第 5 章 连续信源和连续信道

CHAPTER 5

5.1 连续信源

5.1.1 连续信源的数学模型

在 2.5 节我们讲过，对连续随机变量来讲，最基本的统计量是概率密度函数。因此连续信源用概率密度函数来描述。对随机变量 X 存在函数 $p(x)$，且

$$p(x) \geqslant 0$$

$$\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1$$

则 $p(x)$就是该连续信源的概率密度函数。

$$F(x) = P(X < x) = \int_{-\infty}^{x} p(\alpha)\mathrm{d}\alpha$$

称为概率分布函数。

因此，简单连续信源的模型写为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x & \displaystyle\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1 \\ p(x) & \end{bmatrix}$$

5.1.2 连续信源的熵和互信息

在 2.5 节我们已经讲过，连续信源的熵是由绝对熵和微分熵两部分构成的，绝对熵为一无穷大项，通常将其舍去，在不引起混淆的情况下，将微分熵简称为连续信源的熵。

【定义 5-1】 对于连续信源 X，其概率密度函数为 $p(x)$，则该连续信源的熵为

$$H(X) = -\int_{-\infty}^{+\infty} p(x)\log p(x)\mathrm{d}x \tag{5-1}$$

连续信源的熵与离散信源的熵具有类似的形式，但其意义不相同。连续信源的熵与离散信源的熵相比，去掉了一个无穷大项。连续信源的不确定性本应该为无穷大，但是由于实际应用中常常关心的是熵之间的差值，无穷大项可相互抵消，故这样定义连续信源的熵不会影响讨论所关心的互信息量、信道容量等。需要强调的是，连续信源的熵只是熵的相对值，不是绝对值，而离散信源的熵是绝对值。

定义 5-1 给出的是单个连续随机变量的熵，如果有多个随机变量，可以定义联合熵、条件熵、互信息等概念。

【定义 5-2】 设有两个连续随机变量 X 和 Y，其联合熵为

$$H(X,Y)=-\iint_{-\infty}^{+\infty} p(xy)\log p(xy)\mathrm{d}x\mathrm{d}y \tag{5-2}$$

式中 $p(xy)$为二维联合概率密度函数。

【定义 5-3】 设有两个连续随机变量 X 和 Y，其条件熵为

$$H(X\mid Y)=-\iint_{-\infty}^{+\infty} p(xy)\log p(x\mid y)\mathrm{d}x\mathrm{d}y \tag{5-3}$$

或者

$$H(Y\mid X)=-\iint_{-\infty}^{+\infty} p(xy)\log p(y\mid x)\mathrm{d}x\mathrm{d}y \tag{5-4}$$

式中 $p(x|y)$和 $p(y|x)$为条件概率密度函数。

【定义 5-4】 两个连续随机变量 X 和 Y 之间的平均互信息量为

$$I(X;Y)=H(X)-H(X\mid Y)=H(Y)-H(Y\mid X) \tag{5-5}$$

由该定义可以得到

$$I(X;Y)=H(X)+H(Y)-H(X,Y)$$

证明：

$$\begin{aligned}
I(X;Y)&=H(X)-H(X\mid Y)\\
&=H(X)+\iint_{-\infty}^{+\infty} p(xy)\log p(x\mid y)\mathrm{d}x\mathrm{d}y\\
&=H(X)+\iint_{-\infty}^{+\infty} p(xy)\log \frac{p(xy)}{p(y)}\mathrm{d}x\mathrm{d}y\\
&=H(X)+\int_{-\infty}^{+\infty} p(y)\log \frac{1}{p(y)}\mathrm{d}y+\iint_{-\infty}^{+\infty} p(xy)\log p(xy)\mathrm{d}x\mathrm{d}y\\
&=H(X)+H(Y)-H(X,Y)
\end{aligned}$$

由上述推导可以看出，有

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

并且

$$\begin{aligned}
H(X,Y)-H(X)-H(Y)&=\iint_{-\infty}^{+\infty} p(xy)\log \frac{p(x)p(y)}{p(xy)}\mathrm{d}x\mathrm{d}y\\
&\leqslant \iint_{-\infty}^{+\infty} p(xy)\left[\frac{p(x)p(y)}{p(xy)}-1\right]\log e\mathrm{d}x\mathrm{d}y=0
\end{aligned}$$

故有

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

上述定义与离散信源的对应关系式完全类似，而且可以证明连续信源的平均互信息也具有非负性。连续信源中各种熵和平均互信息之间的关系也可以用图 2-6 的维拉图表示。但是由于连续信源的熵是微分熵，略掉了一个无穷大项，因此它不具有非负性和极值性。

