《信息论基础》 第1次理论作业

南京大学计算机学院 张奕斐 231250022

2.11 相关性的度量

(1)由2.4节熵与互信息的关系可知，

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

由离散型随机变量熵的定义，

$$H(X) = -\sum_{x\in X} p(x)\log_2 p(x)$$

由题意知，X_1和X_2同分布，则

$$H(X_1) = H(X_2)$$

于是对原式

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} = \frac{H(X_1) - H(X_2|X_1)}{H(X_1)} = \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} = \frac{I(X_2;X_1)}{H(X_1)} = \frac{I(X_1;X_2)}{H(X_1)}$$

得证明。

(2)由互信息的性质，有$I(X;Y) \le min(H(X), H(Y))$，特别地，本例中有$H(X_1) = H(X_2)$，因此

$$0 \le I(X_1;X_2) \le H(X_1)$$

于是

$$0 \le \frac{I(X_1;X_2)}{H(X_1)} \le 1$$

即$0 \leq \rho \leq 1$，得证明。

(3)当$\rho = 0$时，$I(X_1; X_2) = 0$，由互信息的定义，互信息始终性质，$I(X;Y) \geq 0$，当且仅当X和Y独立时取等号$I(X;Y) = 0$。则当$\rho = 0$时，X_1和X_2独立。

(4)当$\rho = 1$时，知$I(X_1; X_2) = H(X_1) = H(X_2)$，由互信息的性质，有$I(X;Y) \leq min(H(X), H(Y))$，当$X = Y$时，$I(X;Y) = H(X)$，于是此时$X_1 = X_2$。

2.46 熵的公理化定义

由熵的定义，

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

熵的定义符合公理化定义的3个性质：

标准化（Normalization）：

$$H_2\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

连续性（Continuity）：熵函数 $H_2(p, 1-p)$ 是关于概率 p 的连续函数。

组合法则（Grouping Law）：

$$H_m(p_1, p_2, \ldots, p_m) = H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$

由标准化 $H_2\left(\frac{1}{2}, \frac{1}{2}\right) = 1$ 和连续性，

$$H_2(p, 1-p) = -p\log p - (1-p)\log(1-p)$$

假设对 $m-1$ 个事件，熵的为唯一形式为$H_{m-1}(p_1, p_2, \ldots, p_{m-1}) = -\sum_{i=1}^{m-1} p_i \log p_i$。

对 m 个事件，由组合法则：

$$H_m(p_1, \ldots, p_m) = H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$

第二项展开后即为 $-p_1\log p_1 - p_2\log p_2$，合并后总表达式为

$$H_m(p_1,\ldots,p_m)=-\sum_{i=1}^{m}p_i\log p_i$$

因此由归纳假设，熵的为唯一形式为$H_m(p_1,p_2,\ldots,p_m)=-\sum_{i=1}^{m}p_i\log p_i$。

3.4 AEP

(a)A^n 是所有满足 $|-\frac{1}{n}\log p(x^n)-H|\le\epsilon$ 的序列 x^n 的集合。 根据渐进均分性的推论（典型集的性质），当 $n\to\infty$ 时，$\Pr\{X^n\in A^n\}\to 1$。

(b)B^n 是所有满足 $|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|\le\epsilon$ 的序列 x^n，其中 $\mu=E[X]$，有$\Pr\{X^n\in B^n\}\to 1$。

对于$n>N_1$时有$\Pr\{X^n\in A^n\}>1-\frac{\epsilon}{2}$，$\Pr\{X^n\in B^n\}>1-\frac{\epsilon}{2}$成立，于是$\Pr\{X^n\notin A^n\}\le\frac{\epsilon}{2}$，$\Pr\{X^n\notin B^n\}\le\frac{\epsilon}{2}$。

$$\Pr\{X^n\in A^n\cap B^n\}=1-\Pr\{\overline{A^n}\cup\overline{B^n}\}\ge 1-\epsilon$$

因此，$\Pr\{X^n\in A^n\cap B^n\}\to 1$ 。

(c)对于任意 $x^n\in A^n$，$p(x^n)\ge 2^{-n(H+\epsilon)}$。

$$\sum_{x^n\in A^n\cap B^n}p(x^n)\ge\sum_{x^n\in A^n\cap B^n}2^{-n(H+\epsilon)}=|A^n\cap B^n|\cdot 2^{-n(H+\epsilon)}$$

因此$|A^n\cap B^n|\le 2^{n(H+\epsilon)}$。

(d)由(b)中结论，对于任意$\epsilon>0$，有$\Pr\{X^n\in A^n\cap B^n\}\ge 1-\epsilon$对于$n>N_\epsilon$成立。

对于 $x^n\in A^n$，有 $p(x^n)\le 2^{-n(H-\epsilon')}$。因此:

$$1-\epsilon\le|A^n\cap B^n|\cdot 2^{-n(H-\epsilon')}$$

$$|A^n\cap B^n|\ge(1-\epsilon)\cdot 2^{n(H-\epsilon')}$$

取 $\epsilon=\frac{1}{2}$ ，得证明。

3.11 定理的证明

定理 设$X_1,X_2,\cdots,X_n$为服从$p(x)$的$i.i.d$序列。对$\delta<\frac{1}{2}$及任意的$\delta'>0$，如果$\Pr\{B_\delta^{(n)}\}>1-\delta$，则

$$\frac{1}{n}\log|B_\delta^{(n)}|>H-\delta'$$

(a)对于给定的集合A,B，考虑

$$\Pr(A \cap B) = 1 - \Pr(\overline{A} \cup \overline{B})$$

其中必然有

$$\Pr(\overline{A} \cup \overline{B}) = \Pr(\overline{A}) + \Pr(\overline{B}) - \Pr(\overline{A} \cap \overline{B}) \leq \Pr(\overline{A}) + \Pr(\overline{B})$$

