信息论第二讲作业解答

中国科学技术大学《信息论 A》006125.01 班助教组

2024 年 4 月 8 日

第 1 题

For random variables X and Y, prove that $H(X+Y) \leq H(X) + H(Y)$ holds.

证明:

$$\begin{aligned} H(X+Y) &\leq H(X+Y) + H(X,Y|X+Y) \\ &= H(X,Y) + H(X+Y|X,Y) \\ &= H(X,Y) \\ &\leq H(X) + H(Y). \end{aligned}$$

□

注 1. 两处取等条件分别为 $X+Y$ 到 (X,Y) 是单射, 以及 X,Y 相互独立. 以及此处的 $X+Y$ 可以替换为 $f(X,Y)$, 因为我们总是有 $H(X,Y) \geq H(f(X,Y))$.

第 2 题

For random variables X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n, when does

$$H(X_1, X_2, \ldots, X_n|Y_1, Y_2, \ldots, Y_n) = H(X_1|Y_1) + H(X_2|Y_2) + \cdots + H(X_n|Y_n)$$

hold?

解: 我们先定义几个命题:

1. $H(X_1, X_2, \cdots, X_n|Y_1, Y_2, \cdots, Y_n) = H(X_1|Y_1) + H(X_2|Y_2) + \cdots + H(X_n|Y_n)$;
2. $H(X_1|Y_1, Y_2, \cdots, Y_n) = H(X_1|Y_1)$ 且对每个 $i \in \{2, 3, \cdots, n\}$ 有

$$H(X_i|X_{i-1}, \cdots, X_1, Y_1, Y_2, \cdots, Y_n) = H(X_i|Y_i);$$

3. 对所有 x_1, y_1, y_2, $\cdots$, y_n, 只要

$$P_{X_1|Y_1,Y_2,\cdots,Y_n}(x_1|y_1,y_2,\cdots,y_n)=P_{X_1|Y_1}(x_1|y_1) \tag{1}$$

中条件的概率大于 0, 1 式就成立; 对所有 $i\in\{2,3,\cdots,n\}$ 和 x_1, x_2, $\cdots$, x_i, y_1, y_2, $\cdots$, y_n, 只要

$$P_{X_i|X_{i-1},\cdots,X_1,Y_1,Y_2,\cdots,Y_n}(x_i|x_{i-1},\cdots,x_1,y_1,y_2,\cdots,y_n)=P_{X_i|Y_i}(x_i|y_i) \tag{2}$$

中条件的概率大于 0, 2 式就成立;

4. 对所有 x_1, x_2, $\cdots$, x_n, y_1, y_2, $\cdots$, y_n, 如果 $P_{Y_1,Y_2,\cdots,Y_n}(y_1,y_2,\cdots,y_n)>0$ 则

$$P_{X_1,X_2,\cdots,X_n|Y_1,Y_2,\cdots,Y_n}(x_1,x_2,\cdots,x_n|y_1,y_2,\cdots,y_n)=\prod_{i=1}^{n}P_{X_i|Y_i}(x_i|y_i).$$

因为

$$\begin{aligned}&H(X_1,X_2,\cdots,X_n|Y_1,Y_2,\cdots,Y_n)\\=\;&H(X_1|Y_1,Y_2,\cdots,Y_n)+\sum_{i=2}^{n}H(X_i|X_{i-1},\cdots,X_1,Y_1,Y_2,\cdots,Y_n),\end{aligned} \tag{3}$$

$H(X_1|Y_1,Y_2,\cdots,Y_n)\le H(X_1|Y_1)$, 对每个 $i\in\{2,3,\cdots,n\}$ 有

$$H(X_i|X_{i-1},\cdots,X_1,Y_1,Y_2,\cdots,Y_n)\le H(X_i|Y_i),$$

所以命题 1 等价于命题 2. 用第 2 讲讲义的定理 6 可以证明命题 2 等价于命题 3. 用形式与 3 相同的概率的链式法则可以证明命题 3 等价于命题 4. 因此命题 1 等价于命题 4. □

第 3 题

We know from Theorem 10 that conditioning reduces entropy. For mutual information $I(X;Y)$ and conditional mutual information $I(X;Y|Z)$, does an analogous property hold?

解: 设 X 和 Y 独立, 取值于 $\{0,1\}$ 且都以 1/2 的概率取 1. 定义随机变量 $Z=X\oplus Y$. 这样 $I(X;Y)=0<1=I(X;Y|Z)$.

