8.3 The Rate-Distortion Theorem

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

- The minimization is taken over the set of all transition matrices $p(\hat{x}|x)$ such that $Ed(X,\hat{X}) \leq D$, namely the set

$$\left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \right\}.$$

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

- The minimization is taken over the set of all transition matrices $p(\hat{x}|x)$ such that $Ed(X,\hat{X}) \leq D$, namely the set

$$\left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \right\}.$$

- Since this set is compact (closed and bounded) in $\Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$ and $I(X;\hat{X})$ is a continuous functional of $p(\hat{x}|x)$, the minimum value of $I(X;\hat{X})$ can be attained.

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

- The minimization is taken over the set of all transition matrices $p(\hat{x}|x)$ such that $Ed(X, \hat{X}) \leq D$, namely the set

$$\left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \right\}.$$

- Since this set is compact (closed and bounded) in $\Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$ and $I(X;\hat{X})$ is a continuous functional of $p(\hat{x}|x)$, the minimum value of $I(X;\hat{X})$ can be attained.

- Equivalently, the minimization can be taken over the set of all joint distributions $p(x,\hat{x})$ with marginal distribution $p(x)$, the given source distribution, such that $Ed(X,\hat{X}) \leq D$.

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

- The minimization is taken over the set of all transition matrices $p(\hat{x}|x)$ such that $Ed(X,\hat{X}) \leq D$, namely the set

$$\left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \right\}.$$

- Since

$$E\tilde{d}(X,\hat{X}) = Ed(X,\hat{X}) - \Delta,$$

where Δ does not depend on $p(\hat{x}|x)$, we can always replace d by $\tilde{d}$ and D by $D - \Delta$ in the definition of $R_I(D)$ without changing the minimization problem.

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X} : E\tilde{d}(X,\hat{X}) \leq D - \Delta} I(X;\hat{X}).$$

- The minimization is taken over the set of all transition matrices $p(\hat{x}|x)$ such that $Ed(X,\hat{X}) \leq D$, namely the set

$$\left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \right\}.$$

- Since

$$E\tilde{d}(X,\hat{X}) = Ed(X,\hat{X}) - \Delta,$$

  where Δ does not depend on $p(\hat{x}|x)$, we can always replace d by $\tilde{d}$ and D by $D - \Delta$ in the definition of $R_I(D)$ without changing the minimization problem.

Definition 8.16 For $D \geq 0$, the information rate-distortion function is defined by

$$R_I(D) = \min_{\hat{X} : E\tilde{d}(X,\hat{X}) \leq D - \Delta} I(X;\hat{X}).$$

- The minimization is taken over the set of all transition matrices $p(\hat{x}|x)$ such that $Ed(X,\hat{X}) \leq D$, namely the set

$$\left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \leq D \right\}.$$

- Since

$$E\tilde{d}(X,\hat{X}) = Ed(X,\hat{X}) - \Delta,$$

  where Δ does not depend on $p(\hat{x}|x)$, we can always replace d by $\tilde{d}$ and D by $D - \Delta$ in the definition of $R_I(D)$ without changing the minimization problem.

- Without loss of generality, we can assume d is normal.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Theorem 8.17 (The Rate-Distortion Theorem) $R(D) = R_I(D)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Theorem 8.15 The following properties hold for the rate-distortion function $R(D)$:

1. $R(D)$ is non-increasing in D.
2. $R(D)$ is convex.
3. $R(D) = 0$ for $D \geq D_{max}$.
4. $R(0) \leq H(X)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$Ed(X, \hat{X}^{(\lambda)})$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \end{aligned}$$

$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ & \quad = \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ & \quad \leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ & \quad = D^{(\lambda)}. \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \;& \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \;& \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \;& D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)})$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ &= \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ &\leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ &= D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ &= \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} &Ed(X, \hat{X}^{(\lambda)}) \\ &= \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ &\leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ &= D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} &\lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ &= \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = \quad & \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq \quad & I(X; \hat{X}^{(\lambda)}) \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = \quad & \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq \quad & I(X; \hat{X}^{(\lambda)}) \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ & = \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ & \leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ & = D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ & = \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ & \geq I(X; \hat{X}^{(\lambda)}) \\ & \geq R_I(D^{(\lambda)}). \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = \quad & \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq \quad & I(X; \hat{X}^{(\lambda)}) \\ \geq \quad & R_I(D^{(\lambda)}). \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ & = \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ & \leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ & = D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ & = \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ & \geq I(X; \hat{X}^{(\lambda)}) \\ & \geq R_I(D^{(\lambda)}). \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = & \quad \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq & \quad \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = & \quad D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = & \quad \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq & \quad I(X; \hat{X}^{(\lambda)}) \\ \geq & \quad R_I(D^{(\lambda)}). \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = \quad & \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq \quad & I(X; \hat{X}^{(\lambda)}) \\ \geq \quad & R_I(D^{(\lambda)}). \end{aligned}$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = \quad & \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq \quad & I(X; \hat{X}^{(\lambda)}) \\ \geq \quad & R_I(D^{(\lambda)}). \end{aligned}$$

3. Let $\hat{X} = \hat{x}^*$ w.p. 1 to show that $(0, D_{max})$ is achievable. Then for $D \geq D_{max}$, $R_I(D) \leq I(X; \hat{X}) = 0$, which implies $R_I(D) = 0$.

