# Chapter 2

Xuan

2020 年 9 月 21 日

## 1　条件互信息

在有三个变量的情况下，符号 $x_i$ 与符号对 $(y_j,z_k)$ 之间的互信息量定义为

$$I(x_i;y_j,z_k)=log\frac{p(x_i|y_j,z_k)}{p(x_i)}$$

条件互信息量是在给定 $z_k$ 条件下，$x_i$ 与 $y_j$ 之间的互信息量，定义为

$$I(x_i;y_j|z_k)=log\frac{p(x_i|y_j,z_k)}{p(x_i|z_k}$$

结合上述公式可写为

$$I(x_i;y_j,z_k)=I(x_i;z_k)+I(x_i;y_j|z_k)$$

### 1.1　相关公式推导

$$\begin{aligned}I(X;Y|Z)&=\sum p(x,y,z)log\frac{p(x,y|z)}{p(x|z)p(y|z)}\\&=\sum_z D(p(x,y|z)||p(x|z)p(y|z))*p(z)\end{aligned}\tag{1}$$

$$\begin{aligned}\frac{p(x,y|z)}{p(x|z)p(y|z)}&=\frac{p(x|y,z)p(y|z)p(x)}{p(x)p(x|z)p(y|z)}\\&=\frac{p(x,y,z)}{p(x)p(y,z)}*\frac{p(x)}{p(x|z)}\end{aligned}\tag{2}$$

将 (2) 带入 (1)

$$(1)=\sum p(x,y,z)log\frac{p(x,y,z)}{p(x)p(y,z)}-\sum p(x,y,z)log\frac{p(x|z)p(z)}{p(x)p(z)} \tag{3}$$

$$\begin{aligned}\sum p(x,y,z)log\frac{p(x|z)p(z)}{p(x)p(z)} &= \sum_{x,y,z}(p(x,z)log\frac{p(x,z)}{p(x)p(z)})p(y|x,z) \\ &= \sum_{x,z}\sum_{y} \\ &= \sum_{x,z}p(x,z)log\frac{p(x,z)}{p(x)p(z)}\sum_{y}p(y|x,z)\end{aligned} \tag{4}$$

将 (4) 带入 (3)

$$\begin{aligned}(3) &= \sum p(x,y,z)log\frac{p(x,y,z)}{p(x)p(y,z)}-\sum p(x,z)log\frac{p(x,z)}{p(x)p(z)} \\ &= I(X;Y,Z)-I(X;Z) \\ &= H(X)-H(X|Y,Z)-(H(X)-H(X|Z)) \\ &= H(X|Z)-H(X|Y,Z)\end{aligned} \tag{5}$$

# 2 信息不增性

对于马尔可夫链 $X->Y->Z$，有

$$I(X;Z)\leq I(Y;Z)$$

## 2.1 公式证明

$$\left.\begin{aligned}&I(X;Y,Z)=I(X;Y)+I(X;Z|Y) \\ &I(X;Z|Y)=\sum_{y}D(p(x,z|y)||p(x|y)p(z|y))p(y)=0\end{aligned}\right\} I(X;Y,Z)=I(X;Y) \tag{6}$$

$$\begin{aligned}I(X;Y,Z)=I(X;Z)+I(X;Y|Z) \\ p(x,y,z)=p(y)p(x|y)p(z|y) \iff p(x,z|y)=p(x|y)p(z|y)\end{aligned} \tag{7}$$

综上所述，$I(X;Y)=I(X;Y,Z)\geq I(X;Z)$

## 2.2 文氏图证明

见 chapter2.jpg

# 3 连续信源的熵和互信息

## 3.1 连续信源熵

$$H_c(x) = -\int_{-\infty}^{\infty} p_x(x) log p_x(x) dx \geq 0$$

## 3.2 能量确定条件下，$\int_{-infty}^{\infty} x^2 p(x) dx$ 在高斯分布情况下最大化熵率

$$H_c(x) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-m)^2}{2\sigma^2}} log[\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-m)^2}{2\sigma^2}}] dx \tag{8}$$

证明过程见 homework1

## 3.3 范围确定条件下，均匀分布最大化熵率

变量 X 的幅度取值限定在 [a,b]，则有 $\int_a^b p_x(x)dx = 1$，当任意 $p_x(x)$ 符合平均分条件，

$$p_x(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & other \end{cases}$$

时，信源达到最大熵

$$\begin{aligned} H_c(x) &= -\int_a^b \frac{1}{b-a} log \frac{1}{b-a} dx \\ &= log(b-a) \end{aligned} \tag{9}$$

# 4 熵的一部分性质

## 4.1 香农辅助定理

对任意的 n 维概率矢量 $P=(p_1,p_2,...,p_n)$ 和 $Q=(q_1,q_2,...,q_n)$，如下不等式成立

$$H(p_1,p_2,...,p_n=-\sum_{i=1}^{n}p_ilogp_i)\leq-\sum_{i=1}^{n}p_ilogq_i$$

该式表明，用不等的概率分布来编码会使熵增

## 4.2 条件熵小于无条件熵

条件熵小于信源熵：$H(Y|X)\leq H(Y)$。当且仅当 y 和 x 相互独立时，$p(y|x)=p(y)$，取等号。

两个条件下的信源熵小于一个条件下的信源熵,即 $H(Z|x,y)\leq H(Z|Y)$，当且仅当 $p(z|x,y)=p(z|y)$ 时取等号。

联合熵小于信源熵之和：$H(X,Y)\leq H(X)+H(Y)$。当且仅当两个集合相互独立时取等号。

# 5 离散有记忆信源的序列熵

对于由两个符号组成的联合信源，有下列结论：

1. $H(X_1,X_2)=H(X_1)+H(X_2|X_1)=H(X_2)+H(X_1|X_2)$(Chain Rule)
2. $H(X_1)\geq H(X_1|X_2),H(X_2)\geq H(X_2|X_1)$

考虑离散平稳信源，有下列结论成立：

1. $H(X_L/X^{L-1})$ 是 L 的单调非增函数
2. $H_L(X)\geq H(X_L|X^{L-1})(H_L(X)$ 为平均符号熵)
3. $H_L(X)$ 是 L 的单调非增序列

4. 当 $L \to \infty$ 时,有 $H_\infty(X) := lim_{L\to\infty} H_L(X) = lim_{L\to\infty} H(X_L|X_1, X_2, ..., X_{L-1})$，其中，$H_\infty(X)$ 称为极限熵，又称为极限信息量

对于齐次，遍历的马尔可夫链，其状态 $s_i$ 由 $(x_{i_1}, ..., x_{i_m})$ 唯一确定，因此有 $p(x_{i_{m+1}}|x_{i_m}, ..., x_{i_1}) = p(x_{i_{m+1}}|s_i)$，可以得到在这种情况下的极限熵为 $H_\infty(x) = \sum_i p(s_i)H(X|s_i)$，其中 $p(s_i)$ 是马尔可夫链的稳态分布。