【例 5-1】 求均值为 m,方差为 σ^2 的高斯分布的熵。

解:高斯随机变量的概率密度为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

则有

$$H(X)=-\int_{-\infty}^{+\infty}p(x)\log p(x)\mathrm{d}x$$
$$=-\int_{-\infty}^{+\infty}p(x)\left[\log\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}\right]\mathrm{d}x$$

如果取对数的底为 e,则

$$H(X)=-\int_{-\infty}^{+\infty}p(x)\left[-\ln\sqrt{2\pi}\sigma-\frac{1}{2\sigma^2}(x-m)^2\right]\mathrm{d}x$$
$$=\ln\sqrt{2\pi}\sigma+\frac{1}{2\sigma^2}\sigma^2=\ln\sqrt{2\pi\mathrm{e}}\sigma$$

【例 5-2】 设随机变量 X 和 Y 的联合概率密度为

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

求:(1) $H(X)$和 $H(Y)$各是多少?

(2) $H(X|Y)$和 $H(Y|X)$各是多少?

(3) $H(X,Y)$是多少?

(4) $I(X;Y)$是多少?

解:

(1)

$$p(x)=\int_{-\infty}^{+\infty}p(xy)\mathrm{d}y=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{(x-m_x)^2}{2\sigma_x^2}\right\}$$

故

$$H(X)=-\int_{-\infty}^{+\infty}p(x)\log p(x)\mathrm{d}x=\ln\sqrt{2\pi\mathrm{e}}\sigma_x$$

同理可得

$$H(Y)=\ln\sqrt{2\pi\mathrm{e}}\sigma_y$$

(2) 由条件熵定义,有

$$H(X\mid Y)=-\iint_{-\infty}^{+\infty}p(xy)\log p(x\mid y)\mathrm{d}x\mathrm{d}y$$
$$=\iint_{-\infty}^{+\infty}p(xy)\left\{\ln\sqrt{2\pi\sigma_x^2(1-\rho^2)}+\frac{1}{2}\left[\frac{(x-m_x)^2}{(1-\rho^2)\sigma_x^2}-\frac{2\rho(x-m_x)(y-m_y)}{(1-\rho^2)\sigma_x\sigma_y}+\right.\right.$$

$$\left.\left.\frac{(y-m_y)^2}{(1-\rho^2)\sigma_y^2}-\frac{(y-m_y)^2}{\sigma_y^2}\right]\ln e\right\}\mathrm{d}x\mathrm{d}y$$

$$=\ln\sqrt{2\pi\sigma_x^2(1-\rho^2)}+\frac{\log e}{2}\left(\frac{1}{1-\rho^2}-\frac{2\rho^2}{1-\rho^2}+\frac{1}{1-\rho^2}-1\right)=\ln\sqrt{2\pi e\sigma_x^2(1-\rho^2)}$$

同理可得

$$H(Y\mid X)=\ln\sqrt{2\pi e\sigma_y^2(1-\rho^2)}$$

（3）由联合熵定义，有

$$H(X,Y)=-\iint_{-\infty}^{+\infty}p(xy)\log p(xy)\mathrm{d}x\mathrm{d}y$$

$$=\iint_{-\infty}^{+\infty}p(xy)\left\{\ln 2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}+\frac{1}{2}\left[\frac{(x-m_x)^2}{(1-\rho^2)\sigma_x^2}-\frac{2\rho(x-m_x)(y-m_y)}{(1-\rho^2)\sigma_x\sigma_y}+\right.\right.$$

$$\left.\left.\frac{(y-m_y)^2}{(1-\rho^2)\sigma_y^2}\right]\ln e\right\}\mathrm{d}x\mathrm{d}y$$

$$=\ln 2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}+\frac{\log e}{2}\left(\frac{1}{1-\rho^2}-\frac{2\rho^2}{1-\rho^2}+\frac{1}{1-\rho^2}\right)$$

$$=\ln\left[2\pi e\sigma_x\sigma_y\sqrt{(1-\rho^2)}\right]$$

（4）由互信息量定义，有

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

$$=\ln\sqrt{2\pi e}\sigma_x-\ln\sqrt{2\pi e\sigma_x^2(1-\rho^2)}=-\ln\sqrt{(1-\rho^2)}$$

