Solutions: Homework Set # 10

Problem 1 (A mutual information game)

(a) Let $a_b = \arg\min_a f(a,b)$, *i.e.*, $f(a,b) \geq f(a_b,b), \forall a,b$. Then by taking the maximum of both sides,

$$\max_b f(a,b) \geq \max_b f(a_b,b) = \max_b \min_a f(a,b).$$

Note that the RHS does not depend on a anymore, while the LHS still depends on a. However, since the inequality holds for all a and b, it also holds for the minimizing a, *i.e.*,

$$\min_a \max_b f(a,b) \geq \max_b \min_a f(a,b).$$

(b)

$$\begin{aligned} I(X;X+Z^*) &= h(X+Z^*) - h(X+Z^*|X) \\ &= h(X+Z^*) - h(Z^*) \\ &\leq h(X^*+Z^*) - h(Z^*) \\ &= I(X^*;X^*+Z^*) \end{aligned}$$

where the inequality follows from the fact that given the variance, the entropy is maximized by the Gaussian distribution.

(c)

1. This is just expansion of mutual information as $I(X;X+Z) = h(X+Z) - h(X+Z|Z) = h(Y) - h(Z)$ since X and Z are independent.
2. Each entropy expression is replaced by its definition.
3. Note that $f_{Y^*}(y) = \frac{1}{\sqrt{2\pi(P+N)}} \exp\left(-\frac{y^2}{2(P+N)}\right)$. Therefore, $\log f_{Y^*}(y) = -\frac{y^2}{2(P+N)} - \frac{1}{2}\log 2\pi(P+N)$.

$$\begin{aligned} \int_y f_{Y^*}(y)\log f_{Y^*}(y)dy &= \int_y f_{Y^*}(y)\left[-\frac{y^2}{2(P+N)} - \frac{1}{2}\log 2\pi(P+N)\right]dy \\ &= -\frac{1}{2(P+N)}\int_y y^2 f_{Y^*}(y)dy - \frac{1}{2}\log 2\pi(P+N)\int_y f_{Y^*}(y)d_y \\ &= -\frac{1}{2(P+N)}\mathbb{E}_{Y^*}[y^2] - \frac{1}{2}\log 2\pi(P+N) \\ &\stackrel{(a)}{=} -\frac{1}{2(P+N)}\mathbb{E}_{Y}[y^2] - \frac{1}{2}\log 2\pi(P+N) \\ &= -\frac{1}{2(P+N)}\int_y y^2 f_{Y}(y)dy - \frac{1}{2}\log 2\pi(P+N)\int_y f_{Y}(y)d_y \\ &= \int_y f_{Y}(y)\left[-\frac{y^2}{2(P+N)} - \frac{1}{2}\log 2\pi(P+N)\right]dy \\ &= \int_y f_Y(y)\log f_{Y^*}(y)dy \end{aligned}$$

where (a) follows from the fact that $\mathbb{E}_{Y^*}[y^2] = \mathbb{E}_Y[y^2]$. The same proof holds for Z and Z^*.

4. Integration is a linear operation, and

$$\begin{aligned} -\int_y f_{Y^*}(y)\log f_{Y^*}(y)dy + \int_y f_Y(y)\log f_Y(y)dy &= \int_y f_Y(y)\left[\log f_Y(y) - \log f_{Y^*}(y)\right]dy \\ &= \int_y f_Y(y)\log\frac{f_Y(y)}{f_{Y^*}(y)}dy. \end{aligned}$$

Similarly, we can rewrite the two integrals on Z.

5.

$$\begin{aligned} &\int_y f_Y(y)\log\frac{f_Y(y)}{f_{Y^*}(y)}dy + \int_z f_Z(z)\log\frac{f_{Z^*}(z)}{f_Z(z)}dz \\ &\quad= \int_y\left(\int_z f_{Y,Z}(y,z)dz\right)\log\frac{f_Y(y)}{f_{Y^*}(y)}dy + \int_z\left(\int_y f_{Y,Z}(y,z)dy\right)\log\frac{f_{Z^*}(z)}{f_Z(z)}dzdy \\ &\quad= \int_y\int_z f_{Y,Z}(y,z)\log\frac{f_Y(y)}{f_{Y^*}(y)}dzdy + \int_y\int_z f_{Y,Z}(y,z)\log\frac{f_{Z^*}(z)}{f_Z(z)}dzdy \\ &\quad= \int_y\int_z f_{Y,Z}(y,z)\log\frac{f_Y(y)f_{Z^*}(z)}{f_{Y^*}(y)f_Z(z)}dzdy \end{aligned}$$

6. By concavity of the function $\log(\cdot)$.
7. Note that $Y = X^* + Z$. Therefore

$$f_{Y,Z}(y,z) = f_{X^*,Z}(y-z,z) = f_{X^*}(y-z)f_Z(z),$$

where the last equality follows form the fact that X^* and Z are independent.

