ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 28 Information Theory and Coding

Solutions to homework 12 December 24, 2013

PROBLEM 1. 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 $E[e^{s(X+Z)}] = E[^{sX}]E[e^{sX}]$. 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 2. 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 3. It is clear that the input distribution that maximizes the capacity is $X \sim \mathcal{N}(0, P)$. Evaluating the mutual information for this distribution,

$$\begin{aligned}
C_2 &= \max I(X; Y_1, Y_2) \\
&= h(Y_1, Y_2) - h(Y_1, Y_2|X) \\
&= h(Y_1, Y_2) - h(Z_1, Z_2|X) \\
&= h(Y_1, Y_2) - h(Z_1, Z_2)
\end{aligned}$$

Now since

$$(Z_1, Z_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}\right),$$

we have

$$h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2 |K_Z| = \frac{1}{2}\log(2\pi e)^2 N^2(1-\rho^2).$$

Since $Y_1 = X + Z_1$, and $Y_2 = X + Z_2$, we have

$$(Y_1, Y_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} P+N & P+\rho N \\ P+\rho N & P+N \end{bmatrix}\right),$$

and

$$h(Y_1, Y_2) = \frac{1}{2}\log(2\pi e)^2 |K_Y| = \frac{1}{2}\log(2\pi e)^2 (N^2(1-\rho^2) + 2PN(1-\rho)).$$

Hence the capacity is

$$\begin{aligned} C_2 &= h(Y_1, Y_2) - h(Z_1, Z_2) \\ &= \frac{1}{2}\log\left(1 + \frac{2P}{N(1+\rho)}\right). \end{aligned}$$

(a) $\rho = 1$. In this case, $C = \frac{1}{2}\log(1+P/N)$, which is the capacity of a single look channel. This is not surprising, since in this case $Y_1 = Y_2$.

(b) $\rho = 0$. In this case,

$$C = \frac{1}{2}\log\left(1 + 2P/N\right),$$

which corresponds to using twice the power in a single look. The capacity is the same as the capacity of the channel $X \to (Y_1 + Y_2)$.

(c) $\rho = -1$. In this case, $C = \infty$, which is not surprising since if we add Y_1 and Y_2, we can recover X exactly, and so is equivanent to having a noiseless channel.

Note that the capacity of the above channel in all cases is the same as the capacity of the channel $X \to Y_1 + Y_2$. This is not true in general.

Problem 4. (a) By the water-filling solution discussed in class, it follows that we will put all the signal power into the channel with less noise until the total power of noise + signal in that channel equals the noise power in the other channel. After that, we will split any additional power evenly between the two channels.

Thus the combined channel begins to behave like a pair of parallel channels when the signal power is equal to the difference of the two noise powers, i.e., when $2P = \sigma_1^2 - \sigma_2^2$.

(b) Since we are interested in the asymptotics $P/\sigma_1^2 \longrightarrow \infty$ without loss of generality we assume the waterpourring level to be greater than σ_1^2. Hence $P_i = \lambda - \sigma_i^2$, $i = 1, 2$. It follows that

$$\begin{aligned} C_1(P) - C_2(P) &= \frac{1}{2}\log\left(1 + \frac{P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1 + \frac{P_2}{\sigma_2^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_1^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_2^2}\right) \\ &= \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2}\right) + \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_1^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_2^2}\right) \end{aligned}$$

Now

$$\begin{aligned}\frac{1}{2}\log\left(1+\frac{P}{\sigma_1^2}\right) &= \frac{1}{2}\log\left(1+\frac{P_1}{\sigma_1^2}+\frac{P-P_1}{\sigma_1^2}\right)\\ &= \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2}+\frac{P-P_1}{\sigma_1^2}\right)\end{aligned}$$

and similarly

$$\frac{1}{2}\log\left(1+\frac{P}{\sigma_2^2}\right) = \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}+\frac{P-P_2}{\sigma_2^2}\right) = \frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}-\frac{P-P_1}{\sigma_2^2}\right) .$$

We then deduce that

$$\begin{aligned}C_1(P)-C_2(P) &= \frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2}\right)+\frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}\right)-\frac{1}{2}\log\left(\frac{\lambda}{\sigma_1^2}+\frac{P-P_1}{\sigma_1^2}\right)-\frac{1}{2}\log\left(\frac{\lambda}{\sigma_2^2}-\frac{P-P_1}{\sigma_2^2}\right)\\ &= -\frac{1}{2}\log\left(1+\frac{P-P_1}{\lambda}\right)-\frac{1}{2}\log\left(1-\frac{P-P_1}{\lambda}\right) .\end{aligned}$$

We conclude by showing that $\frac{P-P_1}{\lambda}$ tends to zero as P/σ_1^2 tends to infinity. Since $P_i=\lambda-\sigma_i^2$, $i=1,2$, and since $2P=2\lambda-\sigma_1^2-\sigma_2^2$, we have that

$$\begin{aligned}\frac{P-P_1}{\lambda} &= \frac{P-(\lambda-\sigma_1^2)}{\lambda}\\ &= \frac{2(\sigma_2^2-\sigma_1^2)}{2P+\sigma_1^2+\sigma_2^2} .\end{aligned}$$

By assumption P/σ_1^2 tends to infinity, and because $\sigma_2^2<\sigma_1^2$, we have that $\frac{2(\sigma_2^2-\sigma_1^2)}{2P+\sigma_1^2+\sigma_2^2}$ tends to zero as P/σ_1^2 tends to infinity. This gives the desired result.

Problem 5. (a) All rates less than $\frac{1}{2}\log_2(1+\frac{P}{\sigma^2})$ are achievable.

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

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

Problem 6. *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

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

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

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

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