设随机变量 X, Z, Y 都取值于 $\{0,1\}$ 且形成一条 Markov 链, $P_X(1)=1/2$, 对所有 x, $z\in\{0,1\}$ 有

$$P_{Z|X}(z|x)=\begin{cases}0.9, & z=x\\ 0.1, & z\ne x\end{cases},$$

对所有 $z, y \in \{0,1\}$ 有

$$P_{Y|Z}(y|z) = \begin{cases} 0.9, & y = z \\ 0.1, & y \neq z \end{cases}$$

这样 $I(X;Y) > 0 = I(X;Y|Z)$. □

有同学把 $I(X;Y)$ 写成了 $I(x;y)$. 注意大写字母和小写字母的含义是不同的.

有同学给出的理由是: $I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$, 而 $I(X;Y;Z)$ 可能大于 0 也可能小于 0. 如果要用这个结论, 我们需要先证明它. 最好不要用这个结论, 因为我们没有定义过 $I(X;Y;Z)$.

第 4 题

Using the non-negativity of relative divergence, prove the log-sum inequality: for nonnegative numbers $\{a_i\}_{i=1,\dots,n}$ *and* $\{b_i\}_{i=1,\dots,n}$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b},$$

where $a = \sum_{i=1}^{n} a_i$, $b = \sum_{i=1}^{n} b_i$, *with equality holding if and only if there exists* c *such that* $a_i = cb_i$ *for all* i.

证明: 用 $p(i) = a_i/a$ 和 $q(i) = b_i/b$ 定义 $\{1, 2, \cdots, n\}$ 上的概率函数 p 和 q. 这样

$$0 \leq D(p\|q) = \sum_{i=1}^{n} \frac{a_i}{a} \log_2 \left(\frac{a_i/a}{b_i/b} \right) = \frac{1}{a} \sum_{i=1}^{n} a_i \log_2 \left(\frac{a_i}{b_i} \right) + \log_2 \left(\frac{b}{a} \right), \tag{4}$$

所以

$$\sum_{i=1}^{n} a_i \log_2 \left(\frac{a_i}{b_i} \right) \geq a \log_2 \left(\frac{a}{b} \right). \tag{5}$$

如果 5 式成立等号, 则 4 成立等号, $p = q$, 对每个正整数 $i \leq n$ 有 $a_i = (a/b)b_i$. 反之, 如果存在 c 使 $a_i = cb_i$ 对每个正整数 $i \leq n$ 成立, 则 $a = \sum_{i=1}^{n} a_i = c \sum_{i=1}^{n} b_i = cb$,

$$\sum_{i=1}^{n} a_i \log_2 \left(\frac{a_i}{b_i} \right) = \sum_{i=1}^{n} a_i \log_2(c) = a \log_2 \left(\frac{a}{b} \right).$$

□

注 2. 也可以用 $x \log x$ 的凹凸性和 *Jensen* 不等式或者 $ln(1+x) \leq x$ 证明，但此处题目要求使用相对熵的非负性质。

第 5 题

In this exercise we apply Corollary 2 to a guessing problem due to Massey [10]. Suppose that we want to guess the value of a random variable X over $X(\Omega) = \{1, 2, \dots\}$. How many times do we need to guess, on average? Without loss of generality, we can always relabel the random variable so that $P_X(1) \geq P_X(2) \geq \dots$. Prove that, on average, we need to guess no less than $e^{H(X)-1}$ times, where the unit of entropy is nat.

证明: 为了使猜的平均次数最小, 我们第 1 次应该猜 1, 第 2 次应该猜 2, 以此类推. 此时猜的平均次数为 $\sum_{x=1}^{\infty} xP_X(x) = \mathbf{E}[X]$. 用讲义推论 2 和对每个正数 t 成立的不等式 $\ln(t) \leq t - 1$,

$$\begin{aligned} H(X) &\leq \mathbf{E}[X]\ln(\mathbf{E}[X]) - (\mathbf{E}[X]-1)\ln(\mathbf{E}[X]-1) \\ &= \ln(\mathbf{E}[X]) + (\mathbf{E}[X]-1)\ln\left(\frac{\mathbf{E}[X]}{\mathbf{E}[X]-1}\right) \\ &\leq \ln(\mathbf{E}[X]) + (\mathbf{E}[X]-1)\frac{1}{\mathbf{E}[X]-1} \\ &= \ln(\mathbf{E}[X]) + 1, \end{aligned}$$

所以猜的平均次数的最小值 $\mathbf{E}[X] \geq e^{H(X)-1}$. □

第 6 题

Consider a random variable X over $X(\Omega) = \{1, 2, \dots\}$.

