基础信息论

连续信源熵

华中科技大学电信学院

连续信源

信源 → $X_1, X_2, X_3, \ldots\ldots$

信源输出被抽象为随机变量序列（随机过程）

A为(a,b)

- 信源输出的随机变量取值于某一连续区间，为连续信号
- 消息的个数是无穷值。
- 比如：人发出的语音信号$X(t)$、模拟的电信号等

学习目标

- 构建连续信源的数学模型
- 辨识连续信源的分类
- 分析连续熵的性质

连续信源的熵

连续信源

- 实际应用：

信源的输出往往是时间的连续函数，如语音信号、电视图像等。由于它们的取值既是连续的又是随机的，称为连续信源，且信源输出的消息可以用随机过程描述。

- 单变量连续信源的数学模型：

$$X:\begin{Bmatrix} R \\ p(x) \end{Bmatrix}$$

其中 $p(x) \leftarrow p_X(x) = \frac{dF(x)}{dx}$ **边缘概率密度函数**

$F(x):X$ 的概率分布函数

并满足 $\int_{-\infty}^{+\infty} p(x)dx = 1$

连续信源的熵如何计算

回忆离散信源中单符号信源熵的定义：

$$H(X) = -\sum_{i=1}^{n} p(x_i) \cdot \log p(x_i)$$

问题：在连续信源中，能否直接套用离散信源的公式?

回答：不可以。因为在连续信源中，随机变量的取值有无穷多种可能。每个具体取值的概率等于零。

计算连续信源熵的两种方案：

1. 将连续信源数字化，再用离散熵计算。

2. 先进行抽样，成为时间离散信号。再把抽样序列看作量化单位△趋于0时的情况，然后定义计算信源熵。

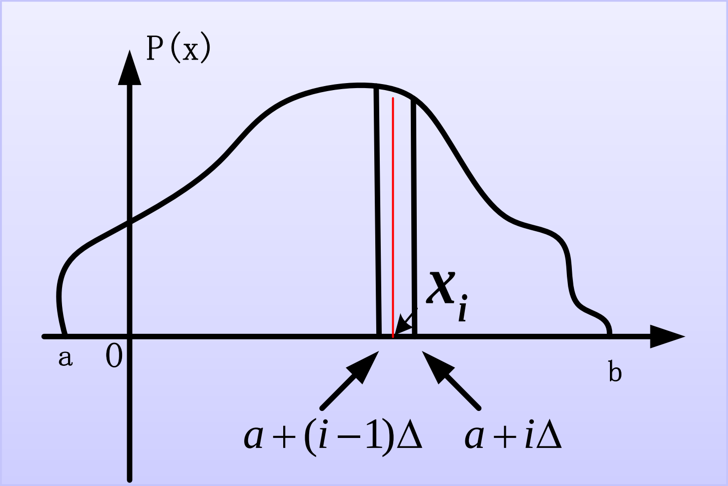
连续信源的熵的计算

思路：将连续信源转化为离散信源后取极限进行计算。

连续信源 　取值范围(a,b)

↓

离散信源 　将(a,b)划分为n个小区间，用小区间内的一个点代表整个小区间。（**将取值离散化**）



$$\Delta = \frac{b-a}{n}$$

则变量落在第i个小区间的概率为：

$$P(a+(i-1)\Delta \le X \le a+i\Delta)$$

$$= \int_{a+(i-1)\Delta}^{a+i\Delta} p(x)dx$$

$$= p(x_i)\Delta$$

中值定理

连续信源的熵的计算(续)

连续信源：

$$X:\begin{Bmatrix} R \\ p(x) \end{Bmatrix}$$

转化为离散信源：

$$\begin{bmatrix} x_1 & \cdots & x_i & \cdots & x_n \\ p(x_1)\cdot\Delta & \cdots & p(x_i)\cdot\Delta & \cdots & p(x_n)\cdot\Delta \end{bmatrix}$$

