2.1 连续随机变量的熵和平均互信息量

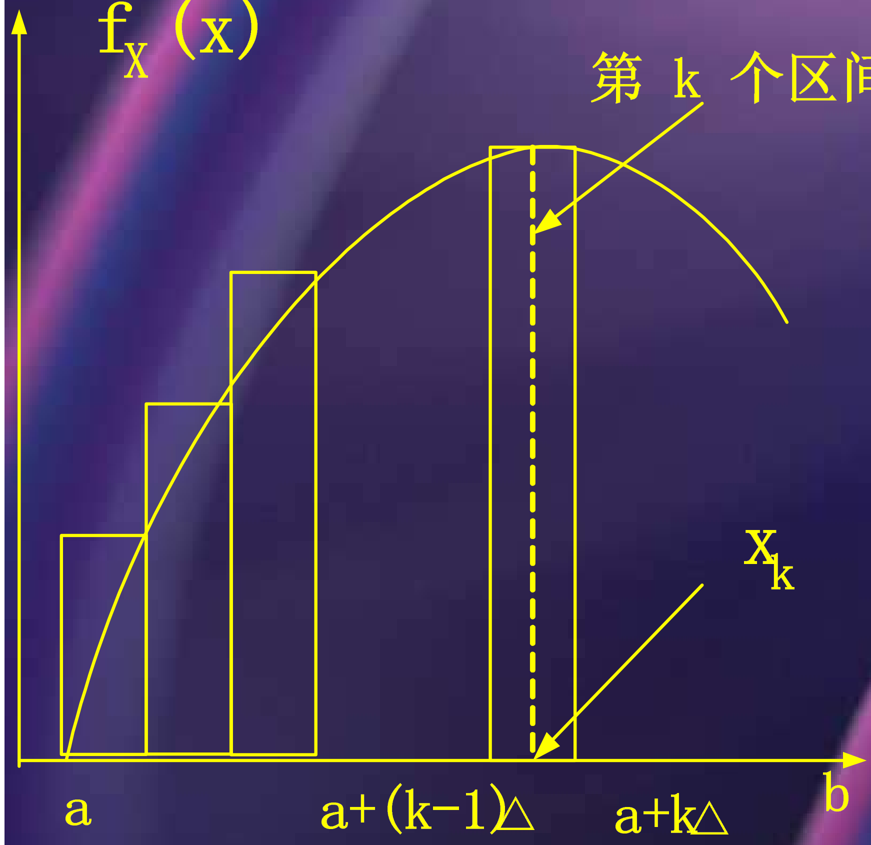
2.1.1 连续随机变量的熵

连续随机变量可以看作是离散随机变量的极限，故可采用离散随机变量来逼近。

下面，将采用这一观点讨论连续信源的信息熵与信息量。

首先类比概率p_i与概率密度$p(x)$，可得单变量连续信源的数学模型：

$$X:\begin{Bmatrix} R \\ f(x) \end{Bmatrix} \quad \text{并满足} \int_R f_X(x)dx = 1$$



令$x \in [a,b]$，且$a<b$,现将它均匀的划分为K份，每份宽度为$\triangle=(b-a)/K$，则x处于第k个区间的概率为p_k，则

$$p_k = \int_{a+(k-1)\triangle}^{a+k\triangle} f_x(x)dx$$

$$= f_X(x_k)\cdot\triangle$$

$$a+(k-1)\triangle \le x_k \le a+k\triangle$$

(积分中值定理)

当$f(x)$为x的连续函数时，由中值定理，必存在一个x_k值，使上式成立。这样得到一个离散随机变量X_Δ，其概率空间为

$$\begin{bmatrix} X_\Delta \\ P_{X_\Delta} \end{bmatrix} = \begin{bmatrix} x_1, & x_2, & \cdots\cdots & x_k \\ f_X(x_1)\Delta, & f_X(x_2)\Delta & \cdots & f_X(x_k)\Delta \end{bmatrix}$$

