Exercises om Information Theory and Entropy, EPFL 2023

1 Few useful equalities and inequalities on Entropies

In what follows, we shall denote the entropy of a random variable X with a distribution $p_X(x)$ as

$$\mathrm{H}(X) = -\int dx\, p_X(x) \log p_X(x)$$

while for many variables, we have e.g

$$\mathrm{H}(X, Y) = -\int dxdy\, p_{X,Y}(x, y) \log p_{X,Y}(x, y)$$

The conditional entropy is defined as

$$\mathrm{H}(X|Y) = -\int dxdy\, p_{X,Y}(x, y) \log p_{X|Y}(x|y)$$

1.1 Entropy of a Gaussian variable

Show that the entropy of a Gaussian variable sampled from $\mathcal{N}(m, \Delta)$ is given by

$$\mathrm{H}(X) = \frac{1}{2} \log 2\pi e \Delta$$

1.2 Gibbs inequality

Given two probability distributions $p(x)$ and $q(x)$, the *relative entropy*, or Kullback-Leibler divergence, is defined as

$$\mathrm{D}_{\mathrm{KL}}(p, q) = \int dx\, p(x) \log \frac{p(x)}{q(x)}$$

Using the fact that $\ln x \leq x - 1$, show the "Gibbs inequality" that states that $D_{\mathrm{KL}}(p, q) \geq 0$.

1.3 Mutual information

The mutual information between two (potentially) correlated variables X and Y is defined as the Kullback-Leibler divergence between their joint distribution and the factorized one. In other words, it reads

$$\mathrm{I}(X; Y) = \int dxdy\, p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x)\, p_Y(y)}$$

Show that the mutual information satisfies the following chain rules :

$$\mathrm{I}(X; Y) = \mathrm{H}(X) - \mathrm{H}(X|Y) = \mathrm{H}(Y) - \mathrm{H}(Y|X)$$

where (show it) the so-called conditional entropy $\mathrm{H}(X|Y)$ can be also be expressed as

$$\mathrm{H}(X|Y) = -\int dy\, p_Y(y) \int dx\, p_{X|Y}(x|y) \log p_{X|Y}(x|y)$$

2 The I-MMSE Theorem for scalars

Consider an unknown variable X^*, distributed as $P_X(x)$. It is observed through a noisy Gaussian channel so that we are given

$$Y = \sqrt{\lambda} X^* + Z \tag{1}$$

where Z is distributed as $\mathcal{N}(0,1)$. We remind that the mutual information between X and Y is given by

$$I(X;Y) = \int dxdy P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \tag{2}$$

1. Consider X distributed according to the posterior $P_{X|Y}(X|Y)$ (where Y is derived from X^*). Using $I(X;Y) = H(Y) - H(Y|X)$ show that the mutual information is given by

$$\begin{aligned} I(X;Y) &= -\mathbb{E}_y \log \int dx P_X(x) \frac{e^{-\frac{1}{2}(y-\sqrt{\lambda}x)^2}}{\sqrt{2\pi}} - \frac{1}{2}\log 2\pi e & (3) \\ &= \text{cst} - \mathbb{E}_{x^*,z} \log \int dx P_X(x) e^{-\frac{\lambda}{2}(x^*-x)^2 - z\sqrt{\lambda}(x^*-x)} & (4) \end{aligned}$$

2. We now define the Gibbs average $\langle . \rangle$, for a given z, x^*, such that

$$\langle f(x) \rangle = \frac{\int dx f(x) P_X(x) e^{-\frac{\lambda}{2}(x^*-x)^2 - z\sqrt{\lambda}(x^*-x)}}{\int dx P_X(x) e^{-\frac{\lambda}{2}(x^*-x)^2 - z\sqrt{\lambda}(x^*-x)}} \tag{5}$$

   Denoting $m = \mathbb{E}[\langle x \rangle x^*]$, $q = \mathbb{E}[\langle x \rangle^2]$ and $\rho = \mathbb{E}[(x^*)^2]$ [1] and using Stein's lemma $\mathbb{E}_z z f(z) = \mathbb{E}_z f'(z)$ prove the so-called I-MMSE theorem (Guo, Shamai, Verdu '05) :

$$\frac{dI}{d\lambda} = \frac{1}{2}\mathbb{E}_{x^*,z}(\langle x \rangle - x^*)^2 = \frac{1}{2}\text{MMSE} = \frac{1}{2}(\rho - m) \tag{6}$$

   where MMSE stands for "Minimal Mean Squared Error".

3 The I-MMSE Theorem for matrices

Consider an unknown vector $\mathbf{X}^* \in \mathbb{R}^d$, with each element sample X_i^* from $P_X(x)$. We construct a rank-one matrix $M = \mathbf{X}^*(\mathbf{X}^*)^T$. The matrix M is observed through a noisy channel, so we are given

$$Y = \sqrt{\frac{\lambda}{d}} M + Z$$

where Z is a symmetric random matrix with element sampled through $\mathcal{N}(0,1)$.

1. The minimal error in reconstruction of the matrix (the M-MMSE) is given by using the posterior means as an estimator $\hat{M} = \mathbb{E}_{X|Y}[\mathbf{X}(\mathbf{X})^T] = \langle \mathbf{X}(\mathbf{X})^T \rangle$. Show that the Matrix-MMSE, defined as $\mathbb{E}_{M,Z} \frac{1}{d^2} \sum_{ij} (\hat{M}_{ij} - M_{ij})^2$, is given by $\rho^2 - m^2$, with $\rho = \mathbb{E}[X_i^2]$ and $m^2 = \mathbb{E}[(\langle X_i \rangle X_i^*)^2]$.
2. Show that the mutual information is given by

$$I(Y,X) = \frac{\lambda \rho^2}{4} - \mathbb{E}_Y[\frac{1}{d} \log Z(Y)] \tag{7}$$

   with $Z(Y) = \mathbb{E}_X[\exp - \sum_{i \leq j} -\frac{\lambda}{2n} x_i^2 x_j^2 + \frac{\lambda}{\sqrt{n}} x_i x_j Y_{ij}]$.
3. Proceed again as in the previous exercise and prove the matrix I-MMSE theorem :

$$\partial_\lambda I(Y,X) = \frac{1}{4}\text{Matrix-MMSE}$$

1. Clue : remember the Nishimori relation $q = m$, and that we proved MMSE$=\rho - m$ in class !