第4章

连续信息与连续信源

本章主要内容

4.1 连续随机变量集合的熵

4.1.1　连续随机变量的离散化
4.1.2　连续随机变量的熵
4.1.3　连续随机变量差熵的性质
4.1.4　连续随机变量的相对熵

4.2 离散时间高斯信源的熵

4.2.1　一维高斯随机变量的熵
4.2.2　多维独立高斯随机矢量的熵
4.2.3　多维相关高斯随机矢量的熵
4.2.4　高斯马尔可夫过程的熵率

本章主要内容

4.3 连续最大熵定理

4.3.1　限峰值最大熵定理
4.3.2　限平均功率最大熵定理
4.3.3　最大熵率定理
4.3.4　熵功率

4.4 连续随机变量之间的平均互信息

4.4.1　连续随机变量之间的平均互信息
4.4.2　连续随机变量之间的平均互信息的性质

本章主要内容

4.5 离散集与连续随机变量之间的互信息

4.5.1　离散事件与连续事件之间的互信息

4.5.2　离散事件与连续随机变量之间的平均互信息

4.6 几种重要的连续信源

4.6.1　音频信源

4.6.2　语音信源

4.6.3　图像信源

4.6.4　视频信源

4. 1连续随机变量的熵

4.1.1　连续随机变量的离散化

4.1.2　连续随机变量的熵

4.1.3　连续随机变量差熵的性质

4.1.4　连续随机变量的相对熵

4.1.1 连续随机变量的离散化

$P=\{S_i,\ i=1,2,\cdots\}$ 其中 S_i 表示离散区间，$\bigcup_i S_i$ 为实数集合，且 S_i 互斥

X的概率分布函数 $F(x)$ 或概率密度 $p(x)$

划分为离散集合[X]

离散化后的随机变量概率分布

$$P_r\{[X]=i\} = P\{x\in S_i\}\approx p(x_i)\Delta x_i \quad (x_i\in S_i) \qquad (4.1)$$

4.1.1 连续随机变量的离散化

对于二维连续随机变量 ，可采用类似方法，得到离散化后对应的二维离散随机变量的联合概率分布：

$$P_r\{[X]=i,[Y]=j\} = P\{x\in S_i ,y\in T_j \}\approx p(x_i ,y_j)\Delta x_i\Delta y_j \qquad (4.2)$$

其中，$\{S_i\},\{T_j\}$ 分别为 X,Y 的某种划分，且 $x_i\in S_i$，$y_j\in T_j$ 。

4.1.2 连续随机变量的熵

设连续随机变量集合 X 在离散化后分别为 $[X]$ ，根据离散化后的离散事件的概率可得

$$H([X]) = -\sum_i p(x_i)\Delta x_i \log[p(x_i)\Delta x_i] \qquad (4.3)$$

取等间隔划分，即令

4.1.2 连续随机变量的熵

当Δx→0 时，(4.4)中的第一和第二项分别用 $h(x)$ 和 $h_0(x)$ 来表示。可得

$$h_0(X) = -\lim_{\Delta x \to 0}(\log \Delta x)\int p(x)dx = -\lim_{\Delta x \to 0}\log \Delta x \to \infty$$

$$h(X) = -\lim_{\Delta x \to 0}\sum_i p(x_i)\Delta x \log p(x_i) = -\int p(x)\log p(x)dx \qquad (4.6)$$

$h_0(x)$ 为 绝对熵， $h(x)$ 为差熵

4.1.2 连续随机变量的熵

通常我们所说的连续信源的熵是差熵，可写成：

$$h(X) = -\underset{p(x)}{E}\{\log p(x)\} = -\int p(x)\log p(x)dx \qquad (4.8)$$

4.1.3 连续随机变量差熵的性质

计算表达式类似。

通过比较可见，由计算离散熵到计算连续熵，不过是将离散概率变成概率密度，将离散求和变成积分。

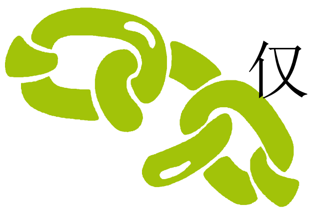
熵的不增性。

连续熵同样满足熵的不增原理，即

$$h(X) \geq h(X/Y) \qquad (4.14)$$

由于 $h(X)-h(X/Y)=\iint p(xy)\log\frac{p(x/y)}{p(x)}dxdy$

$$\geq \iint p(xy)\left(1-\frac{p(x)}{p(x|y)}\right)dxdy = 0$$



仅当X、Y独立时等式成立。

4.1.3 连续随机变量差熵的性质

可加性

设N维高斯随机矢量集合 $\mathbf{X}^{\mathrm{N}} = X_1 X_2 \cdots X_N$

很容易证明

$$h(\mathbf{X}^{\mathrm{N}}) = h(X_1) + h(X_2 / X_1) + \cdots + h(X_N / X_1 \cdots X_{N-1})$$