a) Prove that if $\mathbf{E}X$ is finite, then $H(X)$ is also finite.

b) Prove that if $\mathbf{E}\log X$ is finite, then $H(X)$ is also finite.

c) Prove that if $H(X)$ is finite and $P_X(x)$ is monotonically non-increasing with x, then $\mathbf{E}\log X$ is finite.

d) Give an example to illustrate that the monotonically non-increasing condition of $P_X(x)$ in the previous assertion is necessary.

a) 证明: 如果 $\mathbf{E}[X] = 1$, 则 X 以概率 1 取 1, $H(X) = 0$. 如果 $\mathbf{E}[X] > 1$, 则根据讲义推论 2,

$$H(X) \leq \mathbf{E}[X]\log_2(\mathbf{E}[X]) - (\mathbf{E}[X]-1)\log_2(\mathbf{E}[X]-1) < \infty. \qquad \square$$

b) 证明: 由讲义推论 6 得 $H(\log_2(X)) = H(X)$. 如果 $\mathbf{E}[\log_2(X)]$ 有限, 则 $H(\log_2(X))$ 有限, $H(X)$ 有限. □

c) 证明: 对每个正整数 x, $xP_X(x) \le \sum_{x'=1}^{x} P_X(x') \le 1$, 所以 $x \le 1/P_X(x)$. 这样

$$\mathbf{E}[\log_2(X)] = \sum_{x=1}^{\infty} P_X(x)\log_2(x) \le \sum_{x=1}^{\infty} P_X(x)\log_2\left(\frac{1}{P_X(x)}\right) = H(X) < \infty. \qquad \square$$

d) 证明: 记 $A = \sum_{i=1}^{\infty}(1/i^2)$. 设对每个正整数 i, 随机变量 X 取 2^i 的概率是 $1/Ai^2$. 这样

$$H(X) = -\sum_{i=1}^{\infty} \frac{1}{Ai^2}\log_2\left(\frac{1}{Ai^2}\right) = \sum_{i=1}^{\infty} \frac{\log_2(A) + 2\log_2(i)}{Ai^2} < \infty,$$

但

$$\mathbf{E}[\log_2(X)] = \sum_{i=1}^{\infty} \frac{1}{Ai^2}\log_2(2^i) = \sum_{i=1}^{\infty} \frac{1}{Ai} = \infty. \qquad \square$$

第 7 题

Consider a uniform random variable X over $\{0, 1, \ldots, m-1\}$, and its observation Y is drawn uniformly from $\{(X-1) \bmod m, X, (X+1) \bmod m\}$. Define $P_e = P(Y \neq X)$.

a) Give a lower bound of P_e using the Fano inequality.

b) Find the gap between the lower bound and the exact value of P_e of the MAP decision.

c) Can you resolve the gap by improving the Fano inequality?

这道题应该假设了 $m \ge 3$.

a) 解: 对每个自然数 $y < m$, 可以看出 X 在 $Y = y$ 的条件下服从 $\{(y-1) \bmod m, y, (y+1) \bmod m\}$ 上的均匀分布, 所以 $H(X|Y=y) = \log_2(3)$. 这样 $H(X|Y) = \log_2(3)$, 我们可以把 Fano 不等式写成

