10.2) Since $d(0,1)=\infty$ and $p_X(x)=\begin{cases} 1/2 & x=0 \\ 1/2 & x=1 \end{cases}$

we must have $P_{\hat{X}|X}(1|0)=0$ otherwise

$E[d(X,\hat{X})]=\infty$.

$\Rightarrow P_{\hat{X}|X}(1|0)=0, \quad P_{\hat{X}|X}(0|0)=1$

this leaves $P_{\hat{X}|X}(\hat{x}|1)$ as free.

Let

$$P_{\hat{X}|X}(\hat{x}|1)=\begin{cases} \varepsilon & \hat{x}=0 \\ 1-\varepsilon & \hat{x}=1 \end{cases}$$

Then $E[d(X,\hat{X})]=\frac{1}{2}\times 0+\frac{1}{2}(\varepsilon)\cdot 1$

$=\varepsilon/2$

$$\Rightarrow R(D)=\min_{0\le \varepsilon/2\le D} I(X;\hat{X})$$

Since $P_{\hat{X}|X}(0|1)=\varepsilon$, $\Rightarrow 0\le\varepsilon\le 1$

Now $I(X;\hat{X})$ is decreasing in ε

$$\Rightarrow R(D) = I(X;\hat{X})\Big|_{\frac{\varepsilon}{2}=D}$$

$$= 1 - \tfrac{1}{2}(1-2D)\log(1-2D) + 2D\log 2D$$

for $0 \le \varepsilon \le 1$ or equivalently $0 \le D \le 1/2$.

For $D \ge 1/2$, we can achieve $R(D) = 0$ by choosing $\hat{X} = 0$ since then

$$E[d(X,\hat{X})] = \tfrac{1}{2}d(0,0) + \tfrac{1}{2}d(1,0)$$

$$= 0 + \tfrac{1}{2}1$$

$$= 1/2$$

Now $$p_{\hat{X}}(\hat{x}) = \frac{1}{2} p_{\hat{X}|X}(\hat{x}|0) + \frac{1}{2} p_{\hat{X}|X}(\hat{x}|1)$$

$$= \begin{cases} \frac{1}{2} \times 1 + \frac{1}{2}\varepsilon & \hat{x} = 0 \\ \frac{1}{2} \times 0 + \frac{1}{2}(1-\varepsilon) & \hat{x} = 1 \end{cases}$$

$$\Rightarrow H(\hat{X}) = -\frac{1}{2}(1+\varepsilon)\log\left[\frac{1}{2}(1+\varepsilon)\right] - \frac{1}{2}(1-\varepsilon)\log\left[\frac{1}{2}(1-\varepsilon)\right]$$

and

$$H(\hat{X}|X) = \frac{1}{2} H(\hat{X}|X=0) + \frac{1}{2} H(\hat{X}|X=1)$$
$$= \frac{1}{2} h(0) + \frac{1}{2} h(\varepsilon)$$
$$= -\frac{1}{2}\varepsilon \log \varepsilon - \frac{1}{2}(1-\varepsilon)\log(1-\varepsilon)$$

$$\Rightarrow I(X;\hat{X}) = H(\hat{X}) - H(\hat{X}|X)$$
$$= -\frac{1}{2}(1+\varepsilon)\log\left[\frac{1}{2}(1+\varepsilon)\right] - \frac{1}{2}(1-\varepsilon)\log\left[\frac{1}{2}(1-\varepsilon)\right]$$
$$+ \varepsilon \log \varepsilon + \frac{1}{2}(1-\varepsilon)\log(1-\varepsilon)$$
$$= \frac{1}{2}(1+\varepsilon) - \frac{1}{2}(1+\varepsilon)\log(1+\varepsilon)$$
$$+ \frac{1}{2}(1-\varepsilon) - \frac{1}{2}(1-\varepsilon)\log(1-\varepsilon)$$
$$+ \varepsilon \log \varepsilon + \frac{1}{2}(1-\varepsilon)\log(1-\varepsilon)$$
$$= 1 - \frac{1}{2}(1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log \varepsilon.$$