根据题设条件，$\Pr(A) \geq 1 - \epsilon_1$， $\Pr(\overline{A}) \leq \epsilon_1$，同理$\Pr(\overline{B}) \leq \epsilon_2$。
于是

$$\Pr(A \cap B) \geq 1 - \epsilon_1 - \epsilon_2$$

得证明。

因此可得$\Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) \geq 1 - \epsilon - \delta$。

(b) 不等式的验证

步骤3.34：直接应用(a)的结论，得到

$$\Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) \geq 1 - \epsilon - \delta$$

步骤3.35：展开

$$\sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} p(x^n)$$

步骤3.36：对于典型集$A_\epsilon^{(n)}$，其元素满足概率上界：

$$p(x^n) \leq 2^{-n(H-\epsilon)}$$

步骤3.37：共有$|A_\epsilon^{(n)} \cap B_\delta^{(n)}|$个元素，每个元素的概率不超过$2^{-n(H-\epsilon)}$，总和不超过

$$|A_\epsilon^{(n)} \cap B_\delta^{(n)}| \cdot 2^{-n(H-\epsilon)}$$

步骤3.38：由于$A_\epsilon^{(n)} \cap B_\delta^{(n)} \subseteq B_\delta^{(n)}$，集合大小满足：

$$|A_\epsilon^{(n)} \cap B_\delta^{(n)}| \leq |B_\delta^{(n)}|$$

(c)由步骤(b)可得

$$1 - \epsilon - \delta \leq |B_\delta^{(n)}| \cdot 2^{-n(H-\epsilon)}$$

于是

$$|B_\delta^{(n)}| \geq (1-\epsilon-\delta)\cdot 2^{n(H-\epsilon)}$$

证毕。

4.28 过程的混合

(a) Y_i是平稳的。θ以 $\frac{1}{2}$概率选择 X_1或 X_2，且一旦选定后，Y序列完全由该过程生成。X_1和 X_2均为独立同分布（$i.i.d.$）的伯努利过程，其联合分布在时间平移下不变。

(b) Y_i不是独立同分布（$i.i.d.$）过程。一旦θ确定，Y序列完全由X_1或X_2生成，因此相邻Y_i之间存在隐含的关联性，不满足独立同分布（$i.i.d.$）。

(c)当 θ 确定时，Y_i 的熵率为 $H(p_1)$ 或 $H(p_2)$。 由于 θ 以等概率选择，整体熵率为两者的平均。

$$H(Y) = \frac{1}{2}H(p_1) + \frac{1}{2}H(p_2) = -plogp - (1-p)log(1-p)$$

(d)Y_i不满足

$$-\frac{1}{n}\log p(Y_1, Y_2, \dots, Y_n) \to H(Y).$$

当$n \to \infty$时，观测序列几乎必然确定θ（依题述，最终Y将知道观测的是哪个过程），导致$-\frac{1}{n}\log p(Y^n)$收敛至$H(p_1)$或 $H(p_2)$，而非$H(Y)$。

(e)不存在编码使得平均码长收敛至$H(Y)$（依题述，最终Y将知道观测的是哪个过程，因此平均编码长度应该收敛至$H(X_i)$）。

下面对于$Z_i = X_{\theta_i}i, i = 1, 2, ...$展开讨论。

(a') Z_i是平稳的。

(b') Z_i是独立同分布（$i.i.d.$）过程。Z_i服从伯努利分布。

(c') Z_i满足

$$H(Z) = \frac{1}{2}H(p_1) + \frac{1}{2}H(p_2) = -plogp - (1-p)log(1-p)$$

(d') 由于Z_i是平稳的独立同分布过程，因此满足

$$-\frac{1}{n}\log p(Z_1, Z_2, \dots, Z_n) \to H(Z).$$

(e') 存在编码使得平均码长收敛至$H(Z)$。

4.33 链不等式

设$X_1 \to X_2 \to X_3 \to X_4$构成马尔科夫链。

由马尔科夫链的定义，

$$\Pr(X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \cdots, X_1 = x_1) = \Pr(X_{n+1} = x_{n+1}|X_n = x_n)$$

且由互信息的性质，

$$I(X_1; X_4) = H(X_1) - H(X_1|X_4)$$

$$I(X_2; X_3) = H(X_2) - H(X_2|X_3)$$

$$I(X_1; X_3) = H(X_1) - H(X_1|X_3)$$

$$I(X_2; X_4) = H(X_2) - H(X_2|X_4)$$

于是，要证明题述不等式，只需证明

$$H(X_1|X_3) + H(X_2|X_4) \geq H(X_1|X_4) + H(X_2|X_3)$$

根据条件熵链式法则及马尔科夫链的无记忆性，$H(X_1|X_3) = H(X_1, X_2|X_3) - H(X_2|X_1, X_3) = H(X_1, X_2|X_3) - H(X_2|X_1)$

于是

$$H(X_1|X_3) + H(X_2|X_4) = H(X_1, X_2|X_3) - H(X_2|X_1) + H(X_2|X_4)$$

$$\geq H(X_1, X_2|X_3) - H(X_2|X_4) + H(X_2|X_4) = H(X_1|X_4) + H(X_2|X_3)$$

于是

$$I(X_1; X_4) + I(X_2; X_3) \geq I(X_1; X_3) + I(X_2; X_4)$$

得证明。

2025.04.01