$$\log_2(3) \le h_2(P_e) + P_e\log_2(m-1).$$

所以 P_e 大于等于关于 p 的方程 $h_2(p) + p\log_2(m-1) - \log_2(3) = 0$ 的最小正根 r_m. $\square$

b) 解: 可以看出 $P_e = 2/3$. 图 1 对比了 P_e 和 r_m. $\square$

c) 解: 令随机变量 Z 在 $X \neq Y$ 时取 1, 否则取 0. 类似于讲义中 Fano 不等式的推导,

$$H(X|Y) = H(X|Y) + H(Z|X,Y) = H(X,Z|Y) = H(Z|Y) + H(X|Z,Y),$$
$$H(Z|Y) \le H(Z) = h_2(P_e).$$

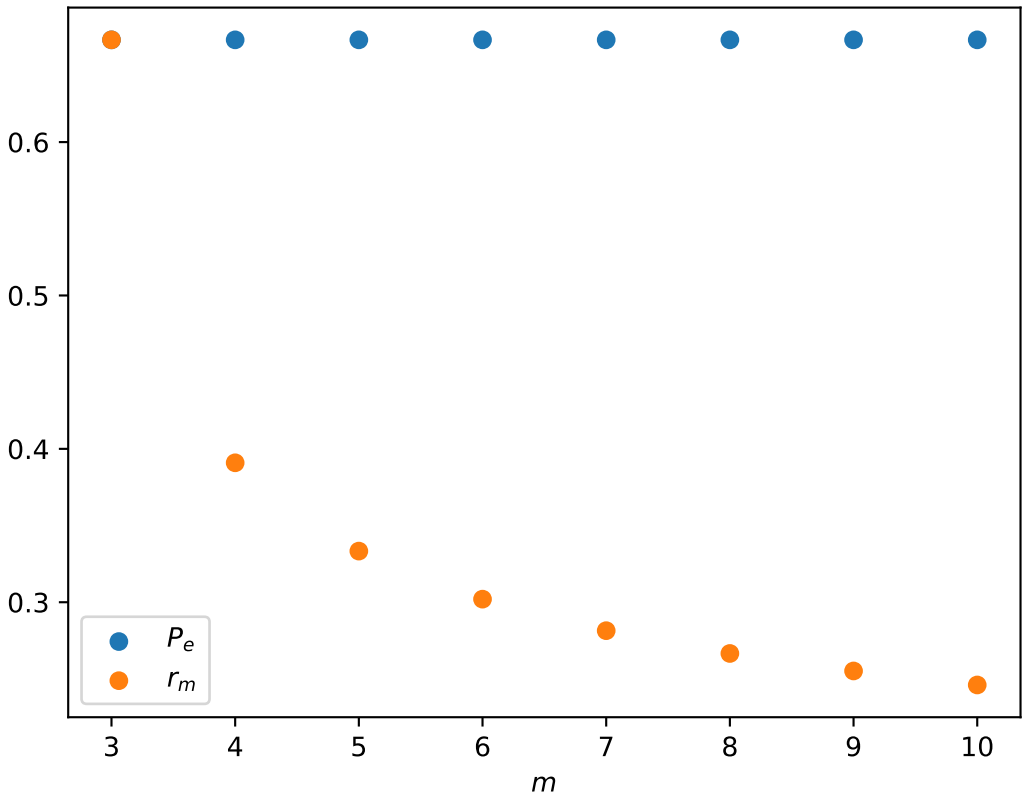

图 1: Fano 不等式给出的 P_e 的下界 r_m 和 P_e.

对使 $P_{Z,Y}(0,y) > 0$ 的每个 y, X 在 $Z = 0$ 且 $Y = y$ 的条件下以概率 1 取 y, 所以 $H(X|Z = 0, Y = y) = 0$. 对使 $P_{Z,Y}(1,y) > 0$ 的每个 y, X 在 $Z = 1$ 且 $Y = y$ 的条件下以概率 1 属于 $\{(y-1) \bmod m, (y+1) \bmod m\}$, 所以 $H(X|Z = 1, Y = y) \le \log_2(2) = 1$. 这样

$$H(X|Z,Y) \le \sum_{y,P_{Z,Y}(1,y)>0} P_{Z,Y}(1,y) \cdot 1 = P_Z(1) = P_e,$$

$$H(X|Y) \le h_2(P_e) + P_e. \tag{6}$$

如果 $m = 3$ 则 6 式就是 Fano 不等式. 如果 $m > 3$ 则 6 式比 Fano 不等式紧. 由 6 式得 P_e 大于等于关于 p 的方程 $h_2(p) + p - \log_2(3) = 0$ 的最小正根 2/3. □

第 8 题

Construct an example where equality holds in the Fano equality.

解: 我们仔细考察 Fano 不等式的推导过程以及取等条件，如下：

$$\begin{aligned} H(X|\hat{X}) + H(E|X,\hat{X}) &= H(E|\hat{X}) + H(X|\hat{X},E) \\ H(X|\hat{X}) + 0 &\leq h_2\left(P_e\right) + (1-P_e)\cdot 0 + P_e \cdot \log(|X(\Omega)|-1) \\ H(X|\hat{X}) &\leq h_2\left(P_e\right) + P_e \log(|X(\Omega)|-1) \end{aligned}$$

从中可以看出，两处放缩分别使用

$$H(E|\hat{X}) \leq H(E) = h_2\left(P_e\right), \qquad H(X|\hat{X},E=1) \leq \log(|X(\Omega)|-1).$$

换言之，即:

