信息论第十讲作业解答

中国科学技术大学《信息论 A》006125.01 班助教组

2024 年 6 月 13 日

第 1题

Consider a sequence of i.i.d. random variables $\{X_i, i = 1, 2, \ldots\}$*, and the sequence of sample means* $\{S_n, n = 1, 2, \ldots\}, S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

a) If X_i *is discrete with entropy* $H(X)$*, calculate* $\frac{1}{n}H(S_1, S_2, \ldots, S_n)$.

b) If X_i *is continuous with differential entropy* $h(X)$*, calculate* $\frac{1}{n}h(S_1, S_2, \ldots, S_n)$.

解答：

a) 对于$H(S_i|S_1, \cdots, S_{i-1})$而言，满足

$$\begin{aligned} H(S_i|S_1, \cdots, S_{i-1}) &= H\left(\frac{X_1 + \cdots + X_{i-1}}{i} + \frac{X_i}{i}|S_1, \cdots, S_{i-1}\right) \\ &= H\left(\frac{i-1}{i}S_{i-1} + \frac{X_i}{i}|S_1, \cdots, S_{i-1}\right) \\ &= H\left(\frac{X_i}{i}|S_1, \cdots, S_{i-1}\right) \\ &= H(X_i). \end{aligned}$$

因此有

$$\begin{aligned} \frac{1}{n}H(S_1, S_2, \cdots, S_n) &= \frac{1}{n}\Big[H(S_1) + H(S_2|S_1) + \cdots + H(S_n|S_1, \cdots, S_{n-1})\Big] \\ &= \frac{1}{n}\sum_{i=1}^{n} H(X_i) \\ &= H(X). \end{aligned}$$

b) 我们按照相同的步骤处理条件熵$h(S_i|S_1, \cdots, S_{i-1})$，即

$$h(S_i|S_1, \cdots, S_{i-1}) = h\left(\frac{i-1}{i}S_{i-1} + \frac{X_i}{i}|S_1, \cdots, S_{i-1}\right)$$

$$= h\left(\frac{X_i}{i}|S_1,\cdots,S_{i-1}\right) \tag{1}$$

$$= h(X_i) - \log i, \tag{2}$$

其中(1)与(2)分别来源于Proposition 3中的$h(X+\underline{b}) = h(X)$与$h(\mathrm{A}\underline{X}) = h(\underline{X})+\log|\mathrm{A}|$。因此，

$$\begin{aligned}\frac{1}{n}h(S_1,S_2,\cdots,S_n) &= \frac{1}{n}\Big[h(S_1)+h(S_2|S_1)+\cdots+h(S_n|S_1,\cdots,S_{n-1})\Big]\\ &= \frac{1}{n}\sum_{i=1}^{n} h(X_i) - \log i\\ &= h(X) - \frac{1}{n}\log n!.\end{aligned}$$

第 2题

For independent continuous random variables X and Y, prove that $h(X+Y) \geq h(X)$.

解答：

$$h(X) = h(X|Y) = h(X+Y|Y) \leq h(X+Y),$$

当且仅当$X+Y$与Y独立，即Y为一个常数时，不等式取等。

第 3题

Consider a k-dimensional continuous random vector $\underline{X}$.

a) If $\underline{X}$ has zero mean, and has covariance matrix $\mathbf{K}$, what is the maximum differential entropy of $\underline{X}$?

b) Prove Hadamard's inequality, $|\mathbf{K}| \leq \prod_{i=1}^{k}\mathbf{K}_{ii}$.

c) Prove that the log-determinant $\ln|\mathbf{K}|$ is concave with respect to $\mathbf{K}$.

解答：

a) 证明高斯分布$f^*_{\underline{X}}(\underline{x})$为$\underline{X}$取最大熵时的分布，即$\underline{X}\sim\mathcal{N}(\mathbf{0},\mathbf{K})$。令 $f_{\underline{X}}(\underline{x})$为满足约束条件$\int f_{\underline{X}}(\underline{x})\,x_ix_j\mathrm{d}\underline{x} = K_{ij}$的任意密度函数，$f^*_{\underline{X}}(\underline{x})$为$\mathcal{N}(\mathbf{0},\mathbf{K})$的密度函数。注意$\log f^*_{\underline{X}}(\underline{x})$为

二次型且满足$\int f^*_{\underline{X}}\left(\underline{x}\right) x_i x_j \mathrm{d}\underline{x} = K_{ij}$。因此有

$$\begin{aligned} D(f_{\underline{X}} \| f^*_{\underline{X}}) &= \int_{\mathcal{S}} f_{\underline{X}}(\underline{x}) \log \frac{f_{\underline{X}}(\underline{x})}{f^*_{\underline{X}}(\underline{x})} \mathrm{d}\underline{x} \\ &= -h(f_{\underline{X}}) - \int_{\mathcal{S}} f_{\underline{X}}(x) \log f^*_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} \\ &\overset{(a)}{=} -h(f_{\underline{X}}) - \int_{\mathcal{S}} f^*_{\underline{X}}(x) \log f^*_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} \\ &= -h(f_{\underline{X}}) + h(f^*_{\underline{X}}) \\ &\geq 0, \end{aligned}$$

其中等号(a)是因为二次型$\log f^*_{\underline{X}}\left(\underline{x}\right)$在$f^*_{\underline{X}}$与$f_{\underline{X}}$中的矩相同。

b) 即证协方差矩阵满足哈达玛不等式。令$\underline{X} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}\right)$，$\underline{X} = (X_1, X_2, \cdots .X_k)$。有高斯分布差分熵可得

$$\begin{aligned} \frac{1}{2} \ln\left(2\pi e\right)^n |K| &= h(X_1, X_2, \cdots, X_n) \\ &\leq \sum_{i=1}^n h(X_i) = \sum_{i=1}^n \frac{1}{2} \ln 2\pi e |K_{ii}|, \end{aligned}$$