上述结果表明，两个高斯随机变量的各自熵只与各自的方差有关。条件熵与相关系数 ρ 有关，当 $\rho=0$ 时，即 X 和 Y 互不相关时，或者说互相独立时，则有 $H(X)=H(X|Y)$ 和 $H(Y)=H(Y|X)$。联合熵也与 ρ 有关，而互信息量仅与 ρ 有关，与方差无关，当 $\rho=0$ 时，$I(X;Y)=0$。

5.2 连续信道及其信道容量

对于连续信道，其输入和输出均为连续的随机信号，但从时间关系上来分，可以分为时间离散信号和时间连续信号两大类。当信道输入输出只能在特定时刻变化，即时间为离散值时，称信道为离散时间信道(或时间离散信道)。当信道的输入输出是随时间变化的，即时间为连续值时，称为连续信道或者波形信道。我们在1.3节已经介绍过这些信号。下面将分别讨论这两种类型的信道。

5.2.1 时间离散信道

设时间离散信道的输入和输出集分别为 X 和 Y。N 个单元时间信道的输入序列为 $x=\{x_1, x_2, \cdots, x_N\}$；输出序列为 $y=\{y_1, y_2, \cdots, y_N\}$。信道转移概率密度为 $p(y|x)$。类似于离散信道，若

$$p(y \mid x)=\prod_{n=1}^{N} p(y_n \mid x_n)$$

就称该信道是无记忆信道。

这样，对于时间离散信道的研究就可以归结为对单位时间段上信道的输入、输出和干扰之间的关系的研究，也就是单个符号的传输问题的研究了。信道容量 C 定义为

$$C=\max_{p(x)}\{I(X;Y)\}$$

其中 $p(x)$ 为输入事件 x 的概率密度。

由于输入和干扰是相互独立的，对于一维随机变量，其信道模型可以表示为

$$Y=X+N$$

式中，X 为输入随机变量，Y 为输出随机变量，N 为随机噪声，且 X 和 N 统计独立。

设随机变量 X 和 N 的概率密度分别为 $p_X(x)$ 和 $p_n(z)$，不难求得随机变量 Y 在 X 条件下的概率密度为

$$p(y \mid x)=p_N(y-x)=p_N(z)$$

则有

$$\begin{aligned}
H(Y \mid X) &= -\iint_{-\infty}^{+\infty} p(xy)\log p(y \mid x)\,\mathrm{d}x\,\mathrm{d}y \\
&= -\iint_{-\infty}^{+\infty} p_X(x)p(y \mid x)\log p(y \mid x)\,\mathrm{d}x\,\mathrm{d}y \\
&= -\iint_{-\infty}^{+\infty} p_X(x)p_N(y-x)\log p_N(y-x)\,\mathrm{d}x\,\mathrm{d}y \\
&= -\int_{-\infty}^{\infty} p_X(x)\int_{-\infty}^{\infty} p_N(z)\log p_N(z)\,\mathrm{d}z\,\mathrm{d}x \\
&= -\int_{-\infty}^{\infty} p_X(x)H(N)\,\mathrm{d}x = H(N)
\end{aligned}$$

式中，$H(N)$ 为信道噪声的熵，因此互信息为

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

由上式可以看出，简单加性噪声信道的互信息由输出熵和噪声熵所决定。若输入信源 X 和噪声信源 N 分别为均值为 0、方差为 σ_X^2 和 σ_N^2 的高斯分布，则随机变量 Y 为均值为 0、方差为 $\sigma_X^2+\sigma_N^2$ 的高斯分布。所以

$$\begin{aligned}
I(X;Y) &= H(Y)-H(N) \\
&= \frac{1}{2}\log\left[2\pi\mathrm{e}(\sigma_X^2+\sigma_N^2)\right]-\frac{1}{2}\log\left[2\pi\mathrm{e}\sigma_N^2\right]=\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right)
\end{aligned}$$

当 σ_X^2/σ_N^2 任意大时，则 $I(X;Y)$ 同样也可以任意大。由于实际中信号和噪声的能量是有限的，所以我们所研究的时间离散连续信道的容量是在功率受限条件下进行的。