$$\leq h(X_1) + h(X_2) + \cdots + h(X_N) \quad (4.15)$$

且仅当 $X_1, X_2, \cdots, X_N$相互独立时，熵的不增性等式成立。

4.1.3 连续随机变量差熵的性质

连续熵和差熵的差别

1. 差熵可以作为信源平均不确定性的相对量度但不是绝对的量度。
2. 差熵不具有非负性。
3. 在连续信源中，在一一对应变换的条件下，差熵可能发生变化。

4.1.3 连续随机变量差熵的性质

定理4.1 设 $\mathbf{X}^N$ 、$\mathbf{Y}^N$为定义在$\mathbf{R}^N$空间中的两个N维矢量，$\vec{y}=f(\vec{x})$ 是一个可微的一对一的从 $\mathbf{R}^N$ 到自身的变换，那么

$$h(\mathbf{Y}^N)=h(\mathbf{X}^N)-\int_{R^N} d\mathbf{x}p(\mathbf{x})\ln\left|J(\frac{\mathbf{x}}{\mathbf{y}})\right| \tag{4.17}$$

其中 $p(\mathbf{x})$ 为 $\mathbf{X}^N$的概率密度，$J(\frac{\mathbf{x}}{\mathbf{y}})$ 为逆变换 f^{-1} 的雅可比行列式，

$$J(\frac{\mathbf{x}}{\mathbf{y}})=\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_N} \\ \cdots & \cdots & \cdots \\ \frac{\partial x_N}{\partial y_1} & \cdots & \frac{\partial x_N}{\partial y_N} \end{vmatrix} \tag{4.18}$$

4.1.3 连续随机变量差熵的性质

如果，$\left|J\left(\frac{\mathbf{x}}{\mathbf{y}}\right)\right|$ 不依赖于 $\mathbf{X}^{\mathrm{N}}$ 或者是一个线性变换，那么(4.17)式变为

$$h(\mathbf{Y}^{\mathrm{N}}) = h(\mathbf{X}^{\mathrm{N}}) - \log\left|J\left(\frac{\mathbf{x}}{\mathbf{y}}\right)\right| \qquad (4.20)$$

4.1.3 连续随机变量差熵的性质

设 $\mathbf{X}^{\mathrm{N}}$、$\mathbf{Y}^{\mathrm{N}}$ 为定义在 $\mathbf{R}^{\mathrm{N}}$ 空间中的两个N维随机矢量集合， $\mathbf{y}=\mathbf{A}\mathbf{x}+\alpha$ ，其中 $\mathbf{A}$是一个 $N\times N$ 的可逆线性变换， $\boldsymbol{\alpha}$ 为N维常数列矢量。这时由于

$$J\left(\frac{\mathbf{x}}{\mathbf{y}}\right)=\det(\mathbf{A}^{-1})=[\det(\mathbf{A})]^{-1}$$

其中，$\det(A)$ 表示矩阵A的行列式，则

$$h(\mathbf{Y}^{\mathrm{N}})=h(\mathbf{X}^{\mathrm{N}})+\log\left|\det(\mathbf{A})\right| \qquad (4.20)$$

4.1.3 连续随机变量差熵的性质

可以写成如下更明显的形式：

$$h(A\mathbf{X}^{\mathrm{N}}+\boldsymbol{\alpha})=h(\mathbf{X}^{\mathrm{N}})+\log\left|\det(\mathbf{A})\right| \qquad (4.21\text{a})$$

如果变换为平移和旋转，即 $\det(\mathbf{A})=1$ ，则

$$h(\mathbf{A}\mathbf{X}^{\mathrm{N}}+\boldsymbol{\alpha})=h(\mathbf{X}^{\mathrm{N}}) \qquad (4.21\text{b})$$

