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

2025 年 5 月 14 日

第 1 题

Calculate the differential entropy of the following probability density functions.

a) *Cauchy:* $f(x) = \frac{\lambda}{\pi}\frac{1}{\lambda^2+x^2}, \mathcal{S} = \mathbb{R}$.
b) *Laplace:* $f(x) = \frac{1}{2\lambda}e^{-|x-\theta|/\lambda}, \mathcal{S} = \mathbb{R}$.
c) *Rayleigh:* $f(x) = \alpha x e^{-\alpha x^2/2}, \mathcal{S} = [0, \infty)$.

解: a) 假设随机变量 V 服从标准 Cauchy 分布，即其概率密度函数为 $f_V(v) = \frac{1}{\pi(1+v^2)}$. 所以有 $X = \lambda V, \lambda > 0$, 即有 $h(X) = h(V) + \log\lambda$. 接下来计算 $h(V)$

$$\begin{aligned} h(V) &= -\int_{-\infty}^{\infty} f_V(v)\log f_V(v)dv \\ &= \int_{-\infty}^{\infty}\frac{1}{\pi(1+v^2)}\log\big(\pi(1+v^2)\big)dv \\ &= \log\pi + \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\log(1+v^2)}{1+v^2}dv, \end{aligned} \tag{1}$$

令 $v = \tan\theta, \theta \in (\frac{\pi}{2}, \frac{\pi}{2})$, 则 $dv = \sec^2\theta d\theta$, $1 + v^2 = 1 + \tan^2\theta = \sec^2\theta$, 代入上式便

$$\begin{aligned} \int_{-\infty}^{\infty}\frac{\log(1+v^2)}{1+v^2}dv &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\log(\sec^2\theta)d\theta \\ &= -2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\log(\cos\theta)d\theta \\ &= -4\int_0^{\frac{\pi}{2}}\log(\cos\theta)d\theta, \end{aligned} \tag{2}$$

记 $I = \int_0^{\frac{\pi}{2}}\log(\cos\theta)d\theta$, 可以验证

$$\int_0^{\frac{\pi}{2}}\log(\sin\theta)d\theta = \int_0^{\frac{\pi}{2}}\log(\sin 2\theta)d\theta = I. \tag{3}$$

由此可得

$$
\begin{aligned}
2I &= \int_0^{\frac{\pi}{2}} \log(\cos\theta)d\theta + \int_0^{\frac{\pi}{2}} \log(\sin\theta)d\theta \\
&= \int_0^{\frac{\pi}{2}} \log(\frac{\sin 2\theta}{2})d\theta \\
&= \int_0^{\frac{\pi}{2}} \log(\sin 2\theta)d\theta - \int_0^{\frac{\pi}{2}} \log(2)d\theta \\
&= I - \frac{\pi}{2}\log 2, \qquad (4)
\end{aligned}
$$

可得 $I = -\frac{\pi}{2}\log(2)$, 代入 (1) 式可得 $h(V) = \log(4\pi)$, 即可得 $h(X) = \log(4\pi\lambda)$.

b) 假设随机变量 V 的概率密度函数为 $\frac{1}{2\lambda}e^{-|v|/\lambda}$, 所以有 $X = V + \theta$, $h(X) = h(V)$. 假设对数底为 e, 计算 $h(V)$

$$
\begin{aligned}
h(V) &= -\int_{-\infty}^{\infty} f_V(v)\ln f_V(v)dv \\
&= -\int_{-\infty}^{\infty} f_V(v)\ln\left(\frac{1}{2\lambda}e^{-|v|/\lambda}\right)dv \\
&= \ln(2\lambda) + \frac{1}{\lambda}\int_{-\infty}^{\infty} f_V(v)|v|dv, \qquad (5)
\end{aligned}
$$

接下来计算 $\mathbf{E}[|V|]$, 即

$$
\begin{aligned}
\mathbf{E}[|V|] &= \int_{-\infty}^{\infty} |v|f_V(v)dv \\
&= \int_0^{\infty} \frac{v}{\lambda}e^{-v/\lambda}dv \\
&= \lambda. \qquad (6)
\end{aligned}
$$

所以 $h(X) = h(V) = 1 + \ln(2\lambda)$.

