第4章

连续信息与连续信源

第4章 连续信息与连续信源

本章主要内容：

1. 连续随机变量集合的熵
2. 离散时间高斯信源的熵
3. 连续最大熵定理
4. 连续随机变量集的平均互信息
5. 离散集与连续集之间的互信息

- 本章在研究第3章离散信源的基础上研究连续信源的信息量度量。

- 内容安排如下：

　　首先研究离散时间连续信源的差熵，主要是高斯信源的差熵；然后介绍连续信源最大熵定理；最后介绍连续集合之间的平均互信息、离散集合与连续集合的平均互信息。

§4.1　连续随机变量集合的熵

本节主要内容:

1. 连续随机变量的离散化
2. 连续随机变量集的熵
3. 连续随机变量集的条件熵
4. 连续随机变量集的联合熵
5. 连续随机变量集合差熵的性质
6. 连续随机变量集合的信息散度

4.1.1 连续随机变量的离散化

一个连续随机变量的离散化过程大致如下：

若给定连续随机变量集合X的概率分布$F(x)=P\{X \le x\}$或概率密度$p(x)$；再给定一个由实数集合到有限或可数集合的划分P，使得

$P=\{S_i, i=1,2,\cdots\}$，其中S_i表示离散区间，$\bigcup_i S_i$为实数集合，且S_i互斥；用P将X进行划分，划分后的离散集合表示为$[X]_P$或$[X]$，且使得：

$$P_r\{[X]=i\}=P\{x\in S_i\}\approx p(x_i)\Delta x_i \quad (x_i\in S_i) \qquad (4.1.2)$$

即，把$x_i\in S_i$的概率看成$[X]$取值i的概率，这样就得到离散化后随机变量的概率分布。

4.1.1 连续随机变量的离散化(续)

对于二维连续随机变量XY，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:

$$P_r\{[X]=i,[Y]=j\} = P\{x\in S_i ,y\in T_j \}\approx p(x_i ,y_j)\Delta x_i\Delta y_j \quad (4.1.3)$$

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　连续随机变量集的熵

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

$$H([X]) = -\sum_{i} p(x_i)\Delta x_i \log [p(x_i)\Delta x_i] \qquad (4.1.4)$$

取等间隔划分，即令$\Delta x_i = \Delta x$，则

$$\begin{aligned} H([X]) &= -\sum_{i} p(x_i)\Delta x \log [p(x_i)\Delta x] \\ &= -\sum_{i} p(x_i)\Delta x \log p(x_i) - \sum_{i} p(x_i)\Delta x \log \Delta x \end{aligned} \qquad (4.1.5)$$

4.1.2 连续随机变量集的熵(续)

这样，离散化后信源的熵可看成由（4.1.5）式中的两项组成，当Δx→0 时，第一和第二项分别用$h(x)$和$h_0(x)$来表示。那么

$$h_0(X) = -\lim_{\Delta x \to 0} (\log \Delta x) \int p(x) dx = -\lim_{\Delta x \to 0} \log \Delta x \ \to \infty \qquad (4.1.6)$$

$$h(X) = -\lim_{\Delta x \to 0} \sum_i p(x_i) \Delta x \ \log \ p(x_i) = -\int p(x) \log p(x) dx \qquad (4.1.7)$$

4.1.2 连续随机变量集的熵(续)

可见，连续信源的熵由两部分组成：一部分为绝对熵，其值为无限大，用$h_0(x)$表示；另一部分为差熵（或微分熵），用$h(x)$表示。

通常我们所说的连续信源的熵就是差熵，可写成：

$$h(X) = -\underset{p(x)}{E}\{\log p(x)\} = -\int p(x)\log p(x)dx \qquad (4.1.8)$$

差熵的单位为：比特(奈特)/自由度。

4.1.3　连续随机变量集的条件熵

类似地，可计算离散化后的$H([X]/[Y])$为：

$$H([X]|[Y]) = -\sum_{i,j} p(x_i y_j)\Delta x_i \Delta y_j \log [p(x_i|y_j)\Delta x_i]$$