因此有$|\mathbf{K}| \leq \prod_{i=1}^k \mathbf{K}_{ii}$成立。当$X_1, X_2, \cdots .X_k$相互独立时取等，即$\mathbf{K}_{ij} = 0, i \neq j$。

c) 设 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 是 n 维零均值正态分布向量，$\mathbf{X}_i \sim \phi_{K_i}(\mathbf{x})$，$i = 1, 2$。随机变量 θ 的分布为 $\Pr\{\theta = 1\} = \lambda, \Pr\{\theta = 2\} = 1 - \lambda, \quad 0 \leq \lambda \leq 1$，且 θ、$\mathbf{X}_1$ 和 $\mathbf{X}_2$ 相互独立。令 $\mathbf{Z} = \mathbf{X}_\theta$，则 $\mathbf{Z}$ 的协方差为 $K_Z = \lambda K_1 + (1-\lambda) K_2$。虽然 $\mathbf{Z}$ 不一定服从多元正态分布，但是由于正态分布对于给定方差具有最大熵，因此我们有

$$\begin{aligned} \frac{1}{2} \ln(2\pi e)^n \left|\lambda K_1 + (1-\lambda) K_2\right| &\geq h(\mathbf{Z}) \\ &\geq h(\mathbf{Z} \mid \theta) \\ &= \lambda \frac{1}{2} \ln(2\pi e)^n \left|K_1\right| + (1-\lambda) \frac{1}{2} \ln(2\pi e)^n \left|K_2\right|, \end{aligned}$$

因此，

$$\left|\lambda K_1 + (1-\lambda) K_2\right| \geq \left|K_1\right|^\lambda \left|K_2\right|^{1-\lambda}$$

对不等式两边取对数后可得待证结论。

第 4题

Prove the following generalization of the maximum entropy principle: for any given probability density function $g_X(x), x \in \mathcal{S}$, the probability density function $f_X(x)$ that minimizes

$D\left(f_X \| g_X\right)$ and satisfying $\int_{\mathcal{S}} f_X(x) r_i(x) \mathrm{d}x = \alpha_i, i = 1, 2, \ldots, m$, is given by the following form:

$$f_X(x) = g_X(x) e^{\lambda_0 + \sum_{i=1}^{m} \lambda_i r_i(x)}$$

where $\{\lambda_i\}_{i=0,1,\ldots,m}$ are parameters.

证明：令$f_X^* = g_X(x) e^{\lambda_0 + \sum_{i=1}^{m} \lambda_i r_i(x)}$表示最小化相对熵的分布，$f_X$表示其他满足约束条件的分布. 则

$$\begin{aligned}
D(f_X \| g_X) - D\left(f_X^* \| g_X\right) &= \int_{\mathcal{S}} f_X \ln \frac{f_X}{g_X} \mathrm{d}x - \int_{\mathcal{S}} f_X^* \ln \frac{f_X^*}{g_X} \mathrm{d}x \\
&= \int_{\mathcal{S}} f_X \ln \frac{f_X}{g_X} \mathrm{d}x - \int_{\mathcal{S}} f_X^* \left[\lambda_0 + \sum_i \lambda_i r_i(x)\right] \mathrm{d}x \\
&= \int_{\mathcal{S}} f_X \ln \frac{f_X}{g_X} \mathrm{d}x - \int_{\mathcal{S}} f_X \left[\lambda_0 + \sum_i \lambda_i r_i(x)\right] \mathrm{d}x \\
&= \int_{\mathcal{S}} f_X \ln \frac{f_X}{g_X} \mathrm{d}x - \int_{\mathcal{S}} f_X \ln \frac{f_X^*}{g_X} \mathrm{d}x \\
&= \int_{\mathcal{S}} f_X \ln \frac{f_X}{f_X^*} \mathrm{d}x \\
&= D\left(f_X \| f_X^*\right) \\
&\geq 0,
\end{aligned}$$

因此f_X^*最小化相对熵$D\left(f_X \| g_X\right)$。

第 5题

Prove that the entropy power inequality can be rewritten in the following equivalent form: for independent continuous random vectors $\underline{X}$ and $\underline{Y}$ over $\mathbb{R}^k$, define $\tilde{\underline{X}}$ and $\tilde{\underline{Y}}$ to be independent Gaussian vectors satisfying $h(\tilde{\underline{X}}) = h(\underline{X})$ and $h(\tilde{\underline{Y}}) = h(\underline{Y})$, then it holds that

$$h(\underline{X} + \underline{Y}) \geq h(\tilde{\underline{X}} + \tilde{\underline{Y}}).$$

证明：

由熵幂的定义$N(\underline{X}) = \frac{1}{2\pi e} e^{\frac{2}{k} h(\underline{X})}$可知，只需利用熵幂不等式证明$N(\underline{X} + \underline{Y}) \geq N(\tilde{\underline{X}} + \tilde{\underline{Y}})$即可说明等价性，即

$$N(\tilde{\underline{X}} + \tilde{\underline{Y}}) \overset{(a)}{=} N(\tilde{\underline{X}}) + N(\tilde{\underline{Y}}) \overset{(b)}{=} N(\underline{X}) + N(\underline{Y}) \leq N(\underline{X} + \underline{Y}),$$

其中(a)来源于熵幂不等式取等条件，(b)来源于差分熵相等，不等号来源于熵幂不等式。

注 1. 题目应该增加条件“*$\tilde{\underline{X}}$ and $\tilde{\underline{Y}}$ have proportional covariance matrices*”。