即经过平移和旋转变换后的连续信源的差熵不变。

4.1.4 连续随机变量的相对熵

与离散情况类似，我们可以定义连续随机变量的相对熵（信息散度）。设p和q为定义在同一概率空间的两个概率密度，定义p相对于q的相对熵为：

$$D(p\|q)=\int p(x)\log\frac{p(x)}{q(x)}dx \qquad (4.23)$$

4.2 离散时间高斯信源的熵

4.2.1 一维高斯随机变量的熵

4.2.2 多维独立高斯随机矢量的熵

4.2.3 多维相关高斯随机矢量的熵

4.2.4 高斯马尔可夫过程的熵率

4.2.1 一维高斯随机变量集的熵

设一维高斯随机变量X的分布密度为：

$$g(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\{-\frac{(x-m)^2}{2\sigma^2}\} \qquad (4.25)$$

其中，m，σ^2分别为随机变量X的均值和方差，

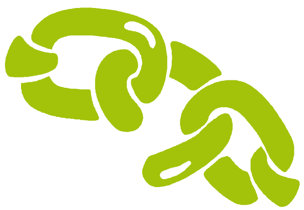
$$\sigma^2=E\{(x-m)^2=E(x^2)-m^2$$

4.2.1 一维高斯随机变量的熵

$$-\log g(x)=\frac{1}{2}\log(2\pi\sigma^2)+(\log e)\frac{(x-m)^2}{2\sigma^2}$$

$$\begin{aligned}h(X)&=-\underset{g(x)}{E}\{\log g(x)\}\\&=\frac{1}{2}\log(2\pi\sigma^2)+(\log e)\frac{E\{(x-m)^2\}}{2\sigma^2}\\&=\frac{1}{2}\log(2\pi e\sigma^2)\end{aligned} \qquad (4.26)$$

可见，高斯信源的熵仅与方差有关而与均值无关。



4.2.2 多维独立高斯随机矢量的熵

设N维独立高斯随机变量的分布密度为：

$$g(\mathbf{x})=\prod_{i=1}^{N}(2\pi\sigma_i^2)^{-\frac{1}{2}}\exp\{-\frac{(x_i-m_i)^2}{2\sigma_i^2}\} \qquad (4.27)$$

其中，m_i, σ_i^2 分别为随机矢量 X_i 的均值和方差。根据熵的可加性，可求得多维独立高斯随机矢量集合的熵：

$$h(\mathbf{X}^{\mathbf{N}})=\sum_{i=1}^{N}h(X_i)=\frac{1}{2}\sum_{i=1}^{N}\log(2\pi e\sigma_i^2)$$

$$=\frac{N}{2}\log[2\pi e(\sigma_1^2\sigma_2^2\cdots\sigma_n^2)^{1/N}] \qquad (4.28)$$

4.2.3 多维独立高斯随机矢量的熵

定理4.3

设N维高斯随机矢量$\mathbf{X}^N$的分布密度为：

$$g(\mathbf{x})=\frac{1}{(2\pi)^{N/2}\det(\Sigma)^{\frac{1}{2}}}\exp[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\Sigma^{-1}(\mathbf{x}-\mathbf{m})] \qquad (4.29)$$

其中，$\mathbf{\Sigma}=(\sigma_{ij})$ 为 $\mathbf{X}^N$ 协方差矩阵，$\sigma_{ij}=\int(x_i-m_i)(x_j-m_j)p(\mathbf{x})d\mathbf{x}$
$\mathbf{m}=(m_1,m_2,\cdots,m_N)^t$为$\vec{x}$ 的均值矢量，那么随机矢量集的熵为：

$$h(\mathbf{X}^N)=\frac{N}{2}\log[2\pi e\det(\mathbf{\Sigma})^{1/N}] \qquad (4.30)$$

4.2.3 多维独立高斯随机矢量的熵

例4.1

设X和Y是分别具有均值m_x, m_y，方差σ_x^2, σ_y^2的两个独立的高斯随机变量集合，且 $V=X-Y, U=X+Y$；试求 $h(UV)$。

解

根据题意有

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

4.2.3 多维独立高斯随机矢量的熵

例4.1

根据（4.20），（4.28)有

$$h(UV)=h(XY)+\log\left|\det\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\right|$$