取等间隔划分，即令$\Delta x_i = \Delta x$，$\Delta y_j = \Delta y$，则

$$H([X]|[Y]) = -\sum_{i,j} p(x_i y_j)\Delta x \Delta y \log[p(x_i|y_j)\Delta x]$$

$$= -\sum_{i,j} p(x_i y_j)\Delta x \Delta y \log p(x_i|y_j) - p(x_i y_j)\Delta x \Delta y \log \Delta x \qquad (4.1.9)$$

4.1.3 连续随机变量集的条件熵(续)

当$\Delta x \to 0, \Delta y \to 0$时，第一和第二项分别用$h(X|Y)$和$h_0(X|Y)$来表示。那么

$$h_0(X|Y) = -\lim_{\Delta x\to 0,\Delta y\to 0} \log \Delta x \int p(x_i y_i) dxdy = -\lim_{\Delta x\to 0} \log \Delta x \to \infty$$

$$h(X|Y) = -\lim_{\Delta x\to 0,\Delta y\to 0} \sum_{i,j} p(x_i\ y_j)\Delta x \Delta y \log p(x_i|y_j)$$

$$= -\iint p(x\ y)\log p(x|y)dxdy \qquad (4.1.11)$$

4.1.3 连续随机变量集的条件熵(续)

与前面类似以，连续信源的条件熵也由两部分组成：一部分为绝对熵，其值为无限大，用 $h_0(X|Y)$ 表示；另一部分为差熵，用 $h(X|Y)$ 表示，可写成：

$$h(X|Y) = -\underset{p(xy)}{E}\{\log p(x|y)\} = -\iint p(xy)\log p(x|y)dxdy \quad (4.1.12)$$

条件差熵的单位也为：比特(奈特)/自由度。

4.1.4 连续随机变量集的联合熵

类似地，可以定义N维连续随机变量集合的联合差熵为：

$$h(\mathbf{X}^{\mathrm{N}}) = \underset{p(\vec{x})}{E}\{-\log p(\mathbf{x})\} = -\int_{\vec{x}} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \quad (4.1.13)$$

其中，N维连续随机变量 $\mathbf{X}^{\mathrm{N}} = X_1 X_2 \cdots X_N$ π，$p(\mathbf{x})$ 为 $\mathbf{X}^{\mathrm{N}}$ 的联合概率密度，积分为在整个概率空间的多重积分。

联合差熵的单位为：比特(奈特)/N自由度。

4.1.4 连续随机变量集的联合熵(续)

对于平稳随机过程或平稳随机序列 $\{X_i\},(i=1,2,\cdots)$，定义熵率为:

$$h(X)=\lim_{N\to\infty}\frac{h(X_1X_2\cdots X_N)}{N} \qquad (4.1.14)$$

实际上，熵率表示每自由度的熵。

注:

（1）一维连续信源的符号含一个自由度，N维连续信源的符号含N个自由度;

（2）一个连续信源的符号可能含多个自由度，所以比特/自由度不一定等于比特/符号;

（3）对于某些信源有时也用比特/符号做单位。

4.1.5 连续随机变量集合差熵的性质
——连续熵与离散熵的类似性

1. 连续熵与离散熵计算表达式类似。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。
2. 熵的不增性。连续熵同样满足熵的不增原理，即

$$h(X) \geq h(X/Y) \qquad (4.1.15)$$

由于

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

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

仅当X、Y独立时等式成立。

有关4.15推导的说明

1、$h(X) = -\int p(x)\log p(x)dx$　　　　（1）

$p(x) = \int p(y)p(x|y)dy$，代入（1）中有:

$$h(X) = -\int[\int p(y)p(x|y)dy]\log p(x)dx$$

$$= -\int\int p(y)p(x|y)\log p(x)dxdy = -\int\int p(x,y)\log p(x)dxdy$$

2、$\log x \geq 1 - \dfrac{1}{x}$　去log不等式

4.1.5 连续随机变量集合差熵的性质(续)
——连续熵与离散熵的类似性

3. 可加性

设N维高斯随机矢量集合 $\mathbf{X}^{\mathbf{N}} = X_1 X_2 \cdots X_N$，很容易证明

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

$$\le h(X_1) + h(X_2) + \cdots + h(X_N) \qquad (4.1.16)$$

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

4.1.5 连续随机变量集合差熵的性质
——连续熵与离散熵的差别

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。

　　如前所述，差熵实际上只是连续信源熵的一部分，因此不能作为信源平均不确性大小的绝对量度。但是每个信源所包含的绝对熵部分都等于 $-\log\Delta x$，与信源的概率分布无关，所以差熵的大小仍然可以作为信源平均不确定性的相对量度，即差熵的大的信源平均不确定性大。

