信息论 3 月 17 号第一次小测解答

中国科学技术大学《信息论 A》006125.01 班助教组

2025 年 3 月 26 日

第 1 题

In this exercise, we provide an information-theoretic proof of the well known number-theoretic result that there are infinitely many prime numbers. For this, consider an arbitrary integer n*, and denote the number of primes no greater than* n *by* $\pi(n)$*. Take a random variable* N *uniformly distributed over* $\{1, 2, \ldots, n\}$*, and write it in its unique prime factorization,* $N = p_1^{X_1} p_2^{X_2} \ldots p_{\pi(n)}^{X_{\pi(n)}}$*, where* $\{p_1, p_2, \cdots, p_{\pi(n)}\}$ *are primes no greater than* n*, and each* X_i *is the largest power* $k \geq 0$ *such that* p_i^k *divides* N*. By inspecting* $H(N)$*, prove that* $\pi(n) \to \infty$ *as* $n \to \infty$*. For further reading, refer to [14].*

证明: 由素数分解的唯一性得知，一个 N 的抽样结果可以和一组 $x_1, x_2, ..., x_{\pi(n)}$ 形成一一对应，从而依次根据均匀分布熵的表达、链式法则和条件减少熵的性质有

$$\log_2(n) = H(N) = H(X_1, X_2, \cdots, X_{\pi(n)}) \leq \sum_{i=1}^{\pi(n)} H(X_i).$$

对每个正整数 $i \leq \pi(n)$, 均有 $2^{X_i} \leq p_i^{X_i} \leq N \leq n$, 从而 $0 \leq X_i \leq \lfloor \log_2(n) \rfloor$. 这样对每个正整数 $i \leq \pi(n)$, X_i 字母表大小为 $\lfloor \log_2(n) \rfloor + 1$, 所以有 $H(X_i) \leq \log_2(\lfloor \log_2(n) \rfloor + 1) \leq \log_2(\log_2(n) + 1)$, 即有

$$\log_2(n) \leq \pi(n) \log_2(\log_2(n) + 1).$$

$n \to \infty$ 时, 因为 $\log_2(n)/\log_2(\log_2(n) + 1) \to \infty$, 所以 $\pi(n) \to \infty$. □

第 2 题

Prove the submodularity property of entropy: for any two sets of random variables $\mathbf{S}_1$ *and* $\mathbf{S}_2$*,* $H(\mathbf{S}_1 \cup \mathbf{S}_2) + H(\mathbf{S}_1 \cap \mathbf{S}_2) \leq H(\mathbf{S}_1) + H(\mathbf{S}_2)$*.*

证明: 记 $X = \mathbf{S}_1 \setminus \mathbf{S}_2$, $Y = \mathbf{S}_1 \cap \mathbf{S}_2$, $Z = \mathbf{S}_2 \setminus \mathbf{S}_1$. 这样 $\mathbf{S}_1 = (X, Y)$, $\mathbf{S}_2 = (Y, Z)$, $\mathbf{S}_1 \cup \mathbf{S}_2 = (X, Y, Z)$. 因为

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

所以 $H(\mathbf{S}_1 \cup \mathbf{S}_2) + H(\mathbf{S}_1 \cap \mathbf{S}_2) \le H(\mathbf{S}_1) + H(\mathbf{S}_2)$. □

第 3 题

For random variables X and Y and a mapping f, under what condition does $H(X|f(Y)) = H(X|Y)$ hold?

解: 因为 $I(X; f(Y)) = H(X) - H(X|f(Y))$, $I(X;Y) = H(X) - H(X|Y)$, 所以

$$H(X|f(Y)) = H(X|Y) \tag{1}$$

当且仅当

$$I(X; f(Y)) = I(X;Y). \tag{2}$$

因为 $X \leftrightarrow Y \leftrightarrow f(Y)$, 根据讲义中 Theorem 3.5, 2 式成立当且仅当 $X \leftrightarrow f(Y) \leftrightarrow Y$. 因此 1 式成立当且仅当 $X \leftrightarrow f(Y) \leftrightarrow Y$. □

第 4 题

For the two-state Markov chain in Example 3.5, if we undersample it to obtain a new stochastic process X_1, X_3, X_5, ..., is it still a Markov chain? Under stationarity, evaluate its entropy rate and compare with that of the original Markov chain X_1, X_2, X_3, ...

