第8章 微分熵

- 定义 设X是一个随机变量，其累积分布函数为 $F(x)=\Pr(X\le x)$ 。如果F(x)是连续的，则称该随机变量是连续的。当F(x)的导数存在时，令f(x)=F’(x)。若 $\int_{-\infty}^{\infty} f(x)=1$ ，则称f(x)是X的概率密度函数。另外，使f(x)>0的所有x构成的集合称为X的支撑集。
- 定义 一个以f(x)为密度函数的连续随机变量X的微分熵（differential entropy）定义为

$$h(X) = -\int_S f(x)\log f(x)dx$$

第8章 微分熵

- **定义** 设X是一个随机变量，其累积分布函数为 $F(x)=\Pr(X\le x)$。如果F(x)是连续的，则称该随机变量是连续的。当F(x)的导数存在时，令f(x)=F’(x)。若 $\int_{-\infty}^{\infty} f(x)=1$ ，则称f(x)是X的概率密度函数。另外，使f(x)>0的所有x构成的集合称为X的支撑集。
- **定义** 一个以f(x)为密度函数的连续随机变量X的微分熵（differential entropy）定义为

$$h(X)=-\int_S f(x)\log f(x)dx$$

$$\boxed{H(X)=-\sum_{x\in\mathcal{X}} p(x)\log p(x)}$$

微分熵的例子

- 例 [0,a]上的均匀分布：

$$h(X) = -\int_0^a \frac{1}{a}\log\frac{1}{a}dx = \log a \text{ 比特}$$

  - ✓ a<1时，h(X)<0
- 例 正态分布 $X \sim \mathcal{N}(0,\sigma^2) = \phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}$ ：

$$h(\phi) = \frac{1}{2}\log 2\pi e\sigma^2 \text{ 比特}$$

连续随机变量的AEP

- AEP：对于一个独立同分布的随机变量序列来说，

$$p(X_1,X_2,\cdots,X_n)\to 2^{-nH(X)}$$

- 定理 设 $X_1,X_2,\cdots,X_n$ 是一个服从密度函数f(x)的独立同分布的随机变量序列，则

$$-\frac{1}{n}\log f(X_1,X_2,\cdots,X_n)\to E[-\log f(X)]=h(X)\ \text{依概率}$$

- 定义 对 $\epsilon>0$ 及任意的n，定义f(x)的典型集 $A_\epsilon^{(n)}$ 如下

$$A_\epsilon^{(n)}=\left\{(x_1,x_2,\cdots,x_n)\in S^n:\left|-\frac{1}{n}\log f(x_1,x_2,\cdots,x_n)-h(X)\right|\le\epsilon\right\}$$

其中

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

连续随机变量的典型集性质

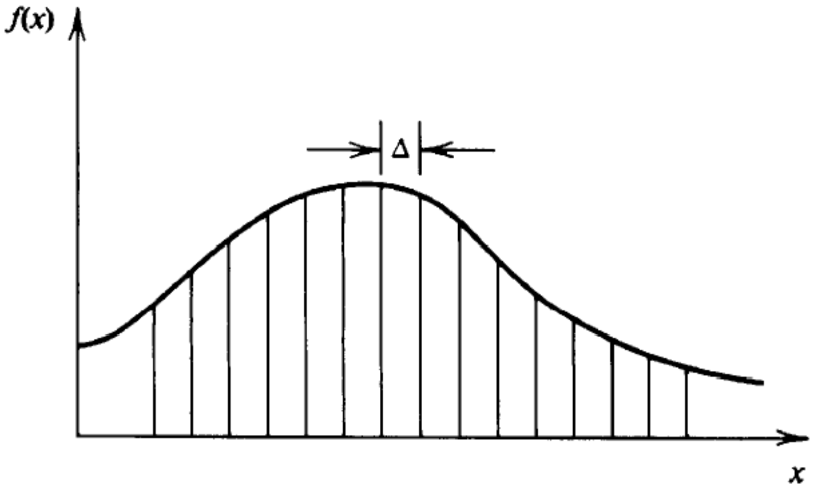
- **定义** 集合的体积Vol(A)定义为

$$\mathrm{Vol}(A)=\int_A dx_1dx_2\cdots dx_n$$

- **定理** 连续随机变量的典型集有如下的性质：

1. 对于充分大的n，$\Pr(A_\epsilon^{(n)})>1-\epsilon$
2. 对于所有的n，$\mathrm{Vol}(A_\epsilon^{(n)})\le 2^{n(h(X)+\epsilon)}$
3. 对于充分大的n，$\mathrm{Vol}(A_\epsilon^{(n)})\ge (1-\epsilon)2^{n(h(X)-\epsilon)}$

微分熵和离散熵的区别



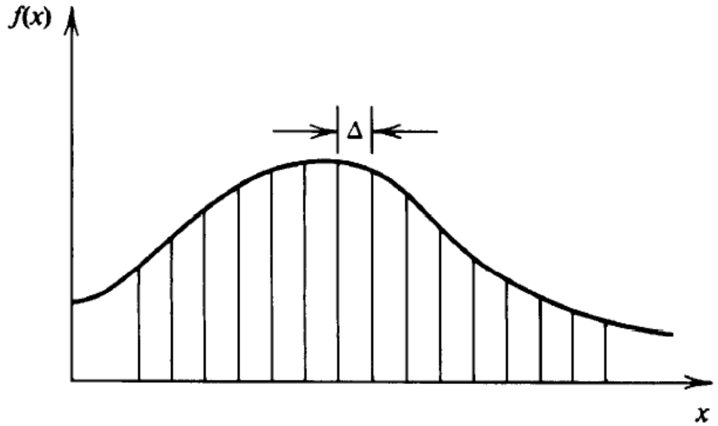
- 定理 如果随机变量X的密度函数f(x)是黎曼可积的，那么

$$H(X^{\Delta}) + \log \Delta \to h(f) = h(X), \quad \text{当} \quad \Delta \to 0$$