4.1.5 连续随机变量集合差熵的性质(续)

——连续熵与离散熵的差别

2. 差熵不具有非负性。

   根据差熵的公式，如果在整个积分区间概率密度的值若大于1，则计算出的差熵的值就小于零。

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。

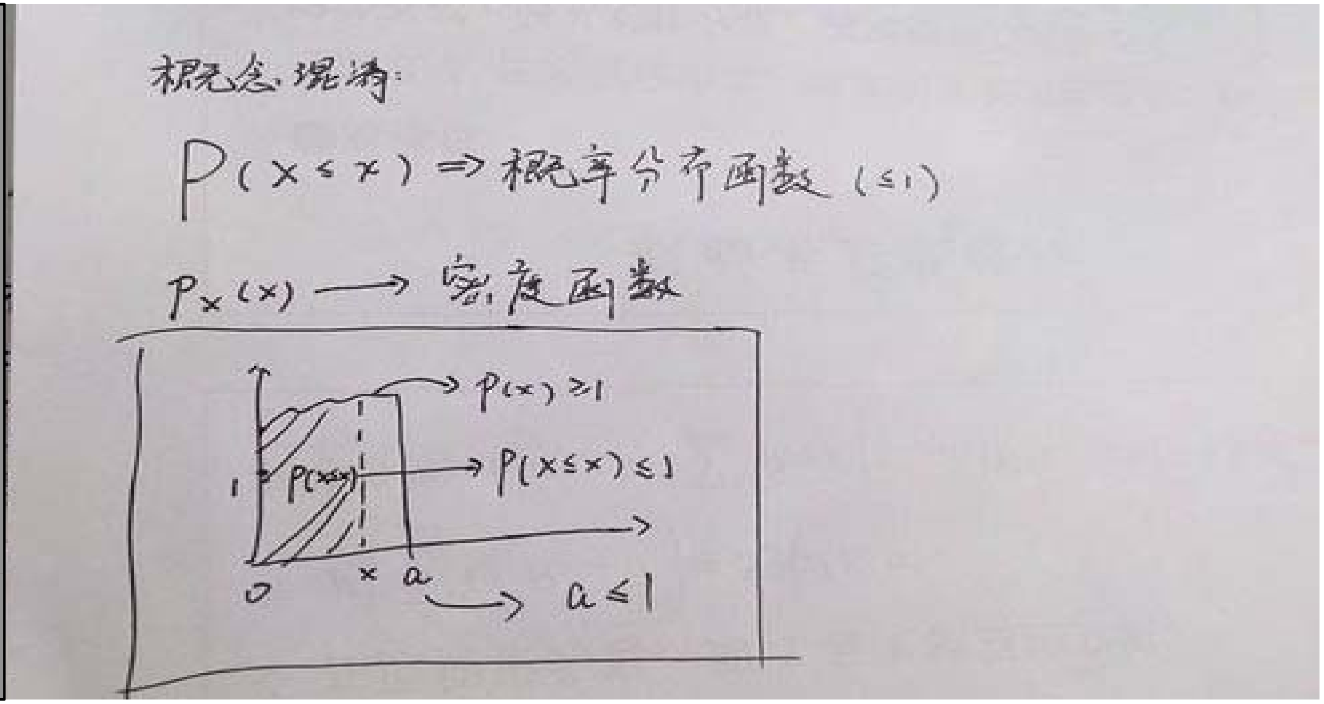
   如果两个离散信源符号的取值有一一对应的变换关系，那么变换后信源的熵是不变的。对于连续信源，差熵可能发生变化

概率密度有可能大于 1 吗?