解: 设 n 是正整数. 定义随机变量 $Y = (X_1, X_3, \cdots, X_{2n-1})$. 如果 $P_{X_{2n+2}, X_{2n+1}, Y}(x_2, x_1, y) > 0$ 且 $x_3 \in \{0, 1\}$ 则

$$\begin{aligned} & P_{X_{2n+3}, X_{2n+2} | X_{2n+1}, Y}(x_3, x_2 | x_1, y) \\ = & P_{X_{2n+2} | X_{2n+1}, Y}(x_2 | x_1, y) P_{X_{2n+3} | X_{2n+2}, X_{2n+1}, Y}(x_3 | x_2, x_1, y) \\ = & P_{X_{2n+2} | X_{2n+1}}(x_2 | x_1) P_{X_{2n+3} | X_{2n+2}, X_{2n+1}}(x_3 | x_2, x_1) \\ = & P_{X_{2n+3}, X_{2n+2} | X_{2n+1}}(x_3, x_2 | x_1). \end{aligned}$$

等式两边对 x_2 求和得 $P_{X_{2n+3}|X_{2n+1},Y}(x_3|x_1,y) = P_{X_{2n+3}|X_{2n+1}}(x_3|x_1)$. 因此 X_1, X_3, X_5, $\cdots$ 是 Markov 链.

根据平稳 Markov 链的熵率的定义, Markov 链 X_1, X_2, $X_3,\dots$ 和 X_1, X_3, X_5, $\cdots$ 的熵率分别为 $H(X_3|X_2)$ 和 $H(X_3|X_1)$. 依据数据处理不等式，我们有

$$\begin{aligned} I(X_2;X_3) &\geq I(X_1;X_3) \\ H(X_3)-H(X_3|X_2) &\geq H(X_3)-H(X_3|X_1) \end{aligned}$$

即 $H(X_3|X_2) \leq H(X_3|X_1)$，说明 Markov 链 X_1, X_2, X_3, $\cdots$ 的熵率小于等于 Markov 链 X_1, X_3, X_5, $\cdots$ 的熵率.

我们可以通过以下方法进一步计算 Markov 链 X_1, X_3, X_5, $\cdots$ 的熵率. 用 Q 表示 Markov 链 X_1, X_2, X_3, $\cdots$ 的一步转移概率矩阵

$$\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

用 π 表示它的平稳分布. X_1, X_3, X_5, $\cdots$ 的一步转移概率矩阵等于

$$Q^2 = \begin{bmatrix} 1-2\alpha+\alpha^2+\alpha\beta & 2\alpha-\alpha^2-\alpha\beta \\ 2\beta-\alpha\beta-\beta^2 & 1-2\beta+\alpha\beta+\beta^2 \end{bmatrix}.$$

因为 $[\pi(0),\pi(1)]Q = [\pi(0),\pi(1)]$, 所以 $[\pi(0),\pi(1)]Q^2 = [\pi(0),\pi(1)]$, π 也是 X_1, X_3, X_5, $\cdots$ 的平稳分布. 由于我们假设了 X_1, X_3, X_5, $\cdots$ 是平稳的, X_1 服从 π. 这样 X_1, X_3, X_5, $\cdots$ 的熵率等于

$$\begin{aligned} H(X_3|X_1) &= \pi(0)H(X_3|X_1=0)+\pi(1)H(X_3|X_1=1) \\ &= \frac{\beta}{\alpha+\beta}h_2(2\alpha-\alpha^2-\alpha\beta)+\frac{\alpha}{\alpha+\beta}h_2(2\beta-\alpha\beta-\beta^2). \end{aligned}$$

□

我们也可以对每个正整数 n 证明 $I(X_1,X_3,\cdots,X_{2n-1};X_{2n+3}|X_{2n+1}) = 0$, 从而证明 X_1, X_3, X_5, $\cdots$ 是一条 Markov 链.

$$\begin{aligned} I(Y;X_{2n+2},X_{2n+3}|X_{2n+1}) &= I(Y;X_{2n+2}|X_{2n+1})+I(Y;X_{2n+3}|X_{2n+1},X_{2n+2}) \\ &= I(Y;X_{2n+3}|X_{2n+1})+I(Y;X_{2n+2}|X_{2n+1},X_{2n+3}) \end{aligned}$$