离散信源的熵：

$$H(X) = -\sum_{i=1}^{n} p(x_i)\cdot\Delta\cdot\log[p(x_i)\cdot\Delta]$$

连续信源的熵：

$$\lim_{\substack{n\to\infty \\ \Delta\to 0}} H(X) = -\lim_{\substack{n\to\infty \\ \Delta\to 0}} \sum_{i=1}^{n} p(x_i)\cdot\Delta\cdot\log[p(x_i)\cdot\Delta]$$

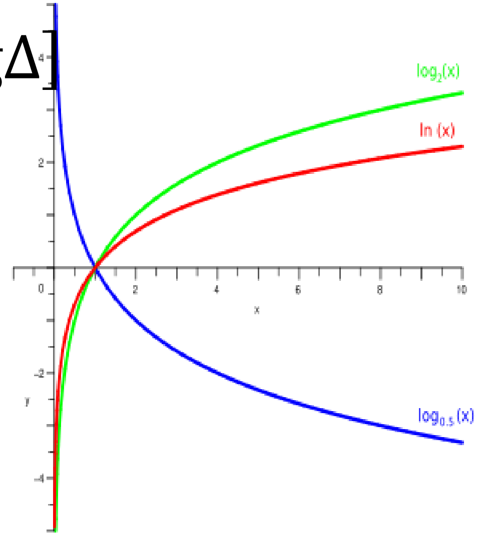
量化单位△趋于0

连续信源的熵的计算(续)

$$\lim_{\substack{n\to\infty\\ \Delta\to 0}} H(X) = -\lim_{\substack{n\to\infty\\ \Delta\to 0}} \sum_{i=1}^{n}[p(x_i)\cdot\Delta\cdot\log p(x_i)] - \lim_{\substack{n\to\infty\\ \Delta\to 0}} \sum_{i=1}^{n}[p(x_i)\cdot\Delta\cdot\log\Delta]$$

$$= -\int_a^b p(x)\cdot\log p(x)dx \qquad - \lim_{\Delta\to 0}\left[\underbrace{\int_a^b p(x)\,dx}_{=1}\right]\cdot\log\Delta$$

$$\boxed{= -\int_a^b p(x)\cdot\log p(x)dx - \lim_{\Delta\to 0}[\log\Delta]}$$



对比单符号离散信源的信源熵定义：

$$H(X) = -\sum_{i=1}^{n} p(x_i)\cdot\log p(x_i)$$

发 现：

- **连续信源的熵，多了一项，且多出来的一项为无穷大项**
- **连续信源的熵为无穷大。**

定义：连续信源的熵

- 连续信源的熵为

$$\boxed{H_c(X) = -\int_{-\infty}^{+\infty} p(x) \cdot \log p(x) dx}$$

前式丢掉无穷大项，只保留第一项

- 说明：

1. $H_C(X)$形式上与离散信源的熵统一，但意义不同：它去掉了一个无穷项，不能代表连续信源所携带的平均信息或平均不确定度，连续信源的不确定性应为无穷大。

差熵　相对熵

2. 实际应用中常常关心的是熵之间的差值，无穷项可相互抵消，故这样定义连续信源的熵不会影响讨论所关心的交互信息量、信息容量和率失真函数。
3. 需要强调的是连续信源熵的值只是熵的相对值，不是绝对值，而离散信源熵的值是绝对值。