如下图所示，定义在区间$\left[0,a\right]$间的均匀分布，$a<1$

概率密度$p(x)=1/a\ >1,\quad 0\leq x\leq a$

均匀分布示意图



4.1.5 连续随机变量集合差熵的性质

——连续信源变换的熵

定理4.1.1　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R^N到自身的变换，那末

$$h(\mathbf{Y}^N)=h(\mathbf{X}^N)-\int_{R^N}d\mathbf{x}p(\mathbf{x})\ln\left|J\left(\frac{\mathbf{x}}{\mathbf{y}}\right)\right| \quad (4.1.17)$$

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，即

$$J\left(\frac{\mathbf{x}}{\mathbf{y}}\right)=\begin{vmatrix}\frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_N}\\ \cdots & \cdots & \cdots\\ \frac{\partial x_N}{\partial y_1} & \cdots & \frac{\partial x_N}{\partial y_N}\end{vmatrix} \quad (4.1.18)$$

4.1.5 连续随机变量集合差熵的性质(续)

——连续信源变换的熵

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(4.1.17)式变为

$$h(\mathbf{Y}^{\mathrm{N}}) = h(\mathbf{X}^{\mathrm{N}}) - \log\left|J\left(\frac{\mathbf{x}}{\mathbf{y}}\right)\right| \qquad (4.1.20)$$

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

$$h(\mathbf{Y}^{\mathrm{N}}) = h(\mathbf{X}^{\mathrm{N}}) + \log\left|\det(\mathbf{A})\right| \qquad (4.1.21)$$

$$-\log|[\det(\mathbf{A})]^{-1}| = +\log|\det(\mathbf{A})|$$

4.1.5 连续随机变量集合差熵的性质(续)
——连续信源变换的熵

可以写成如下更明显的形式:

$$h(A\mathbf{X}^{\mathrm{N}}+\boldsymbol{\alpha})=h(\mathbf{X}^{\mathrm{N}})+\log\left|\det(\mathbf{A})\right| \qquad (4.1.21a)$$

如果变换为平移和旋转，即 $\det(\mathbf{A})=1$，则

$$h(\mathbf{A}\mathbf{X}^{\mathrm{N}}+\boldsymbol{\alpha})=h(\mathbf{X}^{\mathrm{N}}) \qquad (4.1.21b)$$

即经过平移和旋转变换后的连续信源的差熵不变。

4.1.6 连续随机变量集合的信息散度

与离散情况类似，我们可以定义连续随机变量的信息散度。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q于的散度为：

$$D(p//q)=\int p(x)\log\frac{p(x)}{q(x)}dx \qquad (4.1.22)$$

同样，在(4.1.22) 中，概率密度的维数不限，可以是一维，也可以是多维。

4.1.6 连续随机变量集合的信息散度(续)

定理4.1.2（散度不等式）如果两个连续随机矢量概率密度分别为 $p(x)$和 $q(x)$，那么

$$D(p // q) \geq 0 \tag{4.1.23}$$

当且仅当对所有 $x, p(x) = q(x)$ 时，等式成立。

连续随机变量散度不等式的证明

$$-D(p\|q)=\int p(x)\log\frac{q(x)}{p(x)}dx\le\int p(x)\left(\frac{q(x)}{p(x)}-1\right)(\log e)dx=0$$

证明关键点：

（1） 如果x为正，$1-\frac{1}{x}\le\ln x\le x-1$

（2） 对数换底公式：$\log_a x=\frac{\log_b x}{\log_b a}$

（3） $\int p(x)\left(\frac{q(x)}{p(x)}-1\right)dx=\int q(x)dx-\int p(x)dx=0$

§4.2　离散时间高斯信源的熵

本节主要内容:

1. 一维高斯随机变量集的熵
2. 多维独立高斯随机变量集的熵
3. 多维相关高斯随机变量集的熵

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:

$$g(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-m)^2}{2\sigma^2}\} \qquad (4.2.1)$$

其中 m 和 σ^2 分别为随机变量X的均值和方差:

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

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

4.2.1 一维高斯随机变量集的熵（续）

根据（4.2.5）式，可得一维高斯随机矢量集合的熵为：

$$
\begin{aligned}
h(X) &= -\underset{g(x)}{E}\{\log g(x)\} \\
&= \frac{1}{2}\log(2\pi\sigma^2) + (\log e)\frac{E\{(x-m)^2\}}{2\sigma^2} \qquad (4.2.2) \\
&= \frac{1}{2}\log(2\pi e\sigma^2)
\end{aligned}
$$

