Handout 23 — Information Theory and Coding
Solutions to Homework 9 — Nov. 23, 2015

Problem 1.

(a) For all $x, y \in \mathbb{R}$, choosing $\alpha \in [0,1]$, we use the convexity of each f_i, $1 \le i \le n$, to get

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \sum_{i=1}^{n} c_i f_i(\alpha x + (1-\alpha)y) \\ &\le \sum_{i=1}^{n} c_i \left(\alpha f_i(x) + (1-\alpha) f_i(y)\right) \\ &= \alpha \sum_{i=1}^{n} c_i f_i(x) + (1-\alpha) \sum_{i=1}^{n} c_i f_i(y) \\ &= \alpha f(x) + (1-\alpha) f(y). \end{aligned}$$

(b) For all $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, choosing $\alpha \in [0,1]$, observe first that $\alpha x + (1-\alpha)y = (\alpha x_1 + (1-\alpha)y_1, \alpha x_2 + (1-\alpha)y_2, \dots, \alpha x_n + (1-\alpha)y_n)$. We then use the convexity of each f_i, $1 \le i \le n$, to get

$$\begin{aligned} g(\alpha x + (1-\alpha)y) &= \sum_{i=1}^{n} c_i f_i(\alpha x_i + (1-\alpha)y_i) \\ &\le \sum_{i=1}^{n} c_i \left(\alpha f_i(x_i) + (1-\alpha) f_i(y_i)\right) \\ &= \alpha \sum_{i=1}^{n} c_i f_i(x_i) + (1-\alpha) \sum_{i=1}^{n} c_i f_i(y_i) \\ &= \alpha g(x) + (1-\alpha) g(y). \end{aligned}$$

Problem 2. For all $\tilde{x} \in D$, $f(\tilde{x}) = \sup_{i \in I} f_i(x)$ iff (i) $f(\tilde{x}) \ge f_i(\tilde{x})$ for all $i \in I$ and (ii) any $s \in \mathbb{R}$ satisfying $s < f(\tilde{x})$ is such that there exists $i \in I$ satisfying $s < f_i(\tilde{x})$.

Choose $x, y \in D$ and $\alpha \in [0,1]$.

First, pick $i \in I$. With the definition of f (point (i)) and the convexity of each f_i, $i \in I$, we get

$$f_i(\alpha x + (1-\alpha)y) \le \alpha f_i(x) + (1-\alpha) f_i(y) \le \alpha f(x) + (1-\alpha) f(y).$$

Second, since the inequality $f_i(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$ holds for all $i \in I$, we use the definition of f (point (ii)) to claim

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

To see this, observe that, if it was not the case, then $s = \alpha f(x) + (1-\alpha) f(y) < f(\alpha x + (1-\alpha)y)$ would give the contradiction $s < f_i(\tilde{x})$ with $\tilde{x} = \alpha x + (1-\alpha)y$.

PROBLEM 3. 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

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

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 4. 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 5.

(a) By the chain rule

$$I(U,T;V) = I(U;V) + I(T;V|U) = I(U;V),$$

since $I(T;V|U) = 0$ from the Markov property. Also,

$$I(U,T;V) = I(T;V) + I(U;V|T) \ge I(U;V|T),$$

from the non-negativity of the mutual information. These together imply that $I(U;V) \ge I(U;V|T)$.

(b)

$$I(X;Y|W) = \Pr\{W=1\}I(X;Y|W=1) + \Pr\{W=2\}I(X;Y|W=2)$$

Conditional on $W = k$, the distribution of (X,Y) is $p_k(x)p(y|x)$, thus

$$I(X;Y|W) = \lambda I_1 + (1-\lambda)I_2.$$

(c) We obtain $p(x)$ by summing $p(w,x,y)$ over y and w. This gives

$$p(x) = \lambda p_1(x) + (1-\lambda)p_2(x).$$

(d) Note that

$$p(w,x,y) = p(w)p(x|w)p(y|x),$$

that is Y is independent of W when X is given. Thus from (a)

$$I(X;Y) \ge I(X;Y|W). \tag{1}$$

Letting $f(p_X)$ denote the value of $I(X;Y)$ as a function of the distribution of X we can rewrite (1) as

$$f(\lambda p_1 + (1-\lambda)p_2) \ge \lambda f(p_1) + (1-\lambda)f(p_2),$$

which says that f is concave.

Problem 6. Since X and Z are both in the interval $[-1,1]$, their sum $X+Z$ lies in the interval $[-2,+2]$. If we could *choose* the distribution of $X+Z$ as we wished (without the constraint that it has to be the sum of two independent random variables, one of which is uniform) we would have chosen it to be uniform on the interval $[-2,+2]$ to have the largest entropy. Observe now that if we choose X as the random variable that equals $+1$ with probability $1/2$ and -1 with probability $1/2$, then $X+Z$ is uniform in $[-2,+2]$ and thus this distribution maximizes the entropy. An alternate derivation is as follows: note that since X and Z are independent, the moment generating functions of the random variables involved satisfy $E[e^{s(X+Z)}]=E[e^{sX}]E[e^{sZ}]$. Now, we know that $E[e^{sZ}]=\int e^{sz}f_Z(z)\,dz=\int_{-1}^{+1}\frac{1}{2}e^{sz}\,dz=[e^s-e^{-s}]/(2s)$. Similarly, if we want $X+Z$ to be uniform on $[-2,2]$, we can compute $E[e^{s(X+Z)}]=[e^{2s}-e^{-2s}]/(4s)$. This then requires $E[e^{sX}]=\frac{1}{2}[e^{2s}-e^{-2s}]/[e^s-e^{-s}]=\frac{1}{2}[e^s+e^{-s}]$ which is the moment generating function of a random variable which takes on the values $+1$ and -1, each with probability $1/2$.

Similarly, under the constraint XZ lies in the interval $[-1,+1]$, and the best we could hope is that XZ is uniform on this interval. But this can be achieved by making sure that X only takes on the values $+1$ or -1.

Problem 7. 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 8.

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

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

(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

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

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 $p_Z = \mathcal{N}_{\sigma^2}$.