8. This should be in fact equality. The reason is that $f_Z(z)$ can be cancelled from the nominator and the denominator, and then we take out every term does not depend on z from the inner integral.
9. Again since $Y^* = X^* + Z^*$, and X^* and Z^* are independent, we have $f_{Y^*}(y) = \int_z f_{X^*}(y-z)f_{Z^*}(z)dz$.
10. By cancelling $f^{Y^*}(y)$, the remaining would be $\int_y f_Y(y)dy$ which equals 1 since $f_Y(y)$ is a probability distribution.

(d) Using parts (b) and (c) we have,

$$\begin{aligned} \min_{p(z)}\max_{p(x)} I(X;X+Z) &\le \max_{p(x)} \le \max_{p(x)} I(X;X+Z^*) \\ &= I(X^*;X^*+Z^*) \\ &= \min_{p(z)} I(X^*;X^*+Z) \\ &\le \max_{p(x)}\min_{p(z)} I(X;X+Z).. \end{aligned} \tag{1}$$

On the other hand, the result of part (a) for $f(p(z),p(x)) = I(X;X+Z)$ gives us

$$\min_{p(z)}\max_{p(x)} I(X;X+Z) \ge \max_{p(x)}\min_{p(z)} I(X;X+Z). \tag{2}$$

Combining (1) and (2), we have

$$\begin{aligned}\min_{p(z)\in\mathbb{F}_N}\max_{p(x)\in\mathbb{F}_P} I(X;X+Z) &= \max_{p(x)\in\mathbb{F}_P}\min_{p(z)\in\mathbb{F}_P} I(X;X+Z)\\ &= I(X^*;X^*+Z^*)\\ &= \frac{1}{2}\log\left(1+\frac{P}{N}\right).\end{aligned}$$

Problem 2 (Erasure distortion)

The rate distortion function is given by

$$R(D) = \min_{p(\hat{x}|x):\sum p(x,\hat{x})d(x,\hat{x})\le D} I(X;\hat{X}),$$

we proceed by finding the minimizing $p(\hat{x}|x)$. The infinite distortion constrains $p(0|1) = p(1|0) = 0$. By symmetry, $p(E|0) = p(E|1) = \alpha$ and $p(0|0) = p(1|1) = 1-\alpha$.

For this distribution the distortion is $\sum p(x,\hat{x})d(x,\hat{x}) = \alpha \le D$ and $I(X;\hat{X}) = 1-\alpha$ which is minimized for $D=\alpha$. So the rate distortion function is $R(D) = 1-D$ for $0\le D\le 1$, and $R(D) = 0$ for $D > 1$.

To achieve this rate distortion function, we can proceed as follows: if D is rational (e.g. $D = \frac{k}{n}$ then we send only $n-k$ of any block of n bits. We reproduce these bits exactly and reproduce the remaining bits as erasures. Hence we can send information at rate $1-D$ and achieve a distortion D. IF D is irrational, we can get arbitrarily close to D by using longer and longer block lengths.

Problem 3 (Convexity of mutual information as a function of $w(y|x)$)

Consider the following chain of inequalities and equalities:

$$\begin{aligned}I(X;Y_\lambda|Z) =& h(X|Z) - h(X|Y_\lambda,Z)\\ \stackrel{(a)}{=}& h(X) - h(X|Y_\lambda,Z)\\ \stackrel{(b)}{\ge}& h(X) - h(X|Y_\lambda)\\ =& I(X;Y_\lambda)\end{aligned}$$

where (a) follows from independence of X and Z, and (b) follows since conditioning cannot increase entropy $(-h(X|Y,Z)\ge -h(X|Y))$. Also, notice that

$$\begin{aligned}I(X;Y_\lambda|Z) =& I(X;Y_\lambda|Z=1)Pr(Z=1) + I(X;Y_\lambda|Z=2)Pr(Z=2)\\ =& I(X;Y_1)\lambda + I(X;Y_2)(1-\lambda).\end{aligned}$$

Hence, $I(X;Y_1)\lambda + I(X;Y_2)(1-\lambda) \ge I(X;Y_\lambda)$, i.e. mutual information is convex in $w(y|x)$.