10.4)

$$R'(D) = \min_{p_{\hat{X}|X}} I(p_X ; p_{\hat{X}|X})$$
$$\text{s.t. } E[d'(X,\hat{X})] \le D$$

$$= \min_{p_{\hat{X}|X}} I(p_X ; p_{\hat{X}|X})$$
$$\text{s.t. } E[d(X,\hat{X}) - w_X] \le D$$

$$= \min_{p_{\hat{X}|X}} I(p_X ; p_{\hat{X}|X})$$
$$\text{s.t. } E[d(X,\hat{X})] - \bar{w} \le D$$

$$= \min_{p_{\hat{X}|X}} I(p_X ; p_{\hat{X}|X})$$
$$\text{s.t. } E[d(X,\hat{X})] \le D + \bar{w}$$

$$= R(D + \bar{w})$$

10.5) Let $X \oplus \hat{X} = X + \hat{X} \mod m$

$Z = d(X, \hat{X})$

$P[Z=1] \triangleq a \le D$

Then
$$
\begin{aligned}
I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\
&= H(X) - H(X \oplus \hat{X} | \hat{X}) \\
&\geq H(X) - H(X \oplus \hat{X}, Z | \hat{X}) \\
&\geq H(X) - H(X \oplus \hat{X}, Z) \\
&= H(X) - H(Z) - H(X \oplus \hat{X} | Z) \\
&= H(X) - H(Z) - H(X \oplus \hat{X} | Z=0]\, P[Z=0] \\
&\quad - H(X \oplus \hat{X} | Z=1]\, P[Z=1]
\end{aligned}
$$

and $H(Z) = h(a)$

$H(X \oplus \hat{X} | Z=0] = 0$ since $X \oplus \hat{X} = 0$ when $Z=0$

$H(X \oplus \hat{X} | Z=1] \le \log m-1$ since $X \oplus \hat{X} \in \{1, \ldots, m-1\}$ when $Z=1$

$$
\Rightarrow I(X;\hat{X}) \geq \log m - h(a) - a \log(m-1) \triangleq g(a)
$$

and $$R(D) = \min_{\substack{P_{X|\hat{X}} \\ \text{s.t. } E[d(X,\hat{X})] \le D}} I(X;\hat{X})$$

$$\ge \min_{0 \le a \le D} g(a)$$

Since $g(a)$ is decreasing in a when $0 \le a \le 1-\frac{1}{m}$ and $g(1-\frac{1}{m}) = 0$,

$$R(D) \ge \begin{cases} \log m - h(D) - D \log m-1 & 0 \le D \le 1-\frac{1}{m} \\ 0 & 1-\frac{1}{m} \le D \le 1 \end{cases}$$

To show that this can be achieved:

1) For $1-\frac{1}{m} \le D \le 1$, pick $\hat{X} = 1$ which requires no information to be transmitted ($R = 0$).

Then $E[d(X, 1)] = P[X \neq 1] = 1 - \frac{1}{m} \le D$.

2) For $0 \le D \le 1 - \frac{1}{m}$,

Consider $p_{\hat{X}}(\hat{x}) = 1/m$

$$p_{X|\hat{X}}(x|\hat{x}) = \begin{cases} 1-D & x = \hat{x} \\ \frac{D}{m-1} & x \neq \hat{x} \end{cases}$$

Then $\sum_{\hat{x}} p_{\hat{X}}(\hat{x})\, p_{X|\hat{X}}(x|\hat{x}) = 1/m$ and is consistent with $p_X(x) = 1/m$.