$$=\log(2\pi e\sigma_x\sigma_y)+\log 2$$

$$=\log(4\pi e\sigma_x\sigma_y)$$

上面利用了X、Y的独立性。

4.2.3 多维独立高斯随机矢量的熵

例4.2

将变换改为$U=(X+Y)/\sqrt{2}, V=(X-Y)/\sqrt{2}$，试求 $h(UV)$

解：

此时$(x\quad y)$到$(u\quad v)$的变换是正交变换，变换后熵不变，所以

$$h(UV)=\log(2\pi e\sigma_x\sigma_y)$$

4.2.4 高斯马尔可夫过程的熵率

- 平稳有记忆高斯过程可用马尔可夫模型来描述
- 如果状态机的状态 $s_k=(x_{k-p},\cdots,x_{k-1})$ 且过程输出满足

$$x_k=-\sum_{i=1}^{p}a_i x_{k-i}+\varepsilon_k \qquad (4.32)$$

- 那么，过程模型称为p 阶自回归（AR）模型。

§4.3　连续最大熵定理

4.3.1　限峰值最大熵定理

4.3.2　限平均功率最大熵定理

4.3.3　最大熵率定理

4.3.4　熵功率

§4.3.1 限峰值最大熵定理

峰值功率受限为P

等价

信源输出信号的瞬时电压受限于$\pm\sqrt{P}$

等价

信源输出的幅度取值受限于有限区间$[a,b]$

§4.3.1 限峰值最大熵定理

定理4.4

幅度受限的随机变量，当均匀分布时有最大的熵。

详细描述：当N维矢量 $\mathbf{x}=(x_1,x_2,\cdots,x_N)$，具有概率密度 $p(\mathbf{x})$，分布区间为 $(a_1,b_1),(a_2,b_2),\ldots,(a_N,b_N)$ 时，其熵满足

$$h(\mathbf{X}^{\mathbf{N}})\le\sum_{i=1}^{N}\log(b_i-a_i) \qquad (4.45)$$

§4.3.1 限峰值最大熵定理

证明

设 $q(\mathbf{x})$ 是分布区间为 $(a_1,b_1),(a_2,b_2),\ldots,(a_N,b_N)$ 的均匀分布，概率密度为

$$q(\mathbf{x})=\begin{cases}\dfrac{1}{\prod_{i=1}^{N}(b_i-a_i)}, & \vec{x}\in\cap(a_i,b_i)\\ 0 & \text{其它}\end{cases} \qquad (4.46)$$

计算 $-\log q(\mathbf{x})=\sum_{i=1}^{N}\log(b_i-a_i)$, $(x_i\in(a_i,b_i)$, $i=1,\ldots,N)$,

根据 定理4.2，有 $D(p//g)=\int p(\mathbf{x})\log\dfrac{p(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x}\geq 0$

所以: $\mathrm{h}(\mathbf{X}^{\mathrm{N}})\leq \underset{p(\vec{x})}{E}\{-\log q(\mathbf{x})\}=\sum_{i=1}^{N}\log(b_i-a_i)$

仅当 $p(\mathbf{x})$ 等于 $q(\mathbf{x})$时，等式成立，此时的熵就是均匀分布的信源的熵

§4.3.2 限功率最大熵定理

信源输出信号的平均功率受限

等价

一维随机变量方差一定

- 注：一维随机变量的功率即是方差

多维随机变量协方差矩阵一定

§4.3.2 限功率最大熵定理

定理4.5　功率受限的随机变量，当高斯分布时有最大的熵。

详细描述：当N维信源的概率密度为$p(\mathbf{x})$,协方差矩阵为$\mathbf{\Sigma}$，且$\mathbf{\Sigma}=\left(\sigma_{ij}\right)$,其中：

$$\sigma_{ij}=\int(x_i-m_i)(x_j-m_j)p(\mathbf{x})d\mathbf{x}$$

$\mathbf{m}=(m_1,m_2,\cdots,m_N)$为$\mathbf{x}$的均值矢量，那么$\mathbf{X}^{\mathrm{N}}$的熵满足：

$$h(\mathbf{X}^{\mathrm{N}})\le\frac{N}{2}\log[2\pi e\det(\mathbf{\Sigma})^{1/N}] \qquad (4.47)$$

仅当$\mathbf{X}^{\mathrm{N}}$为高斯分布时，等式成立

§4.3.2 限功率最大熵定理

证明

设 $g(\mathbf{x})$ 为式（4.29）所规定的 N 维高斯概率密度，其协方差矩阵也为 $\mathbf{\Sigma}$，

根据定理4.2（散度不等式）有

$$D(p//g)=\int p(\mathbf{x})\log\frac{p(\mathbf{x})}{g(\mathbf{x})}d\mathbf{x}\geq 0$$

所以：

$$\mathrm{h}(\mathbf{X}^{\mathrm{N}})\leq \underset{p(\vec{x})}{E}\{-\log \mathrm{g}(\mathbf{x})\}$$

$$=\frac{N}{2}\log(2\pi\det(\mathbf{\Sigma})^{1/N})+\frac{1}{2}\log e\sum_i\sum_j t_{ij}\underset{p(\vec{x})}{E}\{(x_i-m_i)(x_j-m_j)\}$$