其他连续熵的定义

$$H_c(XY) = -\iint_{R^2} p(xy)\log p\,(xy)dxdy$$

$$H_c(X|Y) = -\iint_{R^2} p(xy)\log p\,(x|y)dxdy$$

$$H_c(Y|X) = -\iint_{R^2} p(xy)\log p\,(y|\mathrm{x})dxdy$$

几种特殊连续信源的熵

均匀分布的连续信源

- 一维均匀分布的信源：

$$p(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其它}\end{cases}$$

代入相对熵的公式：

$$H_c(X)=-\int_{-\infty}^{\infty}p(x)\cdot\log p(x)dx$$

$$=-\int_a^b\frac{1}{b-a}\cdot\log\frac{1}{b-a}dx$$

$$=-\frac{1}{b-a}\cdot\log\frac{1}{b-a}\cdot(b-a)$$

$$=\log(b-a)$$

当$b-a<1$时，
则$H(X)<0$

连续信源的熵不具有非负性

高斯分布的连续信源

- 一维高斯分布：$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$H_c(X)=-\int_{-\infty}^{\infty}p(x)\cdot\log p(x)dx \quad =-\int_{-\infty}^{\infty}p(x)\cdot\log\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}\right]dx$$

$$=-\int_{-\infty}^{\infty}p(x)\cdot\log\frac{1}{\sqrt{2\pi}\sigma}dx+\int_{-\infty}^{\infty}p(x)\cdot\log e\cdot\frac{(x-m)^2}{2\sigma^2}dx$$

$$=-\log\frac{1}{\sqrt{2\pi}\sigma}+\frac{\log e}{2\sigma^2}\cdot\underline{\int_{-\infty}^{\infty}p(x)\cdot(x-m)^2dx} \rightarrow =\sigma^2$$

$$=-\log\frac{1}{\sqrt{2\pi}\sigma}+\frac{\log e}{2\cancel{\sigma^2}}\cdot\cancel{\sigma^2} \quad =\log\sqrt{2\pi}\,\sigma+\frac{1}{2}\log e$$

$$=\frac{1}{2}\log 2\pi e\sigma^2$$

高斯信源的熵仅与方差有关。

方差影响信源整体特性，而均值无影响

指数分布的连续信源

分部积分法：

$$\int udv = uv - \int vdu$$

- 指数分布：

$$p(x) = \frac{1}{m} e^{-\frac{x}{m}} \quad x \ge 0$$

$$H_c(X) = -\int_{-\infty}^{\infty} p(x) \cdot \log p\,(x) dx \quad = -\int_{-\infty}^{\infty} p(x) \cdot \log\left[\frac{1}{m} e^{-\frac{x}{m}}\right] dx$$

$$= -\int_{-\infty}^{\infty} p(x) \cdot \log\frac{1}{m} dx + \int_{-\infty}^{\infty} p(x) \cdot \log e \cdot \frac{x}{m} dx$$

$$= \log m + \frac{\log e}{m} \int_{-\infty}^{\infty} p(x) \cdot x dx \quad = \log m\, e$$

其中：

$$\int_{-\infty}^{\infty} p(x) \cdot x dx = \int_{0}^{\infty} x \cdot \frac{1}{m} e^{-\frac{x}{m}} dx = \frac{1}{\cancel{m}} \int_{0}^{\infty} x \cdot (-\cancel{m}) d e^{-\frac{x}{m}}$$

$$= -e^{-\frac{x}{m}} x \Big|_0^{\infty} + \int_0^{\infty} e^{-\frac{x}{m}} dx = 0 + \left(-m e^{-\frac{x}{m}}\right)\Big|_0^{\infty} \quad = m$$

只取决于均值

连续熵的性质及最大连续熵定理

连续熵的性质

1.连续熵可为负值

离散信源： $$H(X) = -\sum_{i=1}^{n} p(x_i) \cdot \log p(x_i)$$ **非负性**

连续信源： $$H(X) = \underline{H_c(X)} + \infty$$ **绝对熵为无穷大**

可正可负

2. 可加性

$$\begin{cases} H_c(XY) = H_c(X) + H_c(Y|X) \\ H_c(XY) = H_c(Y) + H_c(X|Y) \end{cases}$$

可推广：

$$H_c(X_1X_2 \cdots X_N) = H_c(X_1) + H_c(X_2|X_1) + \cdots + H_c(X_N|X_1 \cdots X_{N-1})$$