可见，高斯信源的熵仅与方差有关而与均值无关。

4.2.2 多维独立高斯随机变量集的熵

设N维独立高斯随机变量的分布密度为：

$$g(\mathbf{x}) = \prod_{i=1}^{N} (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\{-\frac{(x_i - m_i)^2}{2\sigma_i^2}\} \qquad (4.2.3)$$

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：

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

$$= \frac{N}{2}\log[2\pi e(\sigma_1^2\sigma_2^2\cdots\sigma_{\underset{\uparrow}{n}}^2)^{1/N}] \qquad (4.2.4)$$

N

4.2.3 多维相关高斯随机变量集的熵

定理4.2.1　设N维高斯随机矢量 X^N 的分布密度为:

$$g(\mathbf{x})=\frac{1}{(2\pi)^{N/2}\det(\Sigma)^{\frac{1}{2}}}\exp[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\Sigma^{-1}(\mathbf{x}-\mathbf{m})] \quad (4.2.5)$$

其中 $\boldsymbol{\Sigma}=\left(\sigma_{ij}\right)$ 为 $\mathbf{X}^{\mathbf{N}}$ 协方差矩阵，其中 $\sigma_{ij}=\int(x_i-m_i)(x_j-m_j)p(\mathbf{x})d\mathbf{x}$ ，

$\mathbf{m}=(m_1,m_2,\cdots,m_N)^t$为 $\vec{x}$ 的均值矢量，那么随机矢量集的熵为:

$$h(\mathbf{X}^{\mathbf{N}})=\frac{N}{2}\log[2\pi e\det(\boldsymbol{\Sigma})^{1/N}] \quad (4.2.6)$$

多维相关高斯随机向量熵的计算（1）

证明关键点：

（1） $$-\log g(x)=\frac{N}{2}\log\left[2\pi\det(\Sigma)^{1/N}\right]+\frac{1}{2}(\log e)\left[\left(\mathbf{x}-\mathbf{m}\right)^{\tau}\Sigma^{-1}(\mathbf{x}-\mathbf{m})\right]$$

（2）设$\Sigma^{-1}=\left(t_{ij}\right)$，$\left(\mathbf{x}-\mathbf{m}\right)^{\tau}\Sigma^{-1}(\mathbf{x}-\mathbf{m})=\sum_i\sum_j t_{ij}(x_i-m_i)(x_j-m_j)$

（3） $$h(\mathbf{X}^N)=\underset{g(\mathbf{x})}{E}\left\{-\log g(\mathbf{x})\right\}=\frac{N}{2}\log\left[2\pi\det(\Sigma)^{1/N}\right]+\frac{1}{2}(\log e)\sum_i\sum_j t_{ij}E\left\{(x_i-m_i)(x_j-m_j)\right\}$$

$$\sum_i\sum_j t_{ij}\sigma_{ij}=\sum_i\sum_j t_{ij}\sigma_{ji}$$

多维相关高斯随机向量熵的计算（2）

证明关键点：

（4）Σ为对称矩阵$\Rightarrow \sigma_{ij} = \sigma_{ji}$

（5）$\because\ \Sigma^{-1}\Sigma = \mathbf{I}$，$\therefore \sum_i \underbrace{\sum_j t_{ij}\sigma_{ji}}_{\text{对角线上第i个分量}} = \sum_i 1 = N$

单位阵对角线为 1

对角线上第i个分量

例4.2.1　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 $U=X+Y$，$V=X-Y$；试求 $h(UV)$ 。

解

根据题意有

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

根据（4.1.21），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。

例4.2.2（续） 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

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

§ 4.3 连续最大熵定理

主要内容

1、限峰值最大熵定理

2、限功率最大熵定理

3、熵功率和剩余度

- 对于离散信源，当信源符号等概率分布时信源的熵取最大值。对于连续信源，差熵也可以通过改变信源的概率密度求最大值，但情况有所不同：

- 除一般情况下对概率密度的非负 $p(x) \geq 0$ 和归一化 $\int_{-\infty}^{\infty} p(x)dx = 1$ 的约束条件之外，还必须附加其他的约束条件。这些附加约束通常是对随机变量矩的约束，最重要的约束是对信源输出的峰值约束和功率约束，即在一阶矩和二阶矩的约束条件下，

