью такой метрики является то, что её среднее значение равно вероятности ошибки при аппроксимации сигнала:

$$\mathbf{E}[d(X, \hat{X})] = 1 \times \mathrm{P}[X \neq \hat{X}] + 0 \times \mathrm{P}[X = \hat{X}] = \mathrm{P}[X \neq \hat{X}] = P_e$$

Эпсилон-энтропия:

$$R(D) = \begin{cases} H_b(p) - H_b(D), & 0 \le D \le \min\{p, 1-p\}, \\ 0, & \text{иначе.} \end{cases}$$

Как и ожидалось, при $D \to 0$ имеем $R(D) \to H_b(p)$

Эпсилон-энтропия двоичного источника для хемминговой метрики искажения

Требуется выполнить кодирование двоичного симметричного источника со скоростью 0,75 бит/символ. Какая достижимая вероятность ошибки?

$$p = 0,5 \to H_b(p) = 1$$

Следовательно, будет выполнено кодирование с потерями.

$$R(D) = H_b(p) - H_b(D) = 0,75$$

$$H_b(D) = 0,25$$

$$D = P_e$$

$$P_e = 0,04169$$