连续熵的性质

3. 平均互信息量的非负性、对称性

连续信源：

$$
\begin{cases}
I_c(X;Y) = H_c(X) - H_c(X|Y) \\
I_c(X;Y) = H_c(Y) - H_c(Y|X) \\
I_c(X;Y) = [H_c(X) + H_c(Y)] - H_c(XY)
\end{cases}
$$

非负性：

$I_c(X;Y) \geq 0$ **当** X **和** Y **相互独立时，** $I_c(X;Y) = 0$

对称性：

$$I_c(X;Y) = I_c(Y;X)$$

连续信源的最大熵

离散信源：

等概率分布时熵最大，等于 $\log n$。

连续信源：

当限定条件不同时，结论也不同。

- 在具体应用中，仅讨论连续信源的两种限定情况：
- 1. 信源输出的幅度受限；
- 2. 信源输出的平均功率受限。

峰值功率受限时的最大熵

(1) 峰值功率受限时的最大熵

峰值功率： 设 $X \in [a, b]$， $P_{峰} = \dfrac{[\max(|a|,|b|)]^2}{R}$

$$= [\max(|a|,|b|)]^2 \ (R = 1)$$

峰值功率受限： 随机变量取值范围必须为有限值$[a, b]$。

结论：峰值功率受限时，均匀分布的熵最大。

均匀分布

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{其它} \end{cases}$$

峰值功率受限时的最大熵证明

峰值功率受限

$$H_c[X,q(x)] = -\int_{-\infty}^{\infty} q(x)\cdot\log q(x)dx \quad = -\int_a^b q(x)\cdot\log q(x)dx$$

任意分布

$$= \int_a^b q(x)\cdot\log\frac{1}{q(x)}dx \quad = \int_a^b q(x)\cdot\log[\frac{1}{q(x)}\cdot\frac{p(x)}{p(x)}]dx$$

$$= \int_a^b q(x)\cdot\log[\frac{1}{p(x)}\cdot\frac{p(x)}{q(x)}]dx$$

分母两项颠倒位置

均匀分布

$$p(x) = \frac{1}{b-a}$$

$$= \int_a^b q(x)\cdot\log\frac{1}{p(x)}dx + \int_a^b q(x)\cdot\log\frac{p(x)}{q(x)}dx$$

——为常数

$$\boxed{\ln x \le x-1}$$

$$\le \log(b-a) + \log e\cdot\int_a^b q(x)\cdot\left[\frac{p(x)}{q(x)}-1\right]dx$$

$$= 0$$

$$= \log(b-a)$$

$$= H_c[X,p(x)]$$

平均功率受限时的最大熵

(2) 平均功率受限时的最大熵

平均功率：随机变量 X 某次实验的结果是 x

功率为： $p = x^2$

平均功率为： $P_{avg} = E(x^2)$

均值 m 为零时： $P_{avg} = E[x^2] = \sigma^2 = E[(x-m)^2]$

平均功率受限：均值为0，方差受限的随机变量。

结论: 平均功率受限时, 高斯(正态)分布的熵最大。

$$H_c[X, q(x)] = \int_{-\infty}^{\infty} q(x) \cdot \log[\frac{1}{p(x)} \cdot \frac{p(x)}{q(x)}]dx$$

任意分布　　　　高斯分布

$$= \int_{-\infty}^{\infty} q(x) \cdot \log \frac{1}{p(x)} dx + \int_{-\infty}^{\infty} q(x) \cdot \log \frac{p(x)}{q(x)} dx$$

$\boxed{\ln x \leq x - 1}$

$$\leq -\int_{-\infty}^{\infty} q(x) \cdot \log p(x)\, dx + \log e \cdot \underbrace{\int_{-\infty}^{\infty} q(x) \cdot [\frac{p(x)}{q(x)} - 1]dx}_{=0}$$

$$= -\int_{-\infty}^{\infty} q(x) \cdot \log[\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{x^2}{2\sigma^2}}]dx$$

$$= \log \sqrt{2\pi}\, \sigma + \frac{1}{2\sigma^2} \cdot \log e \cdot \underbrace{\int_{-\infty}^{\infty} q(x) \cdot x^2 dx}_{=\sigma^2}$$

$$= \log \sqrt{2\pi}\, \sigma + \frac{1}{2} \cdot \log e$$

$$= \frac{1}{2} \log 2\,\pi e \sigma^2 \quad = H_c[X, p(x)] \quad = \frac{1}{2} \log 2\,\pi e P_{avg} \qquad \therefore \text{得证}$$