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ = \quad & \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ \leq \quad & \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ = \quad & D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ = \quad & \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ \geq \quad & I(X; \hat{X}^{(\lambda)}) \\ \geq \quad & R_I(D^{(\lambda)}). \end{aligned}$$

3. Let $\hat{X} = \hat{x}^*$ w.p. 1 to show that $(0, D_{max})$ is achievable. Then for $D \geq D_{max}$, $R_I(D) \leq I(X; \hat{X}) = 0$, which implies $R_I(D) = 0$.

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ &= \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ &\leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ &= D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ &= \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ &\geq I(X; \hat{X}^{(\lambda)}) \\ &\geq R_I(D^{(\lambda)}). \end{aligned}$$

3. Let $\hat{X} = \hat{x}^*$ w.p. 1 to show that $(0, D_{max})$ is achievable. Then for $D \geq D_{max}$, $R_I(D) \leq I(X; \hat{X}) = 0$, which implies $R_I(D) = 0$.

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.

2. $R_I(D)$ is convex.

3. $R_I(D) = 0$ for $D \geq D_{max}$.

4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ & \quad = \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ & \quad \leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ & \quad = D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ & \quad = \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ & \quad \geq I(X; \hat{X}^{(\lambda)}) \\ & \quad \geq R_I(D^{(\lambda)}). \end{aligned}$$

3. Let $\hat{X} = \hat{x}^*$ w.p. 1 to show that $(0, D_{max})$ is achievable. Then for $D \geq D_{max}$, $R_I(D) \leq I(X; \hat{X}) = 0$, which implies $R_I(D) = 0$.

4. Let $\hat{X} = \hat{x}^*(X)$, so that $Ed(X, \hat{X}) = 0$ (since d is normal). Then

$$R_I(0) \leq I(X; \hat{X}) \leq H(X).$$

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Theorem 8.18 The following properties hold for the information rate-distortion function $R_I(D)$:

1. $R_I(D)$ is non-increasing in D.
2. $R_I(D)$ is convex.
3. $R_I(D) = 0$ for $D \geq D_{max}$.
4. $R_I(0) \leq H(X)$.

Proof

1. For a larger D, the minimization is taken over a larger set.

2. Consider any $D^{(1)}, D^{(2)} \geq 0$ and $0 \leq \lambda \leq 1$. Let $\hat{X}^{(i)}$ achieves $R_I(D^{(i)})$ for $i = 1, 2$, i.e.,

$$R_I(D^{(i)}) = I(X; \hat{X}^{(i)}),$$

where

$$Ed(X, \hat{X}^{(i)}) \leq D^{(i)}.$$

Let $\hat{X}^{(\lambda)}$ be jointly distributed with X defined by

$$p_\lambda(\hat{x}|x) = \lambda p_1(\hat{x}|x) + \bar{\lambda} p_2(\hat{x}|x).$$

Then

$$\begin{aligned} & Ed(X, \hat{X}^{(\lambda)}) \\ &= \lambda Ed(X, \hat{X}^{(1)}) + \bar{\lambda} Ed(X, \hat{X}^{(2)}) \\ &\leq \lambda D^{(1)} + \bar{\lambda} D^{(2)} \\ &= D^{(\lambda)}. \end{aligned}$$

Finally consider

$$\begin{aligned} & \lambda R_I(D^{(1)}) + \bar{\lambda} R_I(D^{(2)}) \\ &= \lambda I(X; \hat{X}^{(1)}) + \bar{\lambda} I(X; \hat{X}^{(2)}) \\ &\geq I(X; \hat{X}^{(\lambda)}) \\ &\geq R_I(D^{(\lambda)}). \end{aligned}$$

3. Let $\hat{X} = \hat{x}^*$ w.p. 1 to show that $(0, D_{max})$ is achievable. Then for $D \geq D_{max}$, $R_I(D) \leq I(X; \hat{X}) = 0$, which implies $R_I(D) = 0$.

4. Let $\hat{X} = \hat{x}^*(X)$, so that $Ed(X, \hat{X}) = 0$ (since d is normal). Then

$$R_I(0) \leq I(X; \hat{X}) \leq H(X).$$

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}).$$

Example 3.13 (Convexity of Mutual Information) For fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$D' \quad \ge \quad Ed(X, \hat{X})$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) \underline{p(\hat{x})} d(x,\hat{x}) \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) \underline{p(\hat{x})} d(x, \hat{x}) \\ &= \sum_{\hat{x}} \underline{p(\hat{x})} \sum_x p(x) d(x, \hat{x}) \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \underline{\sum_x p(x) d(x, \hat{x})}
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \underline{Ed(X, \hat{x})} \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \underline{Ed(X, \hat{x})} \\
&\geq \sum_{\hat{x}} p(\hat{x}) \underline{Ed(X, \hat{x}^*)}
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) \underline{Ed(X, \hat{x}^*)} \\
&= \sum_{\hat{x}} p(\hat{x}) \underline{D_{max}}
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\ &\geq \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x})d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x})d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x)d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x})Ed(X,\hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x})Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x})D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \leq D \leq D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