求 $h(x) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx$ 的极值问题

4.3.1 限峰值最大熵定理

- 若信源输出信号的峰值功率受限为P，即信源输出信号的瞬时电压限定在$\pm\sqrt{P}$，等价于信源输出连续随机变量X的取值幅度受限于$[a,b]$内取值，即在约束 $\int_a^b p(x)dx = 1$ 下，求信源熵的极值。

——峰值功率受限等价于将信源输出的幅度限制在一个有限区间内。

定理4.3.1　幅度受限的随机变量，当均匀分布时有最大的熵。

证明：

当N维随机矢量 $\mathbf{x}=(x_1,x_2,\cdots,x_N)$,具有概率密度 $p(\mathbf{x})$，分布区间为（a1，b1），（a2，b2），…（aN，bN）时，其熵满足：

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

设是分布区间为（a1，b1），（a2，b2），…（aN，bN）的均匀分布，概率密度为：

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

证明续:

计算-log $q(\mathbf{x})=\sum_{i=1}^{N}\log(b_i-a_i)$ （xi∈（ai，bi）， i=1, ..., N ），
根据定理4.1.2，有

$$D(p//q)=\int p(\mathbf{x})\log\frac{p(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x}\geq 0$$ —》∫p(x)logp(x)dx+∫p(x)q(x)dx

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

$$h(\mathbf{X}^N)=-\int p(\mathbf{x})\log p(\mathbf{x})d\mathbf{x}$$

即：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

- 若信源输出信号的平均功率受限，对于均值为0的一维信源来说，就是其方差 σ^2 受限。对于均值不为零的N维信源 $\mathbf{X}^{\mathrm{N}}$ 的情况，就是在其协方差矩阵 $\mathbf{\Sigma}$ 受限的约束条件下，求信源熵的极值

- 一维随机变量的功率就是它的方差，功率受限即为方差一定；<u>对于多维随机变量，功率受限即为协方差矩阵一定</u>

定理4.3.2 功率受限的随机变量，当高斯分布时有最大的熵。

- 该定理可详细描述如下：设N 维信源 $\mathbf{X}^{\mathrm{N}}$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：
$\sigma_{ij} = \int (x_i - m_i)(x_j - m_j) p(\mathbf{x}) d\mathbf{x}$, $\mathbf{m} = (m_1, m_2, \cdots, m_N)^{\mathrm{t}}$
为 $\mathbf{x}$的均值矢量， $\mathbf{X}^{\mathrm{N}}$的熵满足

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

仅当为高斯分布时等式成立。

证明：

设$g(\mathbf{x})$ 为 (4.2.5)式所规定的N维高斯概率密度，其协方差矩阵也为 $\mathbf{\Sigma}$ ，根据定理4.1.2有

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

证明续

所以

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

$$= \frac{N}{2}\log(2\pi\det(\boldsymbol{\Sigma})^{1/N}) + \frac{1}{2}(\log e)\sum_i\sum_j t_{ij}\sigma_{ij}$$ 限功率—》2个概率分布具有相同的协方差矩阵

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

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 $\boldsymbol{\Sigma}^{-1} = \left(t_{ij}\right)$，类似于(4.2.6)式的推导，可得到(4.3.1)式，仅当 $p(\vec{x})$ 为高斯分布时等式成立。证毕。

4.3.3 熵功率和剩余度

定义差熵为$h(X)$ 的连续随机变量 X 的熵功率为:

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

从而有

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

高斯随机变量的熵功率等于其平均功率

- 可见，连续信源的熵功率就是具有相同差熵的高斯信源的平均功率。
- 设 X 的实际功率为 σ_x^2，根据限功率最大熵定理，具有相同功率时，高斯分布的熵最大，因此有 $h(X) \le (1/2)\log(2\pi e \sigma_x^2)$ 再根据(4.2.10)，得，即 $\sigma^2 \le \sigma_x^2$，<u>任何一个信源的熵功率不大于其实际平均功率（方差）</u>。

