ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 23 Information Theory and Coding
Solutions to Homework 9 Nov. 29, 2017

Problem 1. Choose $x, y \in U$ and $\alpha \in [0,1]$. The convexity if f associated to the fact that h is an increasing function over $[a,b]$ shows

$$g(\alpha x + (1-\alpha)y) = h\big(f\left(\alpha x + (1-\alpha)y\right)\big) \le h\big(\alpha f(x) + (1-\alpha)f(y)\big).$$

The convexity of h gives finally

$$g(\alpha x + (1-\alpha)y) \le \alpha h(f(x)) + (1-\alpha)h(f(y)) = \alpha g(x) + (1-\alpha)g(y).$$

Problem 2. Let us show that the function $g : \lambda \mapsto f(\lambda v_1 + (1-\lambda)v_2)$ is convex (in λ). Choosing $\lambda_x, \lambda_y \in [0,1]$ and $\alpha \in [0,1]$, we use the convexity of f in v to write

$$\begin{aligned}
g(\alpha\lambda_x + (1-\alpha)\lambda_y) &= f\big((\alpha\lambda_x + (1-\alpha)\lambda_y)v_1 + (1-(\alpha\lambda_x + (1-\alpha)\lambda_y))v_2\big) \\
&= f\big(\alpha\lambda_x v_1 + (1-\alpha)\lambda_y v_1 + v_2 - \alpha\lambda_x v_2 - (1-\alpha)\lambda_y v_2\big) \\
&= f\big(\alpha\lambda_x v_1 + (1-\alpha)\lambda_y v_1 + \big(\alpha + (1-\alpha)\big)v_2 - \alpha\lambda_x v_2 - (1-\alpha)\lambda_y v_2\big) \\
&= f\big(\alpha(\lambda_x v_1 + (1-\lambda_x)v_2) + (1-\alpha)(\lambda_y v_1 + (1-\lambda_y)v_2)\big) \\
&\le \alpha f(\lambda_x v_1 + (1-\lambda_x)v_2) + (1-\alpha)f(\lambda_y v_1 + (1-\lambda_y)v_2) \\
&= \alpha g(\lambda_x) + (1-\alpha)g(\lambda_y).
\end{aligned}$$

Problem 3. Taking the hint:

$$\begin{aligned}
0 &\le D(q\|p) \\
&= \int q(x)\log\frac{q(x)}{p(x)}\,dx \\
&= \int q(x)\log q(x)\,dx + \int q(x)\log\frac{1}{p(x)}\,dx \\
&= -h(q) + \int q(x)\log\frac{1}{p(x)}\,dx.
\end{aligned}$$

Now, note that $\log[1/p(x)]$ is of the form $\alpha + \beta x$, and since densities p and q have the same mean, we conclude that

$$\int q(x)\log\frac{1}{p(x)}\,dx = \int p(x)\log\frac{1}{p(x)}\,dx = h(p).$$

Thus, $0 \le -h(q) + h(p)$, yielding the desired conclusion.

Problem 4.

First Method

(a) It suffices to note that $H(X|Y) = H(X + f(Y)|Y)$ for any function f.

(b) Since among all random variables with a given variance the gaussian maximizes the entropy, we have

$$H(X-\alpha Y) \le \frac{1}{2}\log 2\pi e E((X-\alpha Y)^2)\ .$$

(c) From (a) and (b) we have

$$\begin{aligned} I(X;Y) &= H(X) - H(X - \alpha Y|Y) \\ &\geq H(X) - H(X - \alpha Y) \\ &\geq H(X) - \frac{1}{2}\log 2\pi e E((X-\alpha Y)^2)\ . \end{aligned}$$

(d) We have that $\frac{dE((X-\alpha Y)^2)}{d\alpha} = 0$ is equivalent to $E(Y(X - \alpha Y)) = 0$. Hence $\frac{dE((X-\alpha Y)^2)}{d\alpha}$ is equal to zero for $\alpha = \alpha^* = \frac{E(XY)}{E(Y^2)}$. Now on the one hand $E(XY) = E(X(X+Z)) = E(X^2) + E(XZ)$ and because of the independence between X and Z and the fact that Z has zero mean we have that $E(XZ) = 0$, and hence $E(XY) = P$. On the other hand $E(Y^2) = E((X+Z)^2) = E(X^2) + 2E(XZ) + E(Z^2) = P + 0 + \sigma^2$. Therefore $\alpha^* = P/(P+\sigma^2)$.
Then observing that $E((X-\alpha Y)^2)$ is a convex function of α we deduce that $E((X-\alpha Y)^2)$ is minimized for $\alpha = \alpha^*$. Finally an easy computation yields to $E((X-\alpha^* Y)^2) = \frac{\sigma^2 P}{\sigma^2 + P}$.

(e) Since X is gaussian from (c) and (d) we deduce that

$$\begin{aligned} I(X;Y) &\geq \frac{1}{2}\log 2\pi e P - \frac{1}{2}\log 2\pi e \frac{\sigma^2 P}{\sigma^2+P} \\ &= \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right)\ . \end{aligned} \tag{1}$$

with equality if and only if Z is gaussian with covariance σ^2.

Second Method

(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

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

(c) The steps we need to justify read

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

The first equality is by substitution of parts (a) and (b). The inequality is by $\ln(x) \leq x - 1$. 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 $p_Z = \mathcal{N}_{\sigma^2}$.