§4.3.2 限功率最大熵定理

证明（续）

$$= \frac{N}{2}\log(2\pi \det(\mathbf{\Sigma})^{1/N}) + \frac{1}{2}(\log e)\sum_i \sum_j t_{ij}\sigma_{ij}$$

所以：

上面利用了两概率分布具有相同的自协方差矩阵的条件，其中 $\mathbf{\Sigma}^{-1} = \left(t_{ij}\right)$

仅当 $p(\vec{x})$为高斯分布时等式成立。证毕。

§4.3.4 熵功率

熵功率： $\sigma^2 = \dfrac{1}{2\pi e} \mathrm{e}^{2h(X)}$ （4.50）

$h(X) = (1/2)\log(2\pi e\sigma^2)$ （4.51）

由此可得到以下结论：

◆连续信源的熵功率是具有相同差熵的高斯信源的平均功率

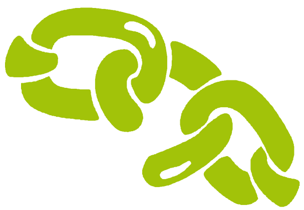
◆任何一个信源的熵功率不大于其实际平均功率 σ_x^2（方差）

限功率最大熵定理

$h(X) \le (1/2)\log(2\pi e\sigma_x^2)$

$h(X) = (1/2)\log(2\pi e\sigma^2)$

推导出 $\sigma^2 \le \sigma_x^2$



§4.4 连续随机变量之间的平均互信息

4.4.1　连续随机变量之间的平均互信息

4.4.2　连续随机变量之间的平均互信息的性质

§4.4.1 连续随机变量之间的平均互信息

设X、Y为两个连续随机变量集合，
它们的平均互信息定义为$I(X;Y)$

$$I([X]_{\mathrm{P}};[Y]_{\mathrm{Q}})=\sum_{i,j}p(u_iv_j)\log\frac{p(u_iv_j)}{p(u_i)q(v_j)} \qquad (4.60)$$

$\mathrm{P}=\{u_i\}$是集合X的划分，$\mathrm{Q}=\{v_j\}$是集合Y的划分。

$$I(X;Y)=\sup_{\mathrm{P,Q}} I([X]_{\mathrm{P}};[Y]_{\mathrm{Q}}) \qquad (4.59)$$

Sup (Supremum)表示上确界，
取遍所有对X、Y的划分P、Q。

§4.4.1 连续随机变量之间的平均互信息

注

◆X、Y的区间划分越细，平均互信息越大。

设P1、P2是X的两种划分，而离散集合 $[X]_{P_2}$ 是 $[X]_{P_1}$ 的细化。

$$I([X]_{\mathrm{P}_1};[Y]_{\mathrm{Q}}) \geq I([X]_{\mathrm{P}_2};[Y]_{\mathrm{Q}})$$

同理可适用于Y。

◆划分区间大小趋近于零时的平均互信息可作为连续随机变量集合X、Y的平均互信息。

§4.4.1 连续随机变量之间的平均互信息

$$I(X;Y)=\iint p(xy)\log\frac{p(xy)}{p(x)q(x)}dxdy$$

设连续集合X、Y，分别由P、Q两划分变成离散集合 $[X]_{\mathrm{P}};[Y]_{\mathrm{Q}}$，且 $[X]_{\mathrm{P}}=\{u_i\}$，$[Y]_{\mathrm{Q}}=\{v_j\}$，那么，根据式(4.2)、式(4.3)可得

$$\begin{cases} p(u_i)=p(x_i)\Delta x_i & (x_i\in u_i)\\ q(v_j)=p(y_j)\Delta y_j & (y_j\in v_j)\\ p(u_iv_j)=p(x_iy_j)\Delta x_i\Delta y_j & (x_i\in u_i,y_j\in v_j)\end{cases}$$

所以 $$I([X]_{\mathrm{P}};[Y]_{\mathrm{Q}})=\sum_{i,j}p(x_iy_j)\Delta x_i\Delta y_j\log\frac{p(x_iy_j)\Delta x_i\Delta y_j}{p(x_i)\Delta x_i q(v_j)\Delta y_j} \tag{4.38}$$

当 $\Delta x\to 0,\Delta y\to 0$ 时,趋近于 $I(X;Y)$，因此，

$$I(X;Y)=\iint p(xy)\log\frac{p(xy)}{p(x)q(x)}dxdy=\underset{p(xy)}{E}\{\log\frac{p(xy)}{p(x)q(y)}\} \tag{4.39}$$

