ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 23 Information Theory and Coding
Solutions to homework 9 December 10, 2007

Problem 1. 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 2. 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 equivalent 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 3. (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 waterpouring 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

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

We have that $\frac{\sigma_1^2-\sigma_2^2}{2P+\sigma_1^2+\sigma_2^2}$ tends to zero as P/σ_1^2 tends to infinity. This gives the desired result.

PROBLEM 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 $\mathrm{E}((Z_1 - Z_2)^2) = \sigma^2(1-\rho^2)$. 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. *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

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

(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) The lower bound for $I(X;Y)$ in part (c) holds for all values of α. In particular, it holds for $\alpha = \alpha^*$. 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

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

(c) The steps we need to justify read

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

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 6.

(a) We have

$$I(X;Y)=h(Y)-h(Y|X)=h(Y)-h(Z|X)=h(Y)-h(Z).$$

where the last equality is because Z is independent of X.

(b) In the natural log basis,

$$h(Z) = -\int f_Z(z) \ln f_Z(z)\, dz = \int_0^\infty z e^{-z}\, dz = 1 \text{ nats.}$$

(c) Since $Y = X + Z$, the expectation of Y, $E[Y]$ equals $E[X] + E[Z]$. Since $E[X]$ is constrained to be less than or equal to P and $E[Z] = 1$, we see that $E[Y] = P + 1$ (since the capacity increases with power). Since X is constrained to be non-negative and so is Z, we see that Y is also constrained to be non-negative.
From problem 2 we know that among non-negative random variables of a given expectation λ, the one with density $p(y) = e^{-y/\lambda}/\lambda$ has the largest differential entropy. This differential entropy in natural units is

$$\int_0^\infty \frac{e^{-y/\lambda}}{\lambda}[\ln \lambda + y/\lambda]\, dy = \ln \lambda + 1 \text{ nats.}$$

Thus, since $E[Y] = 1 + P$ the differential entropy of Y is less than or equal to $1 + \ln(1 + P)$, which implies

$$C \leq \ln(1 + P) \text{ nats}$$

At this point, we do not know if Y can be made to have an exponential distribution with mean $1 + P$ so we cannot know if this above inequality is an equality or not.

(d) The Laplace transform of the random variable Y is $E(e^{sY}) = E(e^{s(X+Z)}) = E(e^{sX})E(e^{sZ})$, where the latter equality follows from the independence of X and Z. Therefore we have that $E(e^{sX}) = \frac{E(e^{sY})}{E(e^{sZ})}$. Computing $E(e^{sY})$,

$$\begin{aligned} E(e^{sY}) &= \int_\infty^\infty e^{sy} f_Y(y) dy \\ &= \int_0^\infty e^{sy} \mu e^{-\mu y} dy \\ &= \frac{\mu}{\mu - s} \qquad \forall s \leq \mu \end{aligned}$$

The expectation is not defined for $s > \mu$ (as the integral blows up). Likewise, we evaluate $E(e^{sZ}) = \frac{1}{1-s}$ (defined for $s \leq 1$). Therefore for $s \leq \min(1, \mu)$, we can evaluate $E(e^{sX})$ as

$$\begin{aligned} E(e^{sX}) &= \frac{E(e^{sY})}{E(e^{sZ})} \\ &= \mu \frac{1-s}{\mu - s} \\ &= \mu + (1-\mu)\frac{\mu}{\mu - s} \end{aligned}$$

Inverting the Laplace transform $E(e^{sX})$ gives us the distribution of the X that gives an exponential distribution for Y. From inspection, we can deduce this distribution of X to be

$$f_X(x) = \mu\delta(x) + (1-\mu)\mu e^{-\mu x} \qquad x \geq 0$$

Notice that the distribution is a convex combination of the exponential distribution and the distribution that puts all the mass on one point (in this case the point $x = 0$).

(e) By taking $\mu = 1/(1+P)$, we see that there is a density on X which makes the density of Y an exponential with mean $1+P$. Furthermore, this density on X makes X non-negative, and, $E[X] = E[Y] - E[Z] = P$. Thus, the bound of part (c) can be achieved.