c) 瑞利分布的概率密度函数为 $f(x) = \alpha x e^{-\alpha x^2/2}, x \geq 0$, 首先计算

$$
\begin{aligned}
\mathbf{E}[X^2] = \int_0^{\infty} x^2 f(x)dx &= \int_0^{\infty} \alpha x^3 e^{-\alpha x^2/2}dx \\
&\overset{u=\alpha x^2/2}{=} \int_0^{\infty} \frac{2}{\alpha}ue^{-u}du \\
&= \frac{2}{\alpha}. \qquad (7)
\end{aligned}
$$

假定对数的底为 e, 计算微分熵

$$
h(X) = -\int_0^{\infty} f(x)\ln f(x)dx
$$

$$
\begin{aligned}
&= -\int_0^\infty f(x)\ln(\alpha x e^{-\alpha x^2/2})dx \\
&= -\ln\alpha - \int_0^\infty f(x)\ln(x)dx + \frac{\alpha}{2}\int_0^\infty f(x)x^2dx \\
&= 1 - \ln\alpha - \int_0^\infty f(x)\ln(x)dx, \qquad (8)
\end{aligned}
$$

令 $y = \sqrt{\alpha}x$, 则 $dy = \sqrt{\alpha}dx$, 代入上式可得

$$
\begin{aligned}
\int_0^\infty f(x)\ln(x)dx &= \int_0^\infty \alpha x e^{-\alpha x^2/2}\ln(x)dx \\
&= \int_0^\infty y e^{-\frac{y^2}{2}} \ln\left(\frac{y}{\sqrt{\alpha}}\right)dy \\
&= -\int_0^\infty y e^{-\frac{y^2}{2}} \ln(\sqrt{\alpha})dy + \int_0^\infty y e^{-\frac{y^2}{2}} \ln(y)dy \\
&= -\ln\sqrt{\alpha} + \int_0^\infty y e^{-\frac{y^2}{2}} \ln(y)dy, \qquad (9)
\end{aligned}
$$

令 $v = \frac{y^2}{2}$, 代入上式可得

$$
\begin{aligned}
\int_0^\infty y e^{-\frac{y^2}{2}} \ln y dy &= \int_0^\infty e^{-v}\ln(\sqrt{2v})dv \\
&= \int_0^\infty e^{-v}\ln(\sqrt{2})dv + \frac{1}{2}\int_0^\infty e^{-v}\ln(v)dv \\
&= \frac{\ln 2}{2} + \frac{-\gamma}{2}, \qquad (10)
\end{aligned}
$$

其中 γ 为欧拉常数，约等于 0.5772. 最后结合 (8), (9) 和 (10) 可得 $h(X) = 1 + \frac{\gamma - \ln 2\alpha}{2}$. □

第 2 题

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

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

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

解:

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

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

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

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

$$\begin{aligned}h(S_i|S_1, \cdots, S_{i-1}) &= h\left(\frac{i-1}{i}S_{i-1} + \frac{X_i}{i}|S_1, \cdots, S_{i-1}\right) \\ &= h\left(\frac{X_i}{i}|S_1, \cdots, S_{i-1}\right) \qquad (11) \\ &= h(X_i) - \log i, \qquad (12)\end{aligned}$$

其中 (11) 与 (12) 分别来源于讲义 Proposition 10.3 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 3 题

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

证明:

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

当且仅当 $X+Y$ 与 Y 独立，即 Y 为一个常数时，不等式取等. □

第 4 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, $|\mathbf{K}| \le \prod_{i=1}^{k} \mathbf{K}_{ii}$.

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

解: a): 约束条件等价为 $\int f_{\underline{X}}(\underline{x}) x_i x_j \mathrm{d}\underline{x} = \mathbf{K}_{ij}, \forall i,j \in \{1,2,\ldots,k\}$, 利用最大熵定理，满足题中条件的最大熵分布的密度函数为

$$f^*_{\underline{X}}(\underline{x}) = e^{\lambda_0 + \sum_{i,j} \lambda_{ij} x_i x_j} = e^{\lambda_0 + \underline{x}^T \mathbf{L} \underline{x}},$$

其中 $\mathbf{L} = [\lambda_{ij}]$, 所以 $f^*_{\underline{X}}$ 是零均值, 协方差矩阵为 $\mathbf{K}$ 的多维高斯分布的密度函数，则有

$$e^{\lambda_0} = \frac{1}{\sqrt{(2\pi)^k |2\mathbf{L}^{-1}|}} = \frac{1}{\sqrt{(2\pi)^k |\mathbf{K}|}}, \mathbf{L} = \frac{1}{2}\mathbf{K}^{-1}.$$