Also, $H(X|\hat{X}) = -(1-D)\log(1-D) - (m-1)\frac{D}{(m-1)}\log\frac{D}{(m-1)}$

$$= -(1-D)\log(1-D) - D\log D + D\log(m-1)$$

$$= h(D) + D\log(m-1)$$

and $I(X;\hat{X}) = H(X) - H(X|\hat{X})$

$$= \log m - h(D) - D\log(m-1)$$

10.6)

a) Let $\mathcal{C}_0 = \mathcal{C}(D_0)$ with optimal $\mathbf{p}_0$

Let $\mathcal{C}_1 = \mathcal{C}(D_1)$ with optimal $\mathbf{p}_1$

Let $D_\lambda = (1-\lambda) D_0 + \lambda D_1$

Let $\mathbf{p}_\lambda = (1-\lambda) \mathbf{p}_0 + \lambda \mathbf{p}_1$

Then
$$\sum_{i=1}^{m} p_\lambda(i) d_i = \sum_{i=1}^{m} \left[(1-\lambda) p_0(i) + \lambda p_1(i)\right] d_i$$
$$= (1-\lambda) \sum_{i=1}^{m} p_0(i) d_i + \lambda \sum_{i=1}^{m} p_1(i) d_i$$
$$\leq (1-\lambda) D_0 + \lambda D_1$$
$$= D_\lambda$$

So
$$\mathcal{C}(D_\lambda) = \max_{\mathbf{p} : \sum_i p(i) d_i \leq D_\lambda} h(\mathbf{p})$$
$$\geq h(\mathbf{p}_\lambda)$$
$$= h((1-\lambda)\mathbf{p}_0 + \lambda \mathbf{p}_1)$$
$$\geq (1-\lambda) h(\mathbf{p}_0) + \lambda h(\mathbf{p}_1)$$
$$= (1-\lambda) \mathcal{C}(D_0) + \lambda \mathcal{C}(D_1)$$

b) $D_{\hat{x}} = \sum_{x} P_{X|\hat{X}=\hat{x}}(x|\hat{x}) d(x, \hat{x})$

<u>10.154</u>: $H(X|\hat{X}=\hat{x}) = H(P_{X|\hat{X}=\hat{x}})$

$$\le \max_{P_{X|\hat{X}=\hat{x}}} H(P_{X|\hat{X}=\hat{x}})$$

$$\text{s.t.} \sum_{x} P_{X|\hat{X}=\hat{x}}(x|\hat{x}) d(x,\hat{x}) \le D_{\hat{x}}$$

$$= \varphi(D_{\hat{x}})$$

<u>10.155</u>: Since $\varphi(D)$ is concave in D,

$$\sum P_{\hat{X}}(\hat{x}) \varphi(D_{\hat{x}}) \le \varphi\left(\sum_{\hat{x}} P_{\hat{X}}(\hat{x}) D_{\hat{x}}\right)$$

<u>10.156</u>:

$$\sum_{x,\hat{x}} P_X(\hat{x}) P_{X|\hat{X}}(x|\hat{x}) d(x,\hat{x})$$

$$= \sum P_X(\hat{x}) D_x \quad \le D$$

and since $\varphi(D)$ is increasing in D:

$$\varphi\left(\sum_{\hat{x}} P_X(\hat{x}) D_{\hat{x}}\right) \le \varphi(D)$$

10.6 c) $$R(D) = \min_{p_{\hat{X}|X}} I(X;\hat{X})$$
$$\text{s.t. } E[d(X,\hat{X})] \le D$$

However, in (b) it was shown that if $E[d(X,\hat{X})] \le D$ then $I(X;\hat{X}) \ge H(X) - \phi(D)$, and this is true regardless of $p_{X\hat{X}}$, provided only $E[d(X,\hat{X})] \le D$.