微分熵和离散熵的区别

- $$\begin{aligned} H(X) &= h(X) - \lim_{\Delta\to 0}\log\Delta \\ &= h(X) + \{\text{无穷大常数}\} \end{aligned}$$
- H(X)是离散意义的熵，是信息熵，无限大
- h(X)是连续意义的熵，是微分熵
- 微分熵h(X)不代表信源X的平均不确定度，也不代表X每取一个数值所提供的平均信息量，不含有信息度量的内涵

微分熵和离散熵的区别



- 连续随机变量X经过精确到小数点后n比特位的量化处理后，熵的值大约是h(X)+n
- 一般情况下，在精确到n位的意义下，h(X)+n是为了描述X所需的平均比特数。
- 一个离散随机变量的微分熵可以看作是 $-\infty$

联合微分熵与条件微分熵

- **定义** 联合密度函数为 $f(x_1, x_2, \cdots, x_n)$ 的一组随机变量 $X_1, X_2, \cdots, X_n$ 的联合微分熵定义为

$$h(X_1, X_2, \cdots, X_n) = -\int f(x^n) \log f(x^n) dx^n$$

- **定义** 如果X, Y的联合密度函数为f(x,y)，定义条件微分熵为

$$h(X|Y) = -\int f(x, y) \log f(x|y) dxdy$$

联合微分熵与条件微分熵

- **定义** 联合密度函数为 $f(x_1, x_2, \cdots, x_n)$ 的一组随机变量 $X_1, X_2, \cdots, X_n$ 的联合微分熵定义为

$$h(X_1, X_2, \cdots, X_n) = -\int f(x^n)\log f(x^n)dx^n$$

- **定义** 如果X, Y的联合密度函数为f(x,y)，定义条件微分熵为

$$h(X|Y) = -\int f(x, y)\log f(x|y)dxdy$$

$$\boxed{h(X|Y) = h(X, Y) - h(Y)}$$

多元正态分布的熵

$h(\phi)=\frac{1}{2}\log 2\pi e\sigma^2$ 比特

➢ 定理 设 $X_1, X_2, \cdots, X_n$ 服从均值为 μ，协方差矩阵为K的多元正态分布 $\mathcal{N}_n(\mu, K)$，则

$$h(X_1, X_2, \cdots, X_n) = h(\mathcal{N}_n(\mu, K)) = \frac{1}{2}\log(2\pi e)^n|K| \text{ 比特}$$

其中|K|表示K的行列式。

相对熵和互信息

$$h(X) = -\int_S f(x)\log f(x)dx$$

- 定义 两个密度函数f和g之间的相对熵：

$$D(f\|g) = \int f\log\frac{f}{g}$$

$$\boxed{D(p\|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}}$$

- 定义 联合密度函数为f(x,y)的两个随机变量间的互信息：

$$I(X;Y) = \int f(x,y)\log\frac{f(x,y)}{f(x)f(y)}$$

$$\boxed{I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}}$$

$$I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X) = h(X) + h(Y) - h(X,Y)$$

$$I(X;Y) = D(f(x,y)\|f(x)f(y))$$

互信息的例子

➤ 例

$$(X,Y) \sim \mathcal{N}(0,K)$$

$$K = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

$$h(X) = h(Y) = \frac{1}{2}\log(2\pi e)\sigma^2$$

$$h(X,Y) = \frac{1}{2}\log(2\pi e)^2|K| = \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2)$$

$$I(X;Y) = h(X) + h(Y) - h(X,Y) = -\frac{1}{2}\log(1-\rho^2)$$

微分熵、相对熵和互信息的性质

- $D(f\|g) \geq 0$，当且仅当f=g时，等号成立
- $I(X;Y) \geq 0$，当且仅当X与Y相互独立时等号成立
- $h(X|Y) \leq h(X)$，当且仅当X与Y相互独立时等号成立
- $h(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} h(X_i|X_1, X_2, \cdots, X_{i-1})$
- $h(X_1, X_2, \cdots, X_n) \leq \sum h(X_i)$，当且仅当所有随机变量相互独立时等号成立
- 平移变换不会改变微分熵：$h(X+c) = h(X)$
- $h(aX) = h(X) + \log|a|$ $\qquad$ $h(A\mathbf{X}) = h(\mathbf{X}) + \log|\det A|$

$$-\int_{-\infty}^{\infty} q(x)\ln p(x)\mathrm{d}x = -\int_{-\infty}^{\infty} p(x)\ln p(x)\mathrm{d}x$$

微分熵、相对熵和互信息的性质

- 设n维随机变量X均值为0，协方差矩阵 $K = E\{XX^t\}$ 则 $h(X) \le \frac{1}{2}\log(2\pi e)^n|K|$，当且仅当 $X \sim \mathcal{N}(0, K)$ 时等号成立。
- 对任意随机变量X及其估计，$$E(X-\hat{X})^2 \ge \frac{1}{2\pi e}e^{2h(X)}$$

  X为正态分布且估计为其均值时，等号成立
- 微分熵不存在绝对的最大熵。连续随机变量的最大熵与随机变量的限制条件有关。在不同的限制条件下，有不同的最大微分熵。
- 当已知边信息Y时，

$$E(X-\hat{X}(Y))^2 \ge \frac{1}{2\pi e}e^{2h(X|Y)}$$