Problem 4

(a)

$$
\begin{aligned}
I(X^m;\hat{Y}^m) &= h(X^m) - h(X^m|\hat{X}^m) && (3)\\
&= \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i|X^{i-1},\hat{X}^m) && (4)\\
&\geq \sum_{i} h(X_i) - \sum_{i} h(X_i|\hat{X}_i) \quad \text{Follows by the fact that conditioning reduces entropy} && (5)\\
&= \sum_{i} I(X_i;\hat{X}_i) && (6)\\
&\geq \sum_{i} \min_{f(\hat{x}_i|x_i):\mathbb{E}(\hat{X}_i-X_i)^2\leq D_i} I(X_i;\hat{X}_i) && (7)\\
&= R(D_i) \quad \text{by definition} && (8)\\
&= \sum_{i=1}^{m}\left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^{+}. \quad \text{As you have derived in class, where } D_i = \mathbb{E}(\hat{X}_i - X_i)^2. && (9)
\end{aligned}
$$

(b) (5) is tight when $f(x^m|\hat{x}^m) = \prod_{i=1}^{m} f(x_i|\hat{x}_i)$.
(7) is tight for $f(x_i|\hat{x}_i)$ that achieves $I(X_i;\hat{X}_i) = (\frac{1}{2}\log\frac{\sigma_i^2}{D_i})^{+}$. For $D_i \leq \sigma_i$, choose $f(\hat{x}_i, x_i)$ such that $X_i = \hat{X}_i + Z_i$, $\hat{X}_i, Z_i$ being independent and each of $\hat{X}_i \sim \mathcal{N}(0, \sigma_i^2 - D_i)$ and $Z_i \sim \mathcal{N}(0, D_i)$. Thus $I(X_i;\hat{X}_i) = \frac{1}{2}\log\frac{\sigma_i^2}{D_i}$. for $D_i \geq \sigma_i$, choose $\hat{X}_i = 0$ with probability 1.

(c) So finally,

$$R(D) = \min_{f(\hat{x}^m|x^m):\mathbb{E}d(X^m,\hat{X}^m)\leq D} I(X^m;\hat{X}^m),$$

where $\mathbb{E}d(X^m,\hat{X}^m) = \mathbb{E}\sum_i(\hat{X}_i - X_i)^2 = \sum_i \mathbb{E}(\hat{X}_i - X_i)^2$.
For $f(x^m|\hat{x}^m)$ that you found in part(b),

$$I(X^m;\hat{Y}^m) = \sum_{i=1}^{m}\left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^{+},$$

with $D_i = \mathbb{E}(\hat{X}_i - X_i)^2$. Thus the rate distortion function can be reduced to the following optimization problem:

$$R(D) = \min_{\sum_i D_i \leq D} \sum_{i=1}^{m}\left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^{+}.$$

(d) Let us work with ln rather than log in this part. Using Lagrange multipliers, we construct the functional

$$J(D) = \sum_{i=1}^{m}\frac{1}{2}\ln\frac{\sigma_i^2}{D_i} + \lambda\sum_{i=1}^{m} D_i.$$

Differentiating with respect to D_i and setting equal to 0,

$$\frac{\partial J}{\partial D_i} = -\frac{1}{2D_i} + \lambda = 0 \Longrightarrow D_i = \lambda'.$$

Thus the optimum allotment of bits to the various desscriptions results in an equal distortion for each random variable as long as $\lambda' \leq \sigma_i^2$ for all i. As the total allowable

distortion D increases, the constant λ' increases until it exceeds σ_i^2 for some i. At this point, the optimum solution is on the boundary of the allowable region of distortion and Kuhn-Tucker conditions should give the answer:
λ is chosen such that

$$\frac{\partial J}{\partial D_i} = -\frac{1}{2D_i} + \lambda \begin{cases} = 0 & \text{if } D_i < \sigma_i^2 \\ \leq 0 & \text{if } D_i \geq \sigma_i^2 \end{cases}$$

Now we check that

$$D_i = \begin{cases} \lambda' & \text{if } \lambda' < \sigma_i^2 \\ \sigma_i^2 & \text{if } \lambda' \geq \sigma_i^2 \end{cases}$$

in fact satisfy K-T conditions: For $D_i = \lambda' \leq \sigma_i^2$, $-\frac{1}{2D_i} + \lambda = 0$ as required. For $D_i = \sigma_i^2 \leq \lambda'$, $-\frac{1}{2D_i} + \lambda \leq -\frac{1}{2\lambda'} + \lambda \leq 0$ as required again. λ' is chosen so that $\sum_i D_i = D$. By abuse of notation, λ' is λ of part (d) of the homework sheet.