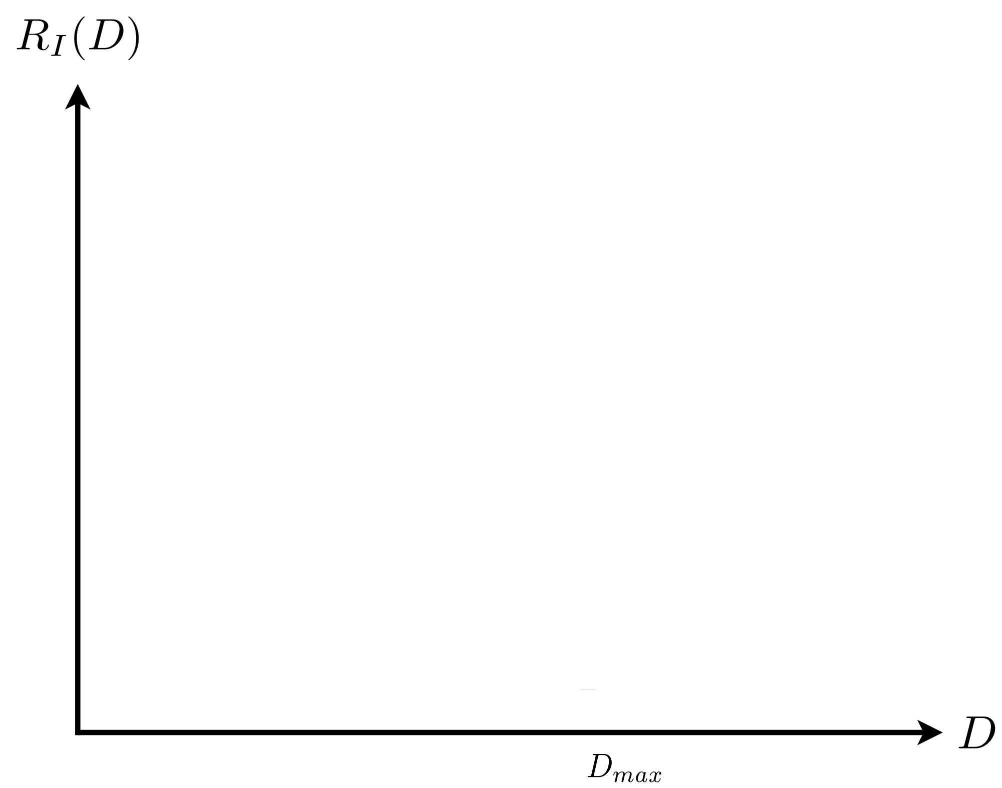
b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \leq D \leq D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.



Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

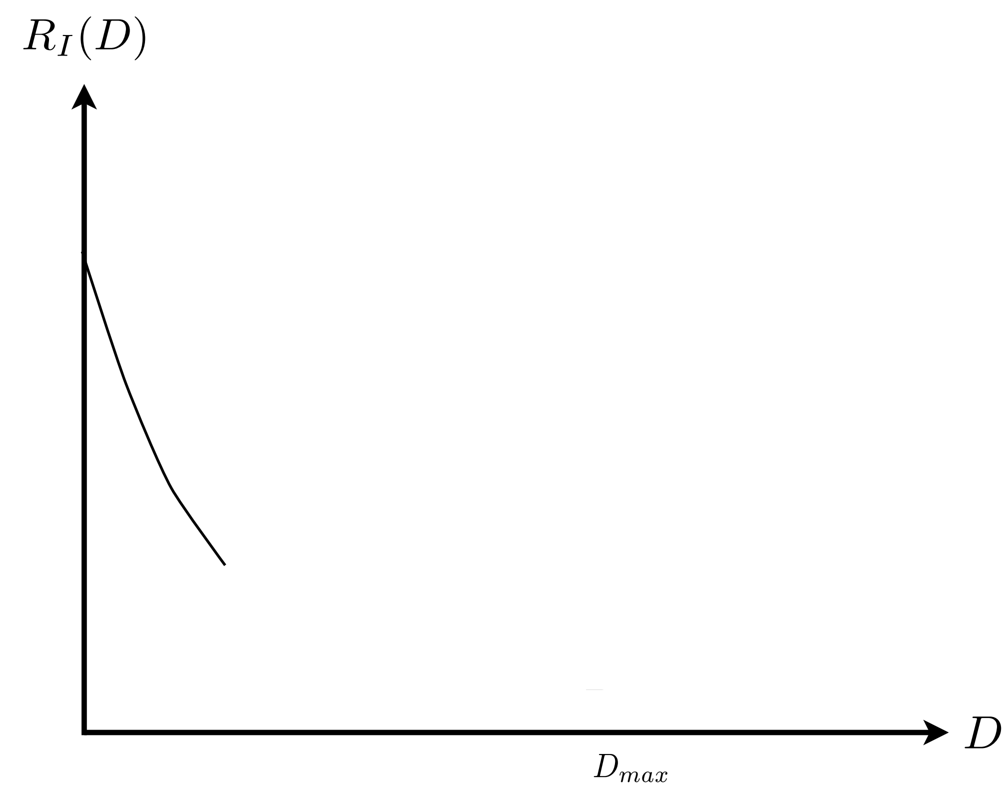
b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.



Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

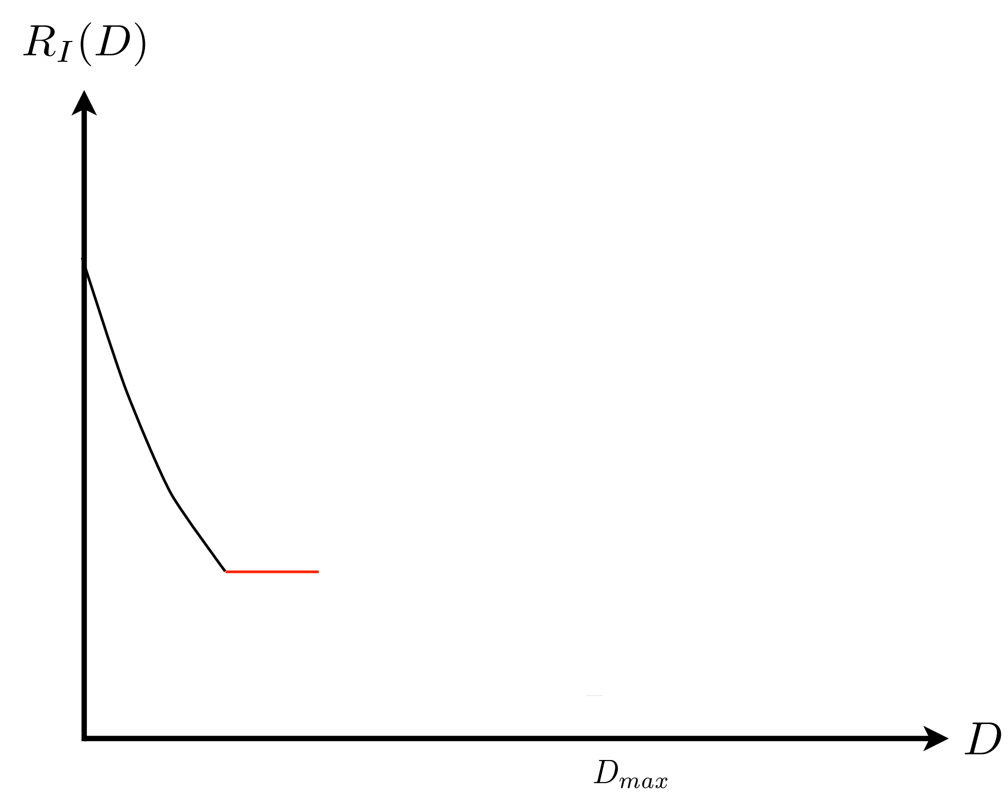
b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.



Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

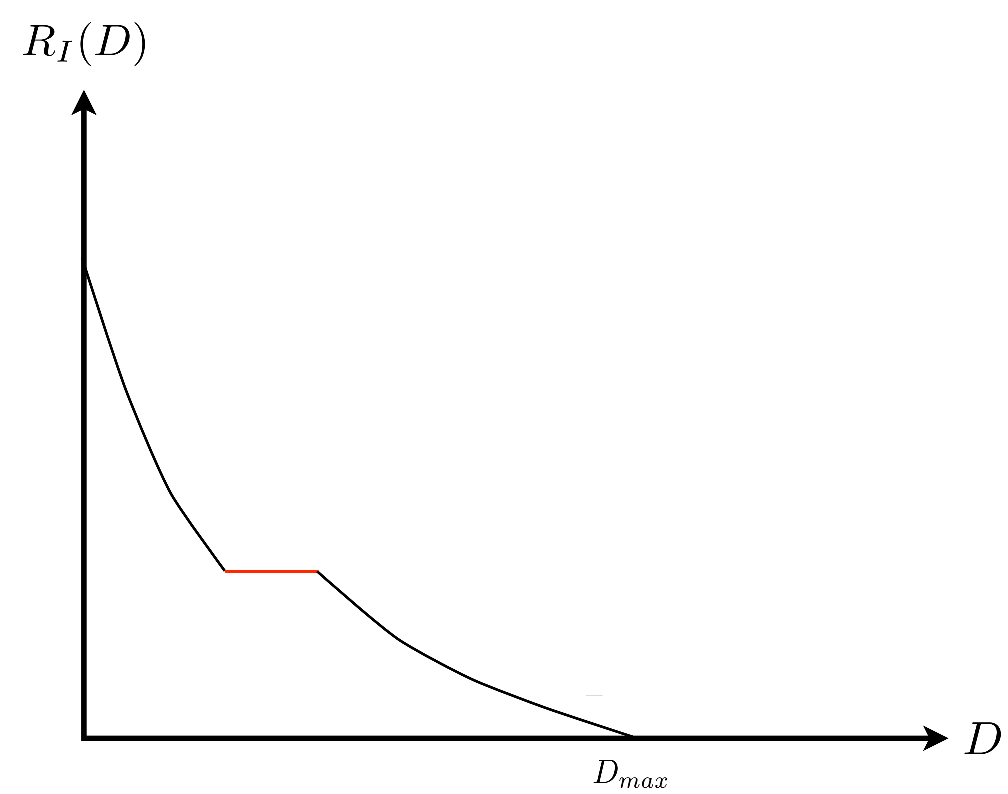
b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.



Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

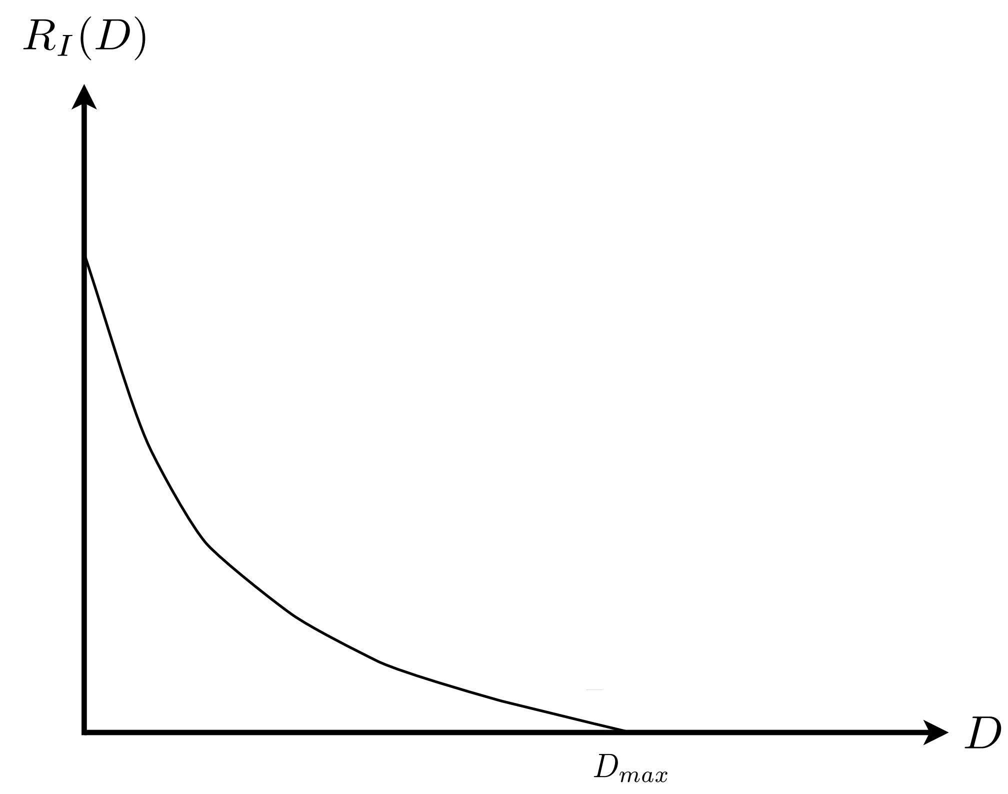
b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.



Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

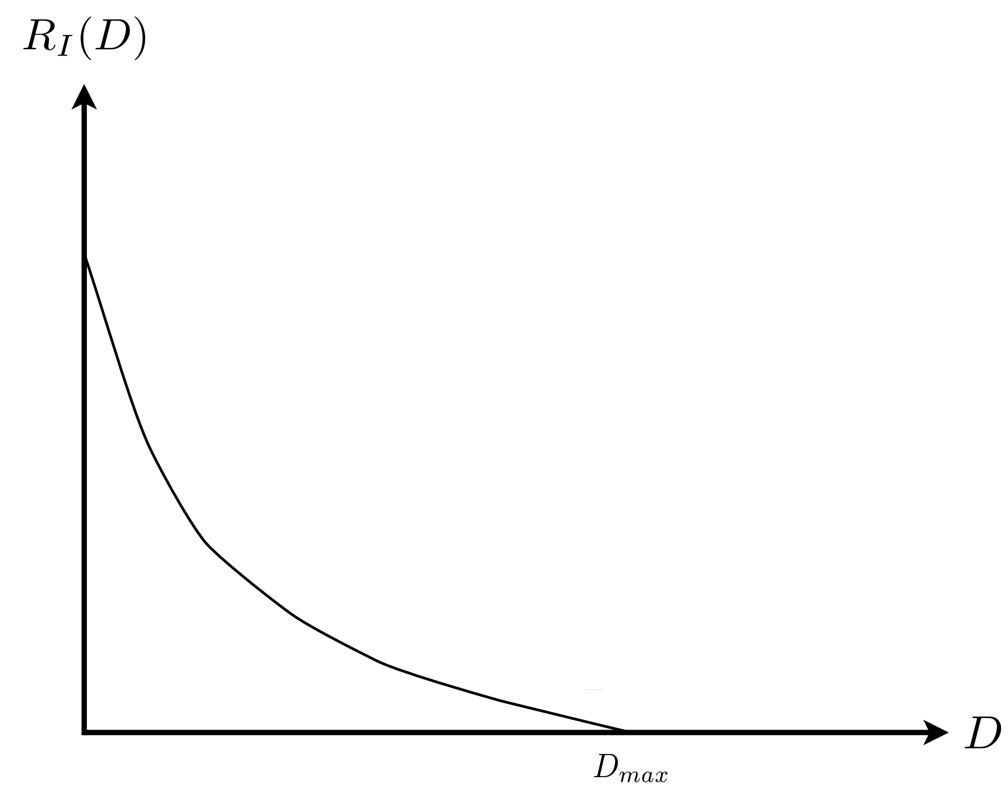
b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.



Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X, \hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \geq Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\geq Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\ &\geq \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \leq D \leq D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X, \hat{X}^*) = D'' < D$.

b. Then

$$R_I(D'') = \min_{\hat{X}: Ed(X, \hat{X}) \leq D''} I(X; \hat{X})$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned}
R_I(D'') &= \min_{\hat{X}: Ed(X,\hat{X}) \le D''} I(X;\hat{X}) \\
&\le I(X;\hat{X}^*)
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned}
R_I(D'') &= \min_{\hat{X}: Ed(X,\hat{X}) \le D''} I(X;\hat{X}) \\
&\le I(X;\hat{X}^*)
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\ge Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned}
R_I(D'') &= \min_{\hat{X}:Ed(X,\hat{X}) \le D''} I(X;\hat{X}) \\
&\le I(X;\hat{X}^*) \\
&= R_I(D),
\end{aligned}$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X, \hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned} R_I(D'') &= \min_{\hat{X}: Ed(X, \hat{X}) \le D''} I(X; \hat{X}) \\ &\le I(X; \hat{X}^*) \\ &= R_I(D), \end{aligned}$$

a contradiction because $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned} R_I(D'') &= \min_{\hat{X}:Ed(X,\hat{X}) \le D''} I(X;\hat{X}) \\ &\le I(X;\hat{X}^*) \\ &= R_I(D), \end{aligned}$$

a contradiction because $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X; \hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X, \hat{X}) \ge Ed(X, \hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X, \hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x) p(\hat{x}) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x, \hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X, \hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X, \hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned} R_I(D'') &= \min_{\hat{X}: Ed(X, \hat{X}) \le D''} I(X; \hat{X}) \\ &\le I(X; \hat{X}^*) \\ &= R_I(D), \end{aligned}$$

a contradiction because $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \le D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \le D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \ge Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned} D' &\ge Ed(X,\hat{X}) \\ &= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\ &= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\ &= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\ &\ge \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\ &= \sum_{\hat{x}} p(\hat{x}) D_{max} \\ &= D_{max}. \end{aligned}$$

c. This is a contradiction because

$$0 \le D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \le D \le D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned} R_I(D'') &= \min_{\hat{X}: Ed(X,\hat{X}) \le D''} I(X;\hat{X}) \\ &\le I(X;\hat{X}^*) \\ &= R_I(D), \end{aligned}$$

a contradiction because $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

c. Therefore, $Ed(X,\hat{X}^*) = D$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x}) d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x}) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x) d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \leq D \leq D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned}
R_I(D'') &= \min_{\hat{X}:Ed(X,\hat{X}) \leq D''} I(X;\hat{X}) \\
&\leq I(X;\hat{X}^*) \\
&= R_I(D),
\end{aligned}$$

a contradiction because $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

c. Therefore, $Ed(X,\hat{X}^*) = D$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Proof

1. Show that $R_I(D) > 0$ for $0 \leq D < D_{max}$ by contradiction.

a. Suppose $R_I(D') = 0$ for some $0 \leq D' < D_{max}$, and let $R_I(D')$ be achieved by some $\hat{X}$. Then

$$R_I(D') = I(X;\hat{X}) = 0$$

implies that X and $\hat{X}$ are independent.

b. Show that such an $\hat{X}$ which is independent of X cannot do better than the constant estimate $\hat{x}^*$, i.e., $Ed(X,\hat{X}) \geq Ed(X,\hat{x}^*) = D_{max}$. This is done by considering

$$\begin{aligned}
D' &\geq Ed(X,\hat{X}) \\
&= \sum_x \sum_{\hat{x}} p(x,\hat{x})d(x,\hat{x}) \\
&= \sum_x \sum_{\hat{x}} p(x)p(\hat{x})d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) \sum_x p(x)d(x,\hat{x}) \\
&= \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}) \\
&\geq \sum_{\hat{x}} p(\hat{x}) Ed(X,\hat{x}^*) \\
&= \sum_{\hat{x}} p(\hat{x}) D_{max} \\
&= D_{max}.
\end{aligned}$$

c. This is a contradiction because

$$0 \leq D' < D_{max}.$$

2. $R_I(D)$ must be strictly decreasing for $0 \leq D \leq D_{max}$ because $R_I(0) > 0$, $R_I(D_{max}) = 0$, and $R_I(D)$ is non-increasing and convex.

3. Show that the inequality constraints in $R_I(D)$ can be replaced by an equality constraint by contradiction.

a. Assume that $R_I(D)$ is achieved by some $\hat{X}^*$ such that $Ed(X,\hat{X}^*) = D'' < D$.

b. Then

$$\begin{aligned}
R_I(D'') &= \min_{\hat{X}:Ed(X,\hat{X}) \leq D''} I(X;\hat{X}) \\
&\leq I(X;\hat{X}^*) \\
&= R_I(D),
\end{aligned}$$

a contradiction because $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

c. Therefore, $Ed(X,\hat{X}^*) = D$.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Remark In all problems of interest,

$$R(0) = R_I(0) > 0.$$

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \le D \le D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Remark In all problems of interest,

$$R(0) = R_I(0) > 0.$$

Otherwise, $R(D) = 0$ for all $D \ge 0$ because $R(D)$ is nonnegative and non-increasing. Therefore,

Corollary 8.19 If $R_I(0) > 0$, then $R_I(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$, and the inequality constraint in the definition of $R_I(D)$ can be replaced by an equality constraint.