§4.4.2 连续随机变量之间平均互信息的性质

对称性 $I(X;Y)=I(Y;X)$ (4.63)

非负性 $I(X;Y)\geq 0$ (4.64)

平均互信息与差熵的关系 $I(X;Y)=h(X)+h(Y)-h(XY)$ (4.65)

线性变换下互信息的不变性

设 $\mathbf{X}^{N}$、$\mathbf{Y}^{N}$为定义在RN空间中的两个N维矢量,$\mathbf{U}^{N}$、$\mathbf{V}^{N}$分别 $\mathbf{X}^{N}$、$\mathbf{Y}^{N}$的可逆线性变换，即,$\mathbf{u}=\mathbf{Ax}+\boldsymbol{\alpha}$，$\mathbf{v}=\mathbf{By}+\boldsymbol{\beta}$

那么 $I(\mathbf{U}^{N};\mathbf{V}^{N})=I(\mathbf{X}^{N};\mathbf{Y}^{N})$ (4.68)

§4.4.2 连续随机变量之间平均互信息的性质

例4.7

二维高斯随机变量集合XY，其中X,Y的均值和方差分别为m_x，m_y和σ_x，σ_y，且相关系数为ρ求：

（1）X,Y的联合分布密度$P_{XY}(xy)$；

（2）$h(XY)$; $h(X)$; $h(Y)$；

（3）$h(Y/X)$; $h(X/Y)$; $I(X;Y)$；

§4.4.2 连续随机变量之间平均互信息的性质

解：

（1）设XY的协方差矩阵Σ，则

$$\Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix},\quad \Sigma^{-1}=\frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)}\begin{pmatrix}\sigma_y^2 & -\rho\sigma_x\sigma_y\\ -\rho\sigma_x\sigma_y & \sigma_x^2\end{pmatrix}$$

利用（4.27)式，得

$$P_{XY}(xy)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-m_x)^2}{\sigma_x^2}-\frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y}+\frac{(y-m_y)^2}{\sigma_y^2}]\}$$

§4.4.2 连续随机变量之间平均互信息的性质

解：

（2）根据高斯变量差熵的公式（4.26）、（4.28），得

$$h(XY)=\log[2\pi e\sigma_x\sigma_y\sqrt{1-\rho^2}]$$

$$h(X)=\frac{1}{2}\log[2\pi e\sigma_x^2]$$

$$h(Y)=\frac{1}{2}\log[2\pi e_x\sigma_y^2]$$

（3）根据公式(4.14)和（4.67），得到

$$h(Y/X)=h(XY)-h(X)=\frac{1}{2}\log[2\pi e\sigma_y^2(1-\rho^2)]$$

$$h(X/Y)=h(XY)-h(Y)=\frac{1}{2}\log[2\pi e\sigma_x^2(1-\rho^2)]$$

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

§4.4.2 连续随机变量之间平均互信息的性质

例4.8

已知X，S为零均值、互相独立的高斯随机变量集合，方差分别为P、Q；Z为独立于X和S的零均值高斯噪声，方差为N；设$Y = X + S + Z, U = X + \alpha S,$

其中，α为常量。求：

（1）$I(U;S)$ ；（2）$I(U;Y)$

§4.4.2 连续随机变量之间平均互信息的性质

解：

（1）
$$I(U;S)=H(U)+H(S)-H(US)$$

$$\rho_{US}=\frac{E(US)}{\sigma_S\sigma_U}=\frac{E(XS+\alpha S^2)}{\sigma_S\sigma_U}=\frac{\alpha Q}{\sigma_S\sigma_U}$$

$$=\frac{1}{2}\log 2\pi e\sigma_U^2+\frac{1}{2}\log 2\pi e\sigma_S^2-\log 2\pi e\sigma_U\sigma_S\sqrt{1-\rho^2}$$

$$=\frac{1}{2}\log\frac{\sigma_U^2\sigma_S^2}{\sigma_U^2\sigma_S^2-(\alpha Q)^2}=\frac{1}{2}\log\frac{P+\alpha^2 Q}{P}$$