【定义 5-5】 对于输入信号平均功率不大于 S 的时间离散信道的容量定义为

$$C=\sup_{n,p_n}\frac{1}{n}I(X;Y)$$

式中上限是对所有的 n 和所有的概率分布 p_n 上求的。在无记忆条件下，时间离散信道容量为

$$C=\max_{p_n} I(X;Y)$$

对于平均功率受限的、最简单的一维时间离散可加高斯噪声信道的互信息为

$$I(X;Y)=H(Y)-H(N)$$

因此信道容量为

$$C=H(Y)-H(N)=\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right)$$

非高斯性加性噪声信道容量的计算相当复杂，只能给出其上下限，有下述定理存在。

【定理 5-1】 假设输入信源的平均功率小于 σ_X^2，信道加性噪声平均功率为 σ_N^2，则可加噪声信道容量 C 满足

$$\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma^2}\right)\leqslant C\leqslant\frac{1}{2}\log\left(\frac{\sigma_X^2+\sigma_N^2}{\sigma^2}\right)$$

式中 σ^2 为噪声的熵功率。

证明：对于加性噪声信道

$$Y=X+N$$

当输入信源和噪声的均值分别为 0 时，信道的输出功率为 $\sigma_X^2+\sigma_N^2$。由于

$$H(Y)\leqslant\frac{1}{2}\log\left[2\pi e(\sigma_X^2+\sigma_N^2)\right]$$

且

$$\sigma^2=\frac{1}{2\pi e}e^{2H(N)}$$

即

$$H(N)=\frac{1}{2}\log(2\pi e\sigma^2)$$

故有

$$\begin{aligned}C&=\max_{p}\left[H(Y)-H(N)\right]=\max_{p}\left[H(Y)\right]-H(N)\\&\leqslant\frac{1}{2}\log\left[2\pi e(\sigma_X^2+\sigma_N^2)\right]-\frac{1}{2}\log(2\pi e\sigma^2)=\frac{1}{2}\log\left(\frac{\sigma_X^2+\sigma_N^2}{\sigma^2}\right)\end{aligned}$$

故不等式右端成立。当噪声满足高斯分布时，则有 $\sigma^2=\sigma_N^2$，等号成立。

由于任何一个信源的熵功率都小于或等于其平均功率，即

$$\sigma^2\leqslant\sigma_N^2$$

所以有

$$\frac{1}{2}\log\left[2\pi e(\sigma_X^2+\sigma_N^2)\right]\geqslant\frac{1}{2}\log\left[2\pi e(\sigma_X^2+\sigma^2)\right]$$

当选择输入信源功率为 σ_X^2 的高斯变量时，有

$$C\geqslant I(X;Y)=H(Y)-H(N)\geqslant\frac{1}{2}\log\left[2\pi e(\sigma_X^2+\sigma^2)\right]$$

$$-\frac{1}{2}\log(2\pi e\sigma^2)=\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma^2}\right)$$

因此

$$\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma^2}\right)\leqslant C\leqslant\frac{1}{2}\log\left(\frac{\sigma_X^2+\sigma_N^2}{\sigma^2}\right)$$

上述定理表明，当噪声功率 σ_N^2 给定后，高斯型干扰是最坏的干扰，此时信道容量 C 最小。因此，在实际应用中和科学研究中，往往把干扰视为高斯分布，研究最坏的情况通常是比较安全的。

5.2.2 连续信道

时间连续信道也称作波形信道。设输入和输出为随机过程 $X(t)$和 $Y(t)$，设 $N(t)$为随机噪声，那么简单的加性噪声信道模型可以表示为

$$Y(t)=X(t)+N(t)$$

输入、输出和噪声都是时间 t 的函数。

由于信道的带宽总是有限的，根据随机信号采样定理，可以将一个时间连续的信道变换成时间离散的随机序列进行处理。设输入、噪声和输出随机序列分别为 X_i、N_i 和 Y_i，$i=1,2,\cdots,n$，则有

$$Y_i=X_i+N_i\quad(i=1,2,\cdots,n)$$

下面将讨论平均功率受限情况下时间连续的高斯信道。

设高斯信道的平均功率为 σ_N^2，即

$$D[N(t)]=\sigma_N^2$$

对于随机序列 N_i，$i=1,2,\cdots,n$，则有

$$D[N_i]=\sigma_N^2$$

因为高斯白噪声的各样本值彼此相互独立，那么 n 维高斯分布的联合概率密度为

$$p(z)=p(z_1z_2\cdots z_n)=\frac{1}{(2\pi\sigma_N^2)^{n/2}}\exp\left\{-\frac{z_1^2+z_2^2+\cdots+z_n^2}{2\sigma_N^2}\right\}$$