- $\hat{X}$ 的具体值对译码错误率没有影响。
- 译码错误时，不论译为任何值，X 的分布总是均匀的。

基于此，我们构造如下两个例子：

例子一：设 $X \in \{1,2,3,...m\}$ 并且 $p_1 \geq p_2 \geq ... \geq p_m$，我们对 X 的最佳估计是 $\hat{X}=1$，此时产生的误差概率为 $p_e = 1-p_1$。此时 Fano 不等式变为：

$$h_2(p_e) + p_e \log(m-1) \geq H(X)$$

并且概率密度函数

$$(p_1,p_2,...,p_m) = \left(1-p_e, \frac{p_e}{m-1}, ..., \frac{p_e}{m-1}\right),$$

此时等号成立，Fano 不等式是精确的。

例子二：设 $X \in \{1,2,3,...m\}$ 并且 $p_1 = p_2 = ... = p_m = \frac{1}{m}$，然后我们构造概率转移矩阵如下：

$$p_{Y|X}(y_j|x_i) = \begin{cases} \frac{p_e}{m-1} & \text{if} i=j \\ p_e & \text{if} i \neq j \end{cases}.$$

并且在译码端保证 $\hat{X}=Y$。此时根据 $p_{X|\hat{X}} = \frac{p_X p_{\hat{X}|X}}{p_{\hat{X}}} = p_{Y|X}$ 有

$$\begin{aligned} H(X|\hat{X}) &= -\frac{p_e}{m-1}\log\frac{p_e}{m-1}(m-1) + (1-p_e)\log\frac{1}{1-p_e} \\ &= p_e \log\frac{1}{p_e} + (1-p_e)\log\frac{1}{1-p_e} + p_e\log(m-1) \\ &= h_2(p_e) + p_e\log(m-1). \end{aligned}$$

□

第 9 题

If the estimate $\hat{X}$ is a size-L subset of $X(\Omega)$, and define the error event to be $\{X \notin \hat{X}\}$, establish an extension of the Fano inequality.

解: 设 X 是离散随机变量, 正整数 $L < |X(\Omega)|$, $\hat{X}$ 是 $X(\Omega)$ 的一个随机的子集, $|X(\Omega)| = L$ 以概率 1 成立. 用 P_e 表示 $X \notin \hat{X}$ 的概率. 我们来证明

$$H(X|\hat{X}) \leq h_2(P_e) + (1 - P_e)\log_2(L) + P_e \log_2(|X(\Omega)| - L). \tag{7}$$

令随机变量 Z 在 $X \notin \hat{X}$ 时取 1, 否则取 0. 类似于讲义中 Fano 不等式的推导,

$$H(X|\hat{X}) = H(X|\hat{X}) + H(Z|X, \hat{X}) = H(X, Z|\hat{X}) = H(Z|\hat{X}) + H(X|Z, \hat{X}),$$
$$H(Z|\hat{X}) \leq H(Z) = h_2(P_e).$$

对使 $P_{Z,\hat{X}}(0, \hat{x}) > 0$ 的每个 $\hat{x}$, X 在 $Z = 0$ 且 $\hat{X} = \hat{x}$ 的条件下以概率 1 属于有 L 个元素的集合 $\hat{x}$, 所以

$$H(X|Z = 0, \hat{X} = \hat{x}) \leq \log_2(L).$$

对使 $P_{Z,\hat{X}}(1, \hat{x}) > 0$ 的每个 $\hat{x}$, X 在 $Z = 1$ 且 $\hat{X} = \hat{x}$ 的条件下以概率 1 属于有 $|X(\Omega)| - L$ 个元素的集合 $X(\Omega) \setminus \hat{x}$, 所以

$$H(X|Z = 1, \hat{X} = \hat{x}) \leq \log_2(|X(\Omega)| - L).$$

这样

$$\begin{aligned} H(X|Z, \hat{X}) &\leq \sum_{\hat{x}, P_{Z,\hat{X}}(0,\hat{x})>0} P_{Z,\hat{X}}(0, \hat{x}) \log_2(L) + \sum_{\hat{x}, P_{Z,\hat{X}}(1,\hat{x})>0} P_{Z,\hat{X}}(1, \hat{x}) \log_2(|X(\Omega) - L|) \\ &= P_Z(0) \log_2(L) + P_Z(1) \log_2(|X(\Omega)| - L) \\ &= (1 - P_e) \log_2(L) + P_e \log_2(|X(\Omega)| - L), \end{aligned}$$

7 式成立. □

第 10 题

Prove the Csiszár identity:

$$\sum_{i=1}^{n} I(X_{i+1}, \ldots, X_n; Y_i|Y_1, \ldots, Y_{i-1}) = \sum_{i=1}^{n} I(Y_1, \ldots, Y_{i-1}; X_i|X_{i+1}, \ldots, X_n),$$

where X_{n+1} and Y_0 are degenerated.