$\Rightarrow \quad R(D) \ge H(X) - \phi(D).$

10.7)

$$d(x,\hat{x}) = \begin{cases} 0 & \hat{x} = x \\ 1 & \hat{x} = \text{"}\varepsilon\text{"} \\ \infty & \hat{x} = 1-x \end{cases}$$

For $E[d(X,\hat{X})]$ to be finite, must have

$$P_{\hat{X}|X}(1|0) = 0 \qquad P_{\hat{X}|X}(0|1) = 0$$

$$\Rightarrow \quad P_{\hat{X}|X}(0|0) = 1-a \qquad P_{\hat{X}|X}(1|1) = 1-b$$

$$P_{\hat{X}|X}(\varepsilon|0) = a \qquad P_{\hat{X}|X}(\varepsilon|1) = b$$

for some $a, b \geq 0$.

Since $$P_{X|\hat{X}}(x|\hat{x}) = \frac{P_{\hat{X}|X}(\hat{x}|x)\, P_X(x)}{\sum_{\hat{x}'} P_{\hat{X}|X}(\hat{x}'|x)\, P_X(x)}$$

and $P_X(x) = 1/2$

$$\Rightarrow \quad P_{X|\hat{X}}(0|0) = 1 \qquad P_{X|\hat{X}}(0|1) = 0$$

$$P_{X|\hat{X}}(0|\varepsilon) = \frac{a}{a+b} \qquad P_{X|\hat{X}}(1|\varepsilon) = \frac{b}{a+b}$$

$$P_{X|\hat{X}}(1|0) = 0 \qquad P_{X|\hat{X}}(1|1) = 1$$

Now $I(X;\hat{X}) = H(X) - H(X|\hat{X})$

$$= 1 - H(X|\hat{X}=0)\, P[\hat{X}=0]$$
$$- H(X|\hat{X}=1)\, P[\hat{X}=1]$$
$$- H(X|\hat{X}=\mathcal{E})\, P[\hat{X}=\mathcal{E}]$$

$$= 1 - h(0)\, P[\hat{X}=0]$$
$$- h(0)\, P[\hat{X}=1]$$
$$- h\left(\frac{a}{a+b}\right)\left(\frac{a}{2}+\frac{b}{2}\right)$$

$$\geq 1 - h(1/2)\left(\frac{a}{2}+\frac{b}{2}\right)$$

$$= 1 - \frac{1}{2}(a+b)$$

$$= 1 - P[\hat{X}=\mathcal{E}]$$

$$= 1 - D$$

where $D = E[d(X,\hat{X})] = P[\hat{X}=\mathcal{E}]$

$\Rightarrow R(D) \geq 1-D$

We can achieve $R(D)=1-D$ by transmitting the first $n(1-D)$ binary symbols of a block of length n, and replacing the last nD symbols with erasures $\mathcal{E}$.

10.8) Let $d(x,\hat{x}) = (x-\hat{x})^2$

$$E[d(x,\hat{x})] = E(x-\hat{x})^2 = \sigma_D^2 \le D$$

$$\begin{aligned}
I(x;\hat{x}) &= h(x) - h(x|\hat{x}) \\
&= h(x) - h(x-\hat{x}|\hat{x}) \\
&\ge h(x) - h(x-\hat{x}) \\
&\overset{(a)}{\ge} h(x) - h(z) \qquad \text{where } z \sim N(0,\sigma_D^2) \\
&= h(x) - \tfrac{1}{2}\log(2\pi e \sigma_D^2) \\
&\ge h(x) - \tfrac{1}{2}(2\pi e D) \qquad \text{since } D \ge \sigma_D^2
\end{aligned}$$

where (a) follows since for a given variance σ_D^2, a normal distribution $N(0,\sigma_D^2)$ maximizes differential entropy (Thm 8.6.5).