信源剩余

- 熵功率的大小可以表示连续信源剩余的大小。如果熵功率等于信号的平均功率，就表示信号没有剩余。熵功率和信号的平均功率相差越大，说明信号的剩余越大。所以信号平均功率和熵功率之差被称为连续信源的剩余度。

- 只有高斯分布的信源的熵功率等于其实际平均功率，剩余度为零。

- 说明：以高斯分布的熵功率（平均功率）表示最大的不确定度，熵功率和信号的平均功率相差越大，表示冗余度越大

以编码的概念理解：最大编码长度由熵功率确定，可编码长度由平均功率确定，说明了信源的可压缩度

定理4.3.3　熵功率不等式

- 如果X和Y都是方差有限的连续随机变量，则

$$e^{2h(X+Y)} \geq e^{2h(X)} + e^{2h(Y)}$$

仅当X和Y均为高斯随机变量时等式成立。
（证明略）

说明：两随机变量集合的熵功率的和不大于两随机变量和的熵功率，<u>除非两者都是高斯随机变量</u>。

§4.4 连续随机变量集的平均互信息

主要内容

1、连续随机变量集的平均互信息

2、连续随机变量集平均互信息的性质

4.4.1 连续随机变量集的平均互信息

- 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：

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

- 其中，Sup（Supremum）为上确界，取遍所有对X、Y的划分P、Q。根据离散平均互信息的定义可得

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

其中，X划分为集合$\mathrm{P}=\{u_i\}$，Y划分为集合$\mathrm{Q}=\{v_j\}$，分别为相应离散集合的概率分布。

- 设对X有两种划分，分别为P_1、P_2，其中P_1中的每一个区间都是P_2中某个区间的子区间，则离散集合$[X]P_1$中的某元素就包含在离散集合$[X]P_2$中的某个元素中。因此$[X]_{P1}$可看成$[X]_{P2}$的细化。根据前面离散互信息的性质有：

$$I([X]_{P_1};[Y]_{Q}) \ge I([X]_{P_2};[Y]_{Q})$$

- 同样的论证也适用于Y。将划分区间大小趋近于零时的平均互信息的极限值作为连续随机变量集合X、Y的平均互信息。

设连续集合X、Y，分别由P、Q两划分变成离散集合 $[X]_P;[Y]_Q$ 且 $[X]_P=\{u_i\}$ ， $[Y]_Q=\{v_j\}$ ，那末，根据（4.1.2）（4.1.3）可得

$$p(u_i)=p(x_i)\Delta x_i \qquad (x_i\in u_i)$$

$$q(v_j)=q(y_j)\Delta y_j \qquad (y_j\in v_j)$$

$$p(u_iv_j)=p(x_iy_j)\Delta x_i\Delta y_j \qquad (x_i\in u_i, y_j\in v_j)$$

所以

$$I([X]_P;[Y]_Q)=\sum_{i,j}p(x_iy_j)\Delta x_i\Delta y_j\log\frac{p(x_iy_j)\Delta x_i\Delta y_j}{p(x_i)\Delta x_i q(y_j)\Delta y_j}$$

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

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

$$=\underset{p(xy)}{E}\{\log\frac{p(xy)}{p(x)q(y)}\}$$

4.4.2 连续随机变量集平均互信息的性质

- 对称性，即

$$I(X;Y) = I(Y;X)$$

- 非负性，即

$$I(X;Y) \geq 0$$

- 平均互信息与差熵的关系

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

- 线性变换下平均互信息的不变性

设$\mathbf{X}^{\mathrm{N}}$、$\mathbf{Y}^{\mathrm{N}}$为定义在R^{N}空间中的两个N维矢量，$\mathbf{U}^{\mathrm{N}}$ 、 $\mathbf{V}^{\mathrm{N}}$分别为$\mathbf{X}^{\mathrm{N}}$、$\mathbf{Y}^{\mathrm{N}}$的可逆线性变换，即，$\mathbf{u} = \mathbf{A}\mathbf{x} + \boldsymbol{\alpha}$ ，$\mathbf{v} = \mathbf{B}\mathbf{y} + \boldsymbol{\beta}$

那么

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

❖ 例题4.4.1：二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，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）$h(Y/X);\ h(X/Y);\ I(X;Y)$ 。

