ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 23 | Information Theory and Coding
Solutions to Homework 9 | Nov. 21, 2023

Problem 1.

$$
\begin{aligned}
h(X) &= \frac{1}{2}\log\left(2\pi e\sigma_x^2\right)\\
h(Y) &= \frac{1}{2}\log\left(2\pi e\sigma_y^2\right)\\
h(X,Y) &= \frac{1}{2}\log\left((2\pi e)^2\det(K)\right) = \frac{1}{2}\log\left((2\pi e)^2\left(\sigma_x^2\sigma_y^2 - \rho^2\sigma_x^2\sigma_y^2\right)\right)\\
I(X,Y) = h(X)+h(Y)-h(X,Y) &= \frac{1}{2}\log\frac{1}{1-\rho^2}
\end{aligned}
$$

Note that the result does not depend on σ_x, σ_y, which says that normalization does not change the mutual information.

Problem 2.

(a) This is by the definition of mutual information once we note that $p_{Y|X}(y|x) = p_Z(y-x)$.

(b) Note that $p_X(x)p_Z(y-x)$ is simply the joint distribution of (x,y), and thus the integral

$$\iint p_X(x)p_Z(y-x)\ln\frac{\mathcal{N}_{\sigma^2}(y-x)}{\mathcal{N}_{\sigma^2+P}(y)}\,dxdy.$$

is an expectation, namely

$$E\ln\frac{\mathcal{N}_{\sigma^2}(Y-X)}{\mathcal{N}_{\sigma^2+P}(Y)}.$$

Substituting the formula for $\mathcal{N}$, this in turn, is

$$
\begin{aligned}
&E\ln\frac{\mathcal{N}_{\sigma^2}(Y-X)}{\mathcal{N}_{\sigma^2+P}(Y)}\\
&\quad= \frac{1}{2}\ln\left(1+P/\sigma^2\right) + \frac{1}{2(\sigma^2+P)}E[Y^2] - \frac{1}{2\sigma^2}E[(Y-X)^2]\\
&\quad= \frac{1}{2}\ln\left(1+P/\sigma^2\right) + \frac{1}{2(\sigma^2+P)}E[(X+Z)^2] - \frac{1}{2\sigma^2}E[Z^2]\\
&\quad= \frac{1}{2}\ln\left(1+P/\sigma^2\right) + \frac{1}{2(\sigma^2+P)}E[X^2+Z^2+2XZ] - \frac{1}{2}\\
&\quad= \frac{1}{2}\ln\left(1+P/\sigma^2\right) + \frac{1}{2(\sigma^2+P)}(P+\sigma^2+0) - \frac{1}{2}\\
&\quad= \frac{1}{2}\ln\left(1+P/\sigma^2\right)
\end{aligned}
$$

(c) The steps we need to justify read

$$\begin{aligned}
\ln(1+P/\sigma^2) - I(X;Y) &= \iint p_X(x)p_Z(y-x)\ln\frac{\mathcal{N}_{\sigma^2}(y-x)p_Y(y)}{\mathcal{N}_{\sigma^2+P}(y)p_Z(y-x)}\,dxdy \\
&\le \iint \frac{p_X(x)\mathcal{N}_{\sigma^2}(y-x)p_Y(y)}{\mathcal{N}_{\sigma^2+P}(y)}\,dxdy - 1 \\
&= \int p_Y(y)\,dy - 1 \\
&= 0.
\end{aligned}$$

The first equality is by substitution of parts (a) and (b). The inequality is by $\ln(x) \le x-1$. The next equality is by noting that

$$\int p_X(x)\mathcal{N}_{\sigma^2}(y-x)dx = (p_X * \mathcal{N}_{\sigma^2})(y) = (\mathcal{N}_P * \mathcal{N}_{\sigma^2})(y) = \mathcal{N}_{P+\sigma^2}(y).$$

The last equality is because any density function integrates to 1.

(d) The conclusion is made by noting that the right hand side of the first equality in (c) is equal to zero if $p_Z = \mathcal{N}_{\sigma^2}$.

Problem 3. Let the input distribution be p. We thus have

$$p(-1)+p(0)+p(1) = 1 \qquad p(-1)\ge 0, p(0)\ge 0, p(1)\ge 0$$

(since p is a distribution) and, to satisfy $E[b(X)] \le \beta$ we must have

$$p(-1)+p(1) = 1-p(0) \le \beta.$$

Moreover,

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&\overset{(a)}{=} H(Y) - p(0) \\
&\overset{(b)}{\le} 1 - p(0) \\
&\overset{(c)}{\le} \min\{1,\beta\}.
\end{aligned}$$

where (a) follows because given $\{X=-1\}$ or $\{X=1\}$ there is no uncertainity in Y while given $\{X=0\}$, Y is uniformly distributed in $\{-1,1\}$, (b) holds since Y is binary with equality if $p(-1)+\frac{1}{2}p(0) = p(1)+\frac{1}{2}p(0) = \frac{1}{2}$ (which happens if we choose $p(1) = p(-1) = \frac{1}{2}(1-p(0))$) and (c) holds because of the cost constraint and is equality if we choose $p(0) = \max\{1-\beta, 0\}$. Hence, the capacity is

$$C = \begin{cases} \beta, & \text{if } \beta \le 1 \\ 1, & \text{if } \beta > 1 \end{cases}.$$

Problem 4. (a) All rates less than $\frac{1}{2}\log_2(1+\frac{P}{\sigma^2})$ are achievable.