§4.4.2 连续随机变量之间平均互信息的性质

解：

（2）

$$\rho_{UY} = \frac{E(YU)}{\sigma_Y \sigma_U} = \frac{E(X^2 + \alpha XS + XS + \alpha S^2 + XZ + \alpha ZS)}{\sigma_Y \sigma_U} = \frac{P + \alpha Q}{\sigma_Y \sigma_U}$$

$$I(U;Y) = H(U) + H(Y) - H(UY)$$

$$= \frac{1}{2}\log 2\pi e \sigma_U^2 + \frac{1}{2}\log 2\pi e \sigma_Y^2 + \log 2\pi e \sigma_U \sigma_Y \sqrt{1-\rho^2}$$

$$= \frac{1}{2}\log \frac{\sigma_U^2 \sigma_Y^2}{\sigma_U^2 \sigma_Y^2 - (P + \alpha Q)^2} = \frac{1}{2}\log \frac{(P+Q+N)(P+\alpha^2 Q)}{PQ(1-\alpha^2) + N(P+\alpha^2 Q)}$$

§4.5　离散集与连续随机变量之间的互信息

4.5.1　离散事件与连续事件之间的互信息

4.5.2　离散事件与连续随机变量之间的平均互信息

§4.5.1离散事件与连续事件之间的互信息

设事件$x \in X$，取自字母表A，
y 为连续集Y中的事件，
定义x 与y 之间的互信息为：

$$I(x;y) = \log \frac{q(x/y)}{p(x)} = \log \frac{p(y/x)}{q(y)} \qquad (4.71)$$

其中： $q(y)$为y 的概率密度，且$q(y) = \sum_x p(x)p(y/x)$

$$q(x/y) = \frac{p(x)p(y/x)}{q(y)}$$

§4.5.2　离散与连续随机变量的平均互信息

连续随机变量 X 连续随机变量 Y 的平均互信息：

$$I(X;Y) = \underset{p(x)p(y/x)}{E}\left[\log\frac{p(y/x)}{q(y)}\right] = \sum_{x} p(x)\int p(y/x)\log\frac{p(y/x)}{q(y)}dy \tag{4.72}$$

§4.5.2　离散与连续随机变量的平均互信息

例4.9

已知一信道的输入和输出分别为X和Y，其中X等概率取值为+1，-1，$Y = X + Z$ ，且Z为在-2与2之间均匀分布的随机变量；

（1）求的概率密度 $q(y)$ ；

（2）求信道输入与输出之间的互信息$I(X;Y)$ 。

§4.5.2　离散与连续随机变量的平均互信息

解：

（1） $q(y) = p(x=+1)P(y/x=+1) + p(x=-1)P(y/x=-1)$

$$= [P(y/x=+1) + P(y/x=-1)]/2$$

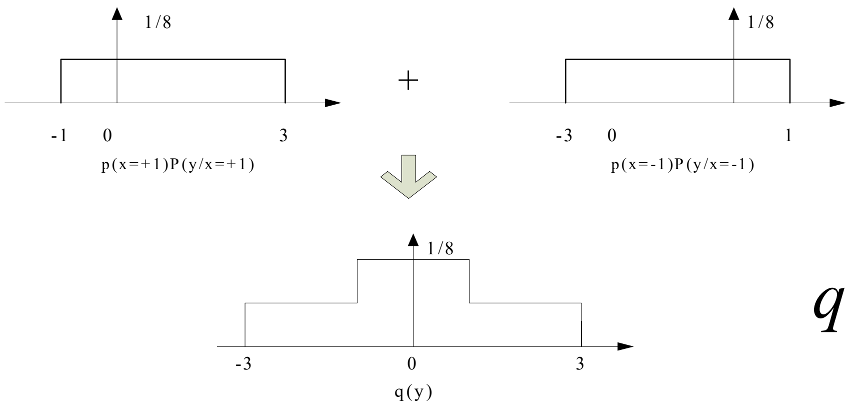
其中,$P(y/x=+1)$和 $P(y/x=-1)$ 为条件概率密度。
设 $p_z(z)$ 为 z 的概率密度，可得

$$P(y/x=+1) = p_Z(y\ -1),\ P(y/x=-1) = p_Z(y\ +1)$$

求 $q(y)$ 的过程，如下图所示

§4.5.2 离散与连续随机变量的平均互信息

解：



如图，可得

$$q(y)=\begin{cases}1/8 & (-3\le y<-1)\\ 1/8 & (1<y\le 3)\\ 1/4 & (-1\le y\le 1)\\ 0 & (y<-3, y>3)\end{cases}$$

（2） $$I(X;Y)=2\left(0.5\times\int_{-3}^{-1}\frac{1}{4}\log\frac{1/4}{q(y)}dy+0.5\times\int_{-1}^{0}\frac{1}{4}\log\frac{1/4}{q(y)}dy\right)$$

$$=2\left(0.5\times\int_{-3}^{-1}\frac{1}{4}\log\frac{1/4}{1/8}dy+0.5\times\int_{-1}^{0}\frac{1}{4}\log\frac{1/4}{1/4}dy\right)=0.5\ bit$$