又因为 $Y \leftrightarrow X_{2n+1} \leftrightarrow X_{2n+2}$ 和 $Y \leftrightarrow X_{2n+2} \leftrightarrow X_{2n+3}$, 所以我们有 $I(Y;X_{2n+2}|X_{2n+1}) = 0$, $I(Y;X_{2n+3}|X_{2n+1},X_{2n+2}) = 0$ 和 $I(Y;X_{2n+2}|X_{2n+1},X_{2n+3}) = 0$, 从而可得 $I(Y;X_{2n+3}|X_{2n+1}) = 0$, 即 $X_1,\cdots,X_{2n-1} \leftrightarrow X_{2n+1} \leftrightarrow X_{2n+2}$ 成立.

用类似的方法可以证明如果正整数 $k_1 \leq n_1 < k_2 \leq n_2 < \cdots$ 则

$$\{(X_{k_j},X_{k_j+1},\cdots,X_{n_j})\}_{j=1}^{\infty}$$

是一条 Markov 链. 见 [1] 推论 3.10.

第 5 题

Define an "almost Markov" relationship for three random variables (X, Y, Z) if they satisfy

$$p(z|x,y) = p(z|y)(1+\epsilon(x,y,z)),$$

where $|\epsilon(x,y,z)| \le \delta$ for any (x,y,z) tuple. Prove that for such an "almost Markov" relationship, we have the following "δ-approximate DPI" hold:

$$I(X;Z) \le I(X;Y) + \delta^2.$$

这道题中互信息的底应该是 e.

证明: 类似于数据处理不等式的推导,

$$I(X;Z) \le I(X;Z) + I(X;Y|Z) = I(X;Y,Z) = I(X;Y) + I(X;Z|Y). \tag{3}$$

根据条件互信息的定义，我们有：

$$\begin{aligned} I(X;Z|Y) &= \sum_{x,y,z} P_{X,Y,Z}(x,y,z) \ln \frac{P_{X,Z|Y}(x,z|y)}{P_{X|Y}(x|y)P_{Z|Y}(z|y)} \\ &= \sum_{x,y,z} P_{X,Y,Z}(x,y,z) \ln \frac{P_{Z|X,Y}(z|x,y)}{P_{Z|Y}(z|y)} \\ &= \sum_{x,y,z} P_{X,Y,Z}(x,y,z) \ln(1+\epsilon(x,y,z)) \end{aligned} \tag{4}$$

在开始后续分析之前，我们可以得到以下事实：

$$\begin{aligned} \sum_{x,y,z} P_{X,Y,Z}(x,y,z) &= 1 \\ \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|X,Y}(z|x,y) &= 1 \\ \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)(1+\epsilon(x,y,z)) &= 1 \\ \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y) + \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)\epsilon(x,y,z) &= 1 \end{aligned}$$

又 $\sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y) = \sum_{x,y} P_{X,Y}(x,y) \sum_z P_{Z|Y}(z|y) = 1$，所以有

$$\sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)\epsilon(x,y,z) = 0 \tag{5}$$

接着从 4 式出发，我们有

$$
\begin{aligned}
I(X;Z|Y) &= \sum_{x,y,z} P_{X,Y,Z}(x,y,z)\ln(1+\epsilon(x,y,z)) \\
&\le \sum_{x,y,z} P_{X,Y,Z}(x,y,z)\epsilon(x,y,z) \\
&= \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)(1+\epsilon(x,y,z))\epsilon(x,y,z) \\
&= \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)\epsilon(x,y,z) + \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)\epsilon^2(x,y,z) \\
&\le \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y)\epsilon(x,y,z) + \delta^2 \sum_{x,y,z} P_{X,Y}(x,y)P_{Z|Y}(z|y) \\
&= \delta^2
\end{aligned} \tag{6}
$$

其中第一个不等式是因为 $\ln(1+x)\le x$，最后一个等号基于 5 式的结果。综合 3 式和 6 式，最终证得 $I(X;Z)\le I(X;Y)+\delta^2$. □

参考文献

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.