且概率空间是完备的：

$$\sum_{k=1}^{K} f_X(x_k)\Delta$$

$$= \sum_{k=1}^{K} \int_{a+(k-1)\Delta}^{a+k\Delta} f_X(x)dx = \int_a^b f_X(x)dx = 1$$

根据离散熵公式，有

$$H(X_\Delta) = -\sum_{i=1}^{K} p_i \cdot \log p_i$$

$$= -\sum_i [f_X(x_i) \cdot \Delta] \cdot \log[f_X(x_i) \cdot \Delta]$$

$$= -\sum_i f_X(x_i) \cdot \Delta[\log f_X(x_i) + \log \Delta]$$

$$= -\sum_i f_X(x_i) \log f_X(x_i)\Delta - (\log \Delta)\sum_i f_X(x_i) \cdot \Delta$$

$$= -\sum_i f_X(x_i) \log f_X(x_i)\Delta - \log \Delta$$

将区间[*a*,*b*]无限细分，即*K*趋于无穷大，即$\varDelta$ 趋于零，对$H(X_{\varDelta})$取极限即得连续熵$H(\mathrm{X})$的实际值

$$H(X)=\lim_{\substack{k\to\infty\\ \Delta\to 0}} H(X_{\Delta})$$

$$=-\int_a^b f_X(x)\log f_X(x)dx-\lim_{\substack{k\to\infty\\ \Delta\to 0}}\log\Delta$$

$$=-\int_a^b f_X(x)\log f_X(x)dx+\infty$$

按离散熵概念推出的连续熵为无穷大，失去意义，但上式第一项作为连续熵的相对值仍有一定意义，为了与连续熵的实际值相区别，称其为随机变量的微分熵，记为***h*(*X*)**

$$h(X) = -\int_a^b f_X(x)\log f_X(x)dx$$

微分熵更一般的定义式为：

$$h(X) = -\int_{-\infty}^{+\infty} f_X(x)\log f_X(x)dx = -\int_R f_X(x)\log f_X(x)dx$$

微分熵的特点：

1、微分熵只是实际熵的有限项，去掉了无穷大项。不能作为连续随机变量不确定性的度量公式

2、不能把微分熵视为自信息量的统计平均：因为连续随机变量取值于连续区间，有无穷多个取值点，每一点的概率均为零，自信息量无意义。

3、微分熵可用于比较两个连续随机变量不确定性

4、不具备非负性

例2.10 （均匀分布随机变量的熵）设连续随机变量X的概率密度函数为

$$f_X(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a,b]\\ 0, & x\notin[a,b]\end{cases}$$

求微分熵。

解：由微分熵定义式得

$$h(X)=-\int_a^b \frac{1}{b-a}\log\frac{1}{b-a}dx=\log(b-a)$$

讨论：若$b-a<1$，则$h(X)<0$,微分熵为负值。可见微分熵不具备非负性

例2.11 （高斯分布随机变量的熵）设随机变量服从高斯分布，即

求微分熵 $$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$$

解：

$$\mathrm{h}(X)=-\int_{-\infty}^{\infty}f_X(x)\log_2 f_X(x)dx$$

$$=-\int_{-\infty}^{\infty}f_X(x)\log_2\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$=-\int_{-\infty}^{\infty}f_X(x)(-\log_2\sqrt{2\pi\sigma^2})dx[\frac{(x-\mu)^2}{2\sigma^2}]dx$$

$$+\int_{-\infty}^{\infty}f_X(x)(\log_2 e)$$

$$\because \log_2 x = \log_2 e \ln x, \quad \int_{-\infty}^{\infty} f_X(x)dx = 1,$$

$$\int_{-\infty}^{\infty} f_X(x)\frac{(x-\mu)^2}{2\sigma^2}dx = \frac{1}{2}$$

$$\therefore H_c(X) = \log_2 \sqrt{2\pi\sigma^2} + \frac{1}{2}\log_2 e = \frac{1}{2}\log_2 2\pi e\sigma^2$$