Remark In all problems of interest,

$$R(0) = R_I(0) > 0.$$

Otherwise, $R(D) = 0$ for all $D \geq 0$ because $R(D)$ is nonnegative and non-increasing. Therefore,

$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})=D} I(X;\hat{X}).$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

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

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

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

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - \underline{H(X|\hat{X})} \\ &= h_b(\gamma) - H(Y|\hat{X}) \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - \underline{H(X|\hat{X})} \\ &= h_b(\gamma) - \underline{H(Y|\hat{X})} \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - \underline{H(Y|\hat{X})} \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - \underline{H(Y|\hat{X})} \\ &\geq h_b(\gamma) - \underline{H(Y)} \qquad (1) \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - \underline{h_b(\Pr\{X \ne \hat{X}\})} \\ &\ge h_b(\gamma) - \underline{h_b(D)}, \qquad (2) \end{aligned}$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \quad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \quad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) && (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \\ &\geq h_b(\gamma) - h_b(D), && (2) \end{aligned}$$

because $\Pr\{X \neq \hat{X}\} = Ed(X, \hat{X}) \leq D$ and $h_b(a)$ is increasing for $0 \leq a \leq \frac{1}{2}$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) & (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \\ &\geq h_b(\gamma) - h_b(D), & (2) \end{aligned}$$

because $\Pr\{X \neq \hat{X}\} = Ed(X, \hat{X}) \leq D$ and $h_b(a)$ is increasing for $0 \leq a \leq \frac{1}{2}$.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \\ &\geq h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \neq \hat{X}\} = Ed(X, \hat{X}) \leq D$ and $h_b(a)$ is increasing for $0 \leq a \leq \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \leq D} I(X; \hat{X}) \geq h_b(\gamma) - h_b(D).$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \\ &\geq h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \neq \hat{X}\} = Ed(X, \hat{X}) \leq D$ and $h_b(a)$ is increasing for $0 \leq a \leq \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}) \geq h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \quad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \quad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \; 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \quad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \quad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - \underline{H(Y|\hat{X})} \\ &\ge h_b(\gamma) - \underline{H(Y)} \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - \underline{H(Y|\hat{X})} \\ &\ge h_b(\gamma) - \underline{H(Y)} \qquad (1) \\ &= h_b(\gamma) - h_b(\underline{\Pr\{X \ne \hat{X}\}}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - \underline{H(Y|\hat{X})} \\ &\ge h_b(\gamma) - \underline{H(Y)} \qquad (1) \\ &= h_b(\gamma) - h_b(\underline{\Pr\{X \ne \hat{X}\}}) \\ &\ge h_b(\gamma) - h_b(\underline{D}), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) \quad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \\ &\geq h_b(\gamma) - h_b(D), \quad (2) \end{aligned}$$

because $\Pr\{X \neq \hat{X}\} = Ed(X, \hat{X}) \leq D$ and $h_b(a)$ is increasing for $0 \leq a \leq \frac{1}{2}$.

5. Therefore,

$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X; \hat{X}) \geq h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

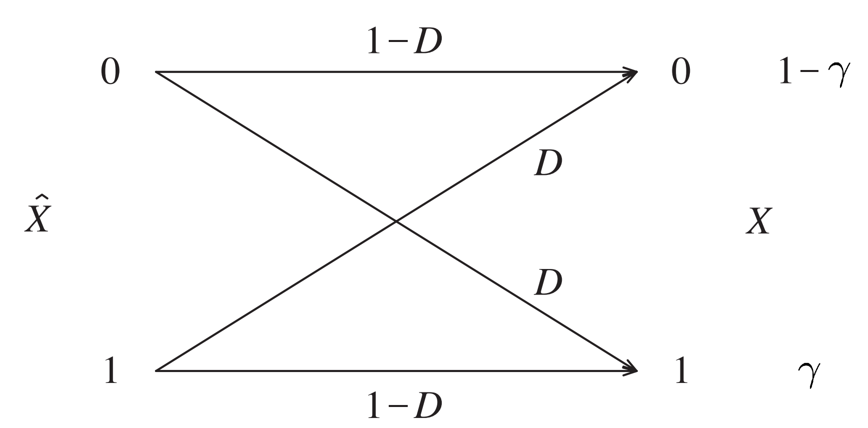
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) && (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), && (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

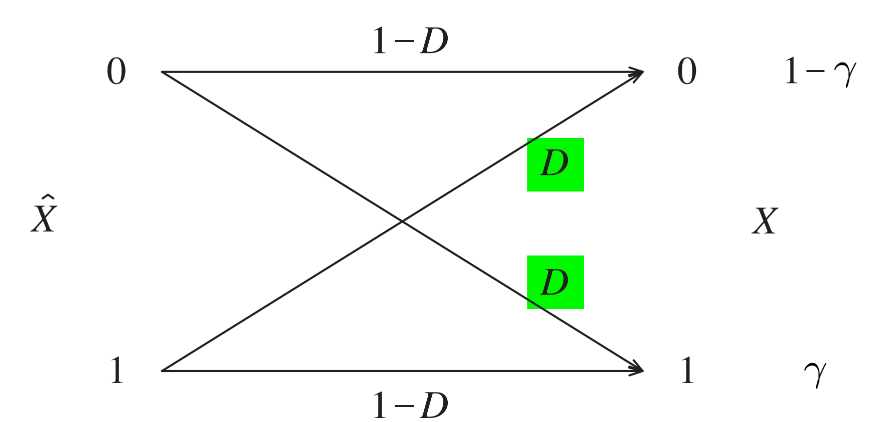
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