所以最大熵分布和最大熵分别为:

$$f^*_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{K}|}} \exp\left(-\frac{1}{2}\underline{x}^T \mathbf{K}^{-1} \underline{x}\right),$$
$$h(f^*_{\underline{X}}) = \frac{1}{2}\ln(2\pi e)^k |\mathbf{K}|.$$

b): 即证协方差矩阵满足哈达玛不等式. 令 $\underline{X} \sim \mathcal{N}(\mathbf{0},\mathbf{K})$, $\underline{X} = (X_1, X_2, \cdots .X_k)$, 则 $varX_i = \mathbf{E}[X_i^2] = \mathbf{K}_{ii}$, 可得

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

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

c): 设 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 是 n 维零均值正态分布向量, $\mathbf{X}_i \sim \phi_{\mathbf{K}_i}(\mathbf{x})$, $i = 1,2$. 随机变量 θ 的分布为 $\Pr\{\theta = 1\} = \lambda, \Pr\{\theta = 2\} = 1 - \lambda, \quad 0 \le \lambda \le 1$, 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 $\mathbf{K}_Z = \mathbf{E}\left[\mathbf{X}_\theta \mathbf{X}_\theta^T\right] = \lambda \mathbf{K}_1 + (1-\lambda)\mathbf{K}_2$. 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

$$\begin{aligned}\frac{1}{2}\ln(2\pi e)^n |\lambda \mathbf{K}_1 + (1-\lambda)\mathbf{K}_2| &\ge h(\mathbf{Z}) \ge h(\mathbf{Z} \mid \theta) \\ &= \lambda \frac{1}{2}\ln(2\pi e)^n |\mathbf{K}_1| + (1-\lambda)\frac{1}{2}\ln(2\pi e)^n |\mathbf{K}_2|,\end{aligned}$$

因此,

$$|\lambda \mathbf{K}_1 + (1-\lambda)\mathbf{K}_2| \ge |\mathbf{K}_1|^{\lambda} |\mathbf{K}_2|^{1-\lambda}.$$

对不等式两边取对数后可得 $\ln|\lambda \mathbf{K}_1 + (1-\lambda)\mathbf{K}_2| \ge \lambda \ln|\mathbf{K}_1| + (1-\lambda)\ln|\mathbf{K}_2|$. □

第 5 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 $D(f_X \| g_X)$ while 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

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

因此 f_X^* 最小化相对熵 $D(f_X \| g_X)$. □

第 6 题

Verify that in the scalar case, the EPI can be rewritten in the following equivalent form: for independent continuous random variables X and Y over $\mathbb{R}$, letting $\tilde{X}$ and $\tilde{Y}$ be independent Gaussian random variables satisfying $h(\tilde{X}) = h(X)$ and $h(\tilde{Y}) = h(Y)$, it holds that

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

解: 记 $\tilde{X}, \tilde{Y}$ 的方差分别为 $\sigma_{\tilde{X}}^2, \sigma_{\tilde{Y}}^2$, 则 $\tilde{X} + \tilde{Y}$ 的方差为 $\sigma_{\tilde{X}}^2 + \sigma_{\tilde{Y}}^2$, 所以有

$$N(\tilde{X} + \tilde{Y}) = \frac{1}{2\pi e} e^{2h(\tilde{X} + \tilde{Y})} = \frac{1}{2\pi e} e^{\ln 2\pi e (\sigma_{\tilde{X}}^2 + \sigma_{\tilde{Y}}^2)} = \sigma_{\tilde{X}}^2 + \sigma_{\tilde{Y}}^2,$$

同理有 $N(X)=N(\tilde{X})=\sigma_{\tilde{X}}^2, N(Y)=N(\tilde{Y})=\sigma_{\tilde{Y}}^2$, 根据 EPI 可得

$$\begin{aligned} N(X+Y) &\geq N(X)+N(Y) \\ &= N(\tilde{X})+N(\tilde{Y}) \\ &= \sigma_{\tilde{X}}^2+\sigma_{\tilde{Y}}^2 \\ &= N(\tilde{X}+\tilde{Y}), \end{aligned}$$

由此可得 $h(X+Y) \geq h(\tilde{X}+\tilde{Y})$. $\square$