输出信号幅度受限条件下的最大熵

- 定理：对于服从均匀分布的随机变量X，具有最大输出熵。

证明：该问题为在约束条件$\int_a^b p(x)dx = 1$下，

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{其他} \end{cases}$$

求$H(X) = -\int_a^b p(x)\log p(x)dx$ 达到最大值的$p(x)$。

令 $F[p(x)] = H(X) + \lambda\left[\int_a^b p(x)dx - 1\right]$

对上式求关于$p(x)$的偏导数，并令其为0，化简后得

$-\log p(x) - 1 + \lambda = 0$ (取e为底的对数)

解得：$p(x) = e^{\lambda-1}$，因为$\int_a^b p(x)dx = \int_a^b e^{\lambda-1}dx = 1$

则有$e^{\lambda-1} = \dfrac{1}{b-a}$，所以$p(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{其他} \end{cases}$

平均功率受限条件下的最大熵

- 定理：对于服从均值为m，方差为σ2的高斯分布的随机变量具有最大输出熵。

证明：该问题为在约束条件

$$\int_{-\infty}^{\infty} p(x)dx = 1\,,\quad \int_{-\infty}^{\infty} xp(x)dx = m,\quad \int_{-\infty}^{\infty} (x-m)^2 p(x)dx = \sigma^2 \text{下,}$$

求$H(X) = -\int_{-\infty}^{\infty} p(x)\log p\,(x)dx$ 达到最大值的$p(x)$。

令$$F[p(x)] = H(X) + \lambda_1\left[\int_{-\infty}^{\infty} p(x)dx - 1\right] + \lambda_2\left[\int_{-\infty}^{\infty} xp(x)dx - m\right] + \lambda_3\left[\int_{-\infty}^{\infty} (x-m)^2 p(x)dx - \sigma^2\right]$$

定理 - 证明 续

令$\dfrac{\partial F[p(x)]}{\partial p(x)}=0$，解得（取$e$为底的对数）

$p(x)=\exp\{\lambda_1-1+\lambda_2 x+\lambda_3(x-m)^2\}$

将上式代入约束条件关系式，可以得到

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

$$H(X)=-\int_{-\infty}^{\infty}p(x)\log p(x)dx$$

$$=\int_{-\infty}^{\infty}p(x)\log\sqrt{2\pi}\,\sigma dx+\int_{-\infty}^{\infty}\frac{(x-m)^2}{2\sigma^2}p(x)dx\log e$$

$$H(X)=\log\sqrt{2\pi}\,\sigma+\frac{1}{2}\log e=\log\sqrt{2\pi e}\bullet\sigma=H(X)$$

总结

- 输出信号幅度受限的连续信源，当满足均匀分布时达到最大输出熵，这与离散信源在以等概率出现达到最大输出熵的结论类似。
- 输出信号平均功率受限条件下，具有高斯分布的连续信源的熵最大，且随平均功率的增加而增加。
- 当峰值功率受限、平均功率受限，连续信源的统计特性分别与两种常见噪声——均匀噪声和高斯噪声的统计特性相一致时，信源具有最大连续熵。
- 因为噪声是一个最不确定的随机过程，而最大的信息量只能从最不确定的事件中获得。

谢谢！

黑晓军
华中科技大学
电子信息与通信学院
Email: heixj@hust.edu.cn
网址： http://eic.hust.edu.cn/aprofessor/heixiaojun

参考资料

- 陈运，信息论与编码，第3版，第6章6.1节，电子工业出版社，2015