这样说可能更好理解: 如果 $X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_n$ 是离散随机变量, X_{n+1} 和 Y_0 是常数, 则

$$\sum_{i=1}^{n} I(X_{i+1}, \cdots, X_{n+1}; Y_i | Y_0, \cdots, Y_{i-1}) = \sum_{i=1}^{n} I(Y_0, \cdots, Y_{i-1}; X_i | X_{i+1}, \cdots, X_{n+1}). \quad (8)$$

方法一:

$$\begin{aligned}
&\sum_{i=1}^{n} I(X_{i+1}, \cdots, X_{n+1}; Y_i | Y_0, \cdots, Y_{i-1}) \\
\overset{\text{(a)}}{=} &\sum_{i=1}^{n-1} I(X_{i+1}, \cdots, X_n; Y_i | Y_0, \cdots, Y_{i-1}, X_{n+1}) \\
\overset{\text{(b)}}{=} &\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(X_j; Y_i | Y_0, \cdots, Y_{i-1}, X_{j+1}, \cdots, X_{n+1}) \\
\overset{\text{(c)}}{=} &\sum_{j=2}^{n} \sum_{i=1}^{j-1} I(X_j; Y_i | Y_0, \cdots, Y_{i-1}, X_{j+1}, \cdots, X_{n+1}) \\
\overset{\text{(d)}}{=} &\sum_{j=2}^{n} I(X_j; Y_1, \cdots, Y_{j-1} | Y_0, X_{j+1}, \cdots, X_{n+1}) \\
\overset{\text{(e)}}{=} &\sum_{j=1}^{n} I(X_j; Y_0, \cdots, Y_{j-1} | X_{j+1}, \cdots, X_{n+1}).
\end{aligned}$$

上式第一行 $i = n$ 的一项等于 0, 对每个正整数 $i < n$ 有

$$I(X_{i+1}, \cdots, X_{n+1}; Y_i | Y_0, \cdots, Y_{i-1}) = I(X_{i+1}, \cdots, X_n; Y_i | Y_0, \cdots, Y_{i-1}, X_{n+1}),$$

所以 (a) 成立. 同理可得 (e). (b) 和 (d) 用了互信息的链式法则. 通过交换求和顺序可以得到 (c). □

方法二: 对每个正整数 $i \leq n-1$, 我们可以以两种方式展开 $I(X_{i+1}, \cdots, X_{n+1}; Y_0, \cdots, Y_i)$ 得到

$$\begin{aligned}
&I(X_{i+1}, \cdots, X_{n+1}; Y_0, \cdots, Y_{i-1}) + I(X_{i+1}, \cdots, X_{n+1}; Y_i | Y_0, \cdots, Y_{i-1}) \\
= &I(X_{i+2}, \cdots, X_{n+1}; Y_0, \cdots, Y_i) + I(X_{i+1}; Y_0, \cdots, Y_i | X_{i+2}, \cdots, X_{n+1}). \quad (9)
\end{aligned}$$

因为 $i \in \{1, n\}$ 时 $I(X_{i+1}, \cdots, X_{n+1}; Y_0, \cdots, Y_{i-1}) = 0$, 所以

$$\begin{aligned}
\sum_{i=1}^{n-1} I(X_{i+1}, \cdots, X_{n+1}; Y_0, \cdots, Y_{i-1}) &= \sum_{i=2}^{n} I(X_{i+1}, \cdots, X_{n+1}; Y_0, \cdots, Y_{i-1}) \\
&= \sum_{i=1}^{n-1} I(X_{i+2}, \cdots, X_{n+1}; Y_0, \cdots, Y_i).
\end{aligned}$$

求 9 式对所有正整数 $i \leq n-1$ 的和得

$$\begin{aligned}\sum_{i=1}^{n-1} I(X_{i+1}, \cdots, X_{n+1}; Y_i | Y_0, \cdots, Y_{i-1}) &= \sum_{i=1}^{n-1} I(X_{i+1}; Y_0, \cdots, Y_i | X_{i+2}, \cdots, X_{n+1}) \\ &= \sum_{i=2}^{n} I(X_i; Y_0, \cdots, Y_{i-1} | X_{i+1}, \cdots, X_{n+1})\end{aligned}$$

即 8 式. □