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（4.2.5)式，得

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

h(X^N)=log(2πe*det(Σ)^(1/N))/N

续解：

（2）根据高斯变量差熵的公式（4.2.6）、（4.2.2），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(4.1.15)和（4. 2.22)，得到

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

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

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

❖ 例4.4.2 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 为常数。求：（1）$I(U;S)$；（2）$I(U;Y)$

污纸信道编码问题

❖ 解(1)：由已知条件 $X \sim N(0, P)$ $S \sim N(0, Q)$ $Z \sim N(0, N)$

注意X和S独立，相关为0

$$Y \sim N(0, P+Q+N) \quad U \sim N(0, P+\alpha^2 Q)$$

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

$$I(U;S) = H(U) + H(S) - H(US)$$

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

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

独立，相关为0

❖ 续解：

(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 离散集与连续集之间的互信息

主要内容

1、离散事件与连续事件之间的互信息

2、离散集合与连续集合的平均互信息

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：

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

互信息对称

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 离散集合与连续集合的平均互信息

- 平均互信息定义:

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

- 例4.5.1　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 $I(X;Y)$ — 离散X和连续Y 。

- 解

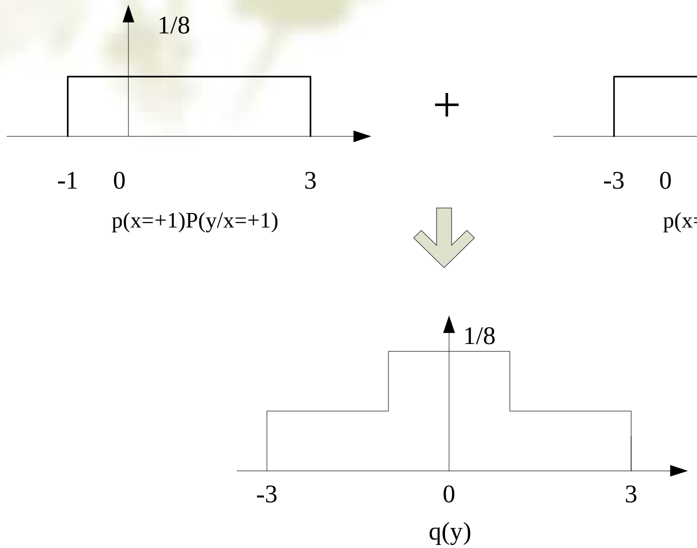
（1）　$q(y)\ =\ p(x=+1)P(y/x=+1)+\ p(x=-1)P(y/x=-1)$

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

其中，$P(y/x=+1)$ 和 $P(y/x=-1)$ 为条件概率密度，有:

$$P(y/x=+1)=\ p_Z(y\ \ \text{-}1),\qquad P(y/x=\ \ \text{-}1)=\ p_Z(y\ \ +1)$$

转化为Z的分布，Z=Y-X=Y-1



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

计算说明

（1） $p_z(z)=\begin{cases}\dfrac{1}{4} & -2\leq z\leq 2\\ 0 & \end{cases}$

（2）对于$p_z(y-1)$，有

$$-2\leq y-1\leq 2\Rightarrow -1\leq y\leq 3$$

$$\frac{1}{2}p(y\,|\,x=+1)=\begin{cases}\dfrac{1}{8} & -1\leq y\leq 3\\ 0 & \end{cases}$$

（3）对于$p_z(y+1)$有类似结果

本章小结

- 连续信源的熵通过对信源输出取值离散化来研究；
- 连续熵与离散熵有很大的区别，特别是连续熵不具非负性，在一一对应的变换下不具熵的不变性；
- 连续熵与离散熵也有类似性，具有熵的可加性；
- 连续随机变量集合的平均互信息保持离散平均互信息的性质；
- 连续最大熵定理：限功率时高斯信源有最大熵；限峰值时均匀分布信源有最大熵；

- 高斯信源的熵:

  N维离散时间高斯信源熵仅与协方差矩阵有关:

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

- 离散集与连续集之间的平均互信息的计算:

$$I(X;Y)=\sum_x p(x)\int_Y p(y/x)\log\frac{p(y/x)}{q(y)}dy$$

课后习题

- P.77
  4.1, 4.4, 4.5, 4.10