Now, let $\hat{x} = \frac{\sigma^2 - D}{\sigma^2}(x+z)$ where $z \sim N\left(0, \frac{D\sigma^2}{\sigma^2 - D}\right)$

Then $VAR(\hat{x}) = \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 VAR(x+z)$

$$= \left(\frac{\sigma^2 - D}{\sigma^2}\right)\left(\sigma^2 + \frac{D\sigma^2}{\sigma^2 - D}\right)$$

$$= \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \left(\frac{\sigma^4}{\sigma^2 - D}\right) = \sigma^2 - D$$

$$\Rightarrow \quad h(\hat{X}) \le \tfrac{1}{2}\log\left[2\pi e(\sigma^2 - D)\right]$$

Also, $h(\hat{X} \mid X) = h\left(\frac{\sigma^2 - D}{\sigma^2}(X + Z) \mid X\right)$

$$= h\left(\frac{\sigma^2 - D}{\sigma^2} \cancel{X}\ \Big|\ X\right)$$

$$= h\left(\frac{\sigma^2 - D}{\sigma^2} Z\right)$$

$$= h\left(N\left(0, \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \frac{D\sigma^2}{\sigma^2 - D}\right)\right)$$

$$= h\left(N\left(0, \frac{D(\sigma^2 - D)}{\sigma^2}\right)\right)$$

$$\Rightarrow I(X;\hat{X}) = \tfrac{1}{2}\log\left[\cancel{2\pi e}\cancel{(\sigma^2 - D)}\right] - \tfrac{1}{2}\log\left[\cancel{2\pi e}\left(\frac{D\cancel{(\sigma^2 - D)}}{\sigma^2}\right)\right]$$

$$= \tfrac{1}{2}\log \frac{\sigma^2}{D}$$

10.12) This decreases the R(D) function since we can ignore the new symbol $\hat{x}_0$ and use the previous scheme still (with the previous rate).

Thus, adding the new symbol has the potential to decrease R(D), but will never increase it since it can be ignored.

10.14 a) $I(X,Y;\hat{X},\hat{Y})$

$$= H(X,Y) - H(X,Y|\hat{X},\hat{Y})$$

$$= H(X) + H(Y) - H(X|\hat{X}\hat{Y}) - H(Y|\hat{X}\hat{Y}X)$$

$$\overset{(a)}{\geq} H(X) + H(Y) - H(X|\hat{X}) - H(Y|\hat{Y})$$

$$= I(X;\hat{X}) + I(Y;\hat{Y})$$

$$\Rightarrow R_{XY}(D_1,D_2) = \min_{\substack{P_{\hat{X}\hat{Y}|XY} \\ \text{s.t. } E[d(X,\hat{X})] \leq D_1 \\ E[d(Y,\hat{Y})] \leq D_2}} I(X,Y;\hat{X},\hat{Y})$$

$$\geq \min_{\substack{P_{\hat{X}\hat{Y}|XY} \\ \text{s.t. } E[d(X,\hat{X})] \leq D_1 \\ E[d(X,\hat{Y})] \leq D_2}} I(X;\hat{X}) + I(Y;\hat{Y})$$

$$= \min_{\substack{P_{\hat{X}|X} \\ \text{s.t. } E[d(X,\hat{X})] \leq D_1}} I(X;\hat{X}) + \min_{\substack{P_{\hat{Y}|Y} \\ \text{s.t. } E[d(Y,\hat{Y})] \leq D_2}} I(Y;\hat{Y})$$

$$= R_X(D_1) + R_Y(D_2)$$

b) Equality holds since optimal $P_{\hat{X}|X}$ & $P_{\hat{Y}|Y}$ can be combined into $P_{\hat{X}\hat{Y}|XY} = P_{\hat{X}|X} P_{\hat{Y}|Y}$, and for $P_{\hat{X}\hat{Y}|XY} = P_{\hat{X}|X} P_{\hat{Y}|Y}$, $P_{XY} = P_X P_Y$:

$$H(X|\hat{X}\hat{Y}) = H(X|\hat{X})$$

$$H(Y|\hat{X}\hat{Y}X) = H(Y|\hat{Y}),$$

resulting in equality in step (a) above