4.6 几种重要的连续信源

4.6.1　音频信源

4.6.2　语音信源

4.6.3　图像信源

4.6.4　视频信源

4.6.1　音频信源

●人类能听到的声音通常称为音频（Audio）

●人听觉所能感受到的频率范围大致是20到20kHz，频率高于20 kHz的称为超声波，频率低于20 Hz的称为次声波。

●对音频进行处理和编码时，需要数字化声音，因此需要将模拟音频进行抽样。

●通常音频样值是通过PCM调制得到的，样值之间有很大的相关性，这就是说音频信源具有较大的剩余度。

4.6.2 语音信源

●语音（Speech）是指人所发出的声音

●语音功率谱频率范围通常从500到4kHz，按每倍频程8到10dB速率衰减。

●语音信号的剩余度表现在如下几方面：

（1）语音信号样本间相关性很强。

（2）浊音具有准周期性；

（3）声管形状及其变化的速率较慢；

（4）数字语音码符号的概率不均匀。

4.6.3　图像信源

●图像（Image）信源主要指的是数字图像，包括静止图像和活动图像。

●数字图像可以分成如下几类：
(1) 二值图像
(2) 灰度图像
(3) 连续色调图像
(4) 离散色调图像
(5) 卡通类图像

●图像冗余包含：
(1) 空间冗余
(2) 频谱冗余

4.6.4 视频信源

●视频（Video）的含义是可视信息，指的是时变的图像。

●视频信号分为三类：分量视频、组合视频和S-视频。

●视频信号的重要参数是：垂直分辨率、分辨率、图像纵横比、帧率以及每像素的所需比特数。

●彩色视频采用RGB三基色模型，即任何图像彩色都用三基色红(R)、绿(G)、蓝(B)的混合来近似。

本章小节

1.连续信息的度量通过对信源离散化后得到离散信息度量取极限得到；

2.差熵

●表达式 $h(X)=-\int p(x)\log p(x)dx$

●不具有非负性，在一一对应变换下不具有不变性

$$h(\mathbf{AX^N}+\boldsymbol{\alpha})=h(\mathbf{X^N})+\log\left|\det(\mathbf{A})\right|$$

●可加性 $h(\mathbf{X^N})=\sum_{i=1}^{N}h(X_i\mid X_1\cdots X_{i-1})\le\sum_{i=1}^{N}h(X_i)$

3.平稳过程的熵率：$\bar{h}(X)=\lim\limits_{N\to\infty}N^{-1}h(X_1X_2\cdots X_N)=\lim\limits_{N\to\infty}h(X_N\mid X_1\cdots X_{N-1})$

本章小节

4. N维高斯矢量的熵：$h(\mathbf{X}^N)=(N/2)\log[2\pi e\det(\Sigma)^{1/N}]$

5. AR高斯过程熵率：$\bar{h}(X)=(1/2)\log(2\pi e\sigma_\varepsilon^2)$

6. 平稳连续过程熵功率：$\tilde{N}(X)=(2\pi e)^{-1}\,\mathrm{e}^{2\bar{h}(X)}$

7. 熵功率不等式（EPI ）$e^{2h(\mathbf{X}^n+\mathbf{Y}^n)/n}\ge e^{2h(\mathbf{X}^n)/n}+e^{2h(\mathbf{Y}^n)/n}$

本章小节

8．连续最大熵定理：

●限平均功率时高斯分布有最大熵；

●限峰值时均匀分布有最大熵；

●满足二阶矩约束的最大熵率过程是p阶高斯马尔可夫过程；

9. 连续平均互信息

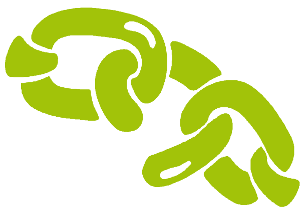
$$I(X;Y)=\iint_{XY} p(xy)\log\frac{p(xy)}{p(x)q(y)}dxdy$$

10. 连续平均互信息与差熵的关系：

$$I(X;Y)=h(X)-h(X\mid Y)=h(Y)-h(Y\mid X)=h(X)+h(Y)-h(XY)$$

11. 离散与连续随机变量间的平均互信息：

$$I(X;Y)=\sum_{x} p(x)\int_{Y} p(y\mid x)\log\frac{p(y\mid x)}{q(y)}dy$$



谢谢!