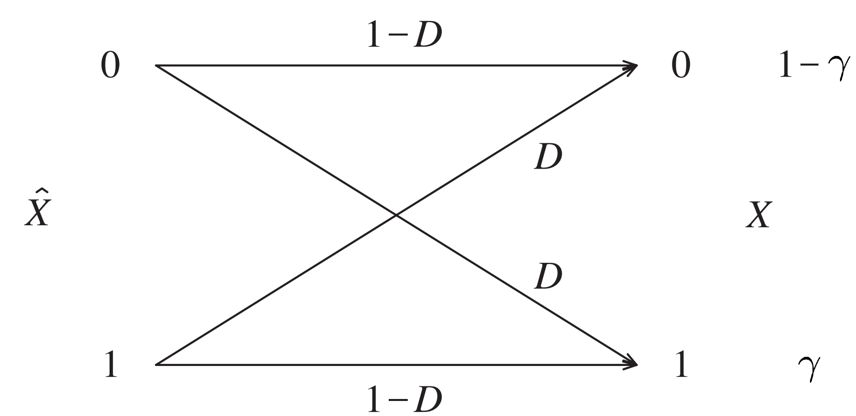
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

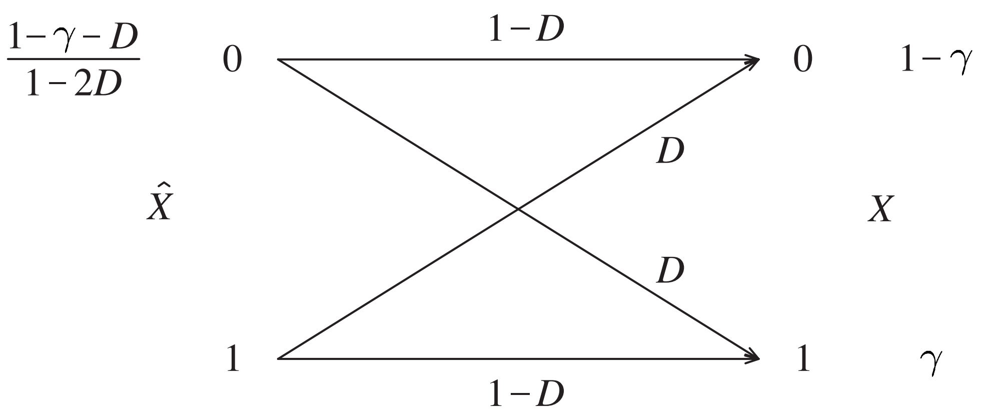
$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

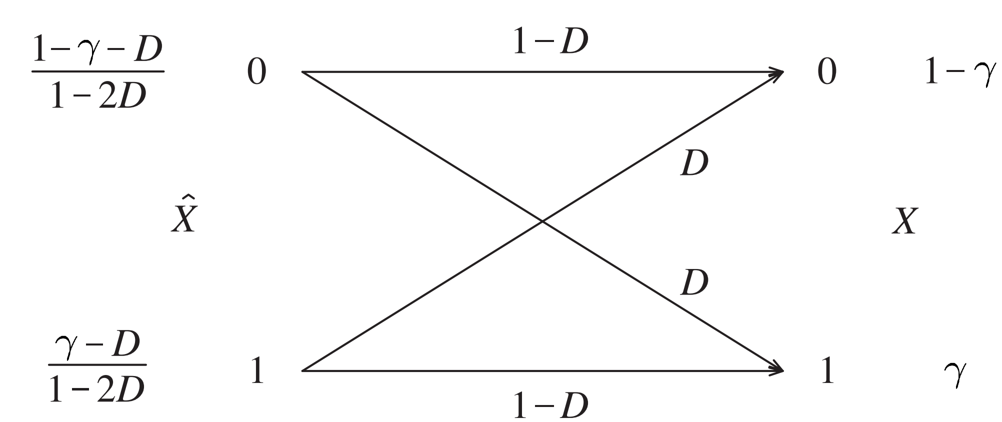
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \quad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \quad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

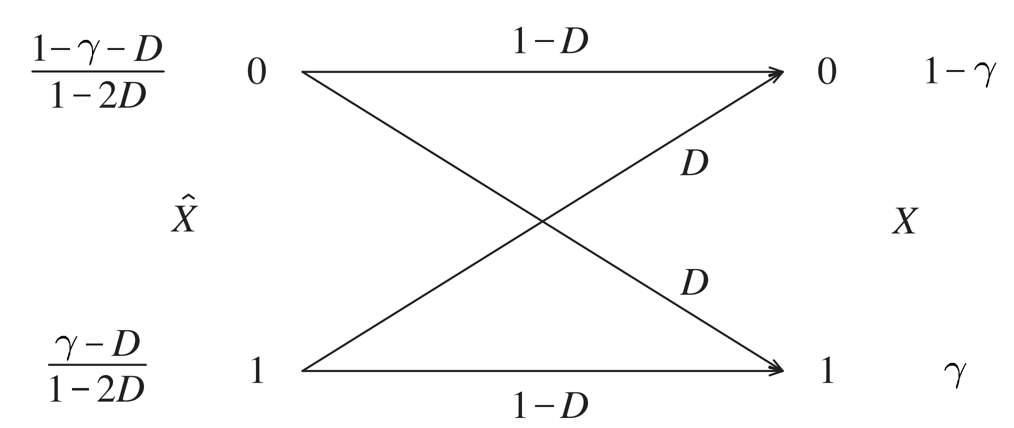
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