对于加性噪声信道，由概率理论可知

$$p(y\mid x)=p(z)=\prod_{i=1}^{n}p(z_i)=\prod_{i=1}^{n}p(y_i\mid x_i)$$

由于信道是无记忆的，那么 n 维随机序列的平均互信息满足

$$I(X;Y)\leqslant\sum_{i=1}^{n}I(X_i;Y_i)$$

因此时间连续信道的信道容量为

$$C=\max_{p(x)}I(X;Y)=\max_{p(x)}\sum_{i=1}^{n}I(X_i;Y_i)$$

若信道为高斯信道，则时间连续信道的信道容量为

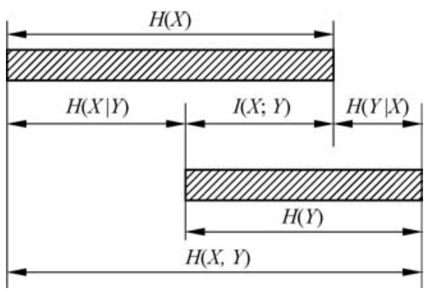
$$C=\frac{n}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right)$$

达到该信道容量的条件是，n 维输入随机序列中的每一分量都必须是均值为 0、方差为 σ_X^2 且相互独立的高斯变量。

5.3 本章小结

本章简要介绍了连续信源和连续信道，见表 5-1。

表 5-1 本章小结

<table>
<tr><td rowspan="6">连续信源</td><td colspan="3">模型：$\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}x & \\ p(x) & \int_{-\infty}^{\infty}p(x)\mathrm{d}x=1\end{bmatrix}$，概率密度</td></tr>
<tr><td colspan="3">熵：$H(X)=-\int_{-\infty}^{\infty}p(x)\log p(x)\mathrm{d}x$</td></tr>
<tr><td colspan="3">共熵：$H(X,Y)=-\iint_{-\infty}^{+\infty}p(xy)\log p(xy)\mathrm{d}x\mathrm{d}y$</td></tr>
<tr><td colspan="3">条件熵：$H(X\mid Y)=-\iint_{-\infty}^{+\infty}p(xy)\log p(x\mid y)\mathrm{d}x\mathrm{d}y$</td></tr>
<tr><td colspan="3">平均互信息：$I(X;Y)=H(X)-H(X\mid Y)=H(Y)-H(Y\mid X)$</td></tr>
<tr><td colspan="3">关系：
</td></tr>
<tr><td rowspan="3">连续信道（加性噪声）</td><td rowspan="2">时间离散信道（单符号连续信道）</td><td>噪声为高斯分布
$C=H(Y)-H(N)=\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right)$</td><td>噪声为一般分布
$\frac{1}{2}\log\left(1+\frac{\sigma_X^2}{\sigma^2}\right)\leqslant C\leqslant\frac{1}{2}\log\left(\frac{\sigma_X^2+\sigma_N^2}{\sigma^2}\right)$</td></tr>
<tr><td colspan="2">这表明，当噪声功率 σ_N^2 给定后，高斯型干扰是最坏的干扰，此时信道容量 C 最小</td></tr>
<tr><td colspan="3">时间连续信道，且噪声为高斯分布：$C=\frac{n}{2}\log\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right)$</td></tr>
</table>

连续信源主要讲了信源的模型、熵、共熵、条件熵和平均互信息。需要注意的是，连续信源的熵不同于离散信源的熵。连续信源的熵是微分熵，与离散信源的熵相比，去掉了一个无穷大项。

连续信道分时间离散和时间连续两种情况讨论，信道上的噪声主要讨论了加性高斯噪声。给出了不同情况下的信道容量。

5.4 习题

5-1 设信源的输出幅度被限定在(a, b)内。证明：在限定范围内，当输出信号的概率密度是均匀分布时，信源具有最大熵

$$H_{\max}[x, p(x)] = \log(b-a)$$

5-2 设随机变量 X 的概率密度为

$$p(x) = \frac{1}{2}\lambda \mathrm{e}^{-\lambda|x|}$$

求随机变量 X 的熵。

5-3 证明连续信源 X 的熵 $H(X)$是关于 X 的概率密度函数 $p(x)$的上凸函数。

5-4 对于连续型随机序列 X_i 和 Y_i $(i=1, 2, \cdots, n)$，证明：当信源 X_i 无记忆时，则有

$$I(X;Y) \geqslant \sum_{i=1}^{n} I(X_i;Y_i)$$

当信道 $p(y|x)$无记忆时，则有

$$I(X;Y) \leqslant \sum_{i=1}^{n} I(X_i;Y_i)$$

并问在什么条件下，上述两式等号成立。