(b) The new noise $Z_1 - \rho Z_2$ has zero mean and variance $\mathrm{E}((Z_1-\rho Z_2)^2) = \sigma^2(1-\rho^2)$. Therefore, all rates less than $\frac{1}{2}\log_2(1+\frac{P}{\sigma^2(1-\rho^2)})$ are achievable.

(c) The capacity is $C = \max I(X; Y_1, Y_2) = \max(h(Y_1, Y_2) - h(Z_1, Z_2)) = \frac{1}{2}\log_2(1 + \frac{P}{\sigma^2(1-\rho^2)})$. This shows that the scheme used in (b) is a way to achieve capacity.

PROBLEM 5. (a) In this exercise we assume all the vectors are column vectors. We know that $(X_1, \cdots, X_n, Y_1, \cdots, Y_m)$ are jointly Gaussian random variables if and only if any linear combination of these variables is normally distributed. This means that any linear combination of $\mathrm{X} = (X_1, X_2, \cdots, X_n)$ is normally distributed and thus X is an n Gaussian random vector. Similarly, the vector $\mathrm{Y} = (Y_1, \cdots, Y_m)$ is an m dimensional random vector.

Moreover, we can write $(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = (\mathrm{X}, \mathrm{Y})$. So its covariance matrix is

$$K = E\left(\begin{bmatrix}\mathrm{X}\\ \mathrm{Y}\end{bmatrix}\begin{bmatrix}\mathrm{X}^T & \mathrm{Y}^T\end{bmatrix}\right) = \begin{bmatrix}E\left(\mathrm{XX}^T\right) & E\left(\mathrm{XY}^T\right)\\ E\left(\mathrm{YX}^T\right) & E\left(\mathrm{YY}^T\right)\end{bmatrix} = \begin{bmatrix}K_{11} & K_{12}\\ K_{21} & K_{22}\end{bmatrix}.$$

So $K_{11} = E\left(\mathrm{XX}^T\right)$ and $K_{22} = E\left(\mathrm{YY}^T\right)$. Thus the vector $\mathrm{X} = (X_1, \cdots, X_n)$ is normally distributed with covariance matrix K_{11} and the vector $\mathrm{Y} = (Y_1, \cdots, Y_m)$ is normally distributed with covariance matrix K_{22}.

Hence, using the results derived in class we get

$$h(X_1, \cdots, X_n) = \frac{1}{2}\ln\left((2\pi e)^n \det\left(K_{11}\right)\right),$$

$$h(Y_1, \cdots, Y_m) = \frac{1}{2}\ln\left((2\pi e)^m \det\left(K_{22}\right)\right)$$

and

$$h(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = \frac{1}{2}\ln\left((2\pi e)^{n+m} \det\left(K\right)\right).$$

(b) Let A_{11} be an $n \times n$ matrix and A_{22} be an $m \times m$ matrix. So A becomes an $(n+m) \times (n+m)$ matrix. Since A is a positive definite matrix then there exists an $n+m$ dimensional Gaussian random vector which covariance matrix is A. Let's denote this vector as $(X_1, \cdots, X_n, Y_1, \cdots, Y_m)$. From question (a) we know that

$$h(X_1, \cdots, X_n) = \frac{1}{2}\ln\left((2\pi e)^n \det\left(A_{11}\right)\right),$$

$$h(Y_1, \cdots, Y_m) = \frac{1}{2}\ln\left((2\pi e)^m \det\left(A_{22}\right)\right)$$

and

$$h(X_1, \cdots, X_n, Y_1, \cdots, Y_m) = \frac{1}{2}\ln\left((2\pi e)^{n+m} \det\left(A\right)\right).$$

Moreover, we know that

$$h\left(X_1, \cdots, X_n, Y_1, \cdots, Y_m\right) \leq h\left(X_1, \cdots, X_n\right) + h\left(Y_1, \cdots, Y_m\right).$$

So,

$$\begin{aligned}\frac{1}{2}\ln\left((2\pi e)^{n+m}\det\left(A\right)\right) &\leq \frac{1}{2}\ln\left((2\pi e)^n\det\left(A_{11}\right)\right) + \frac{1}{2}\ln\left((2\pi e)^m\det\left(A_{22}\right)\right)\\ (2\pi e)^{n+m}\det\left(A\right) &\leq (2\pi e)^n\det\left(A_{11}\right) \times (2\pi e)^m\det\left(A_{22}\right)\\ \det(A) &\leq \det(A_{11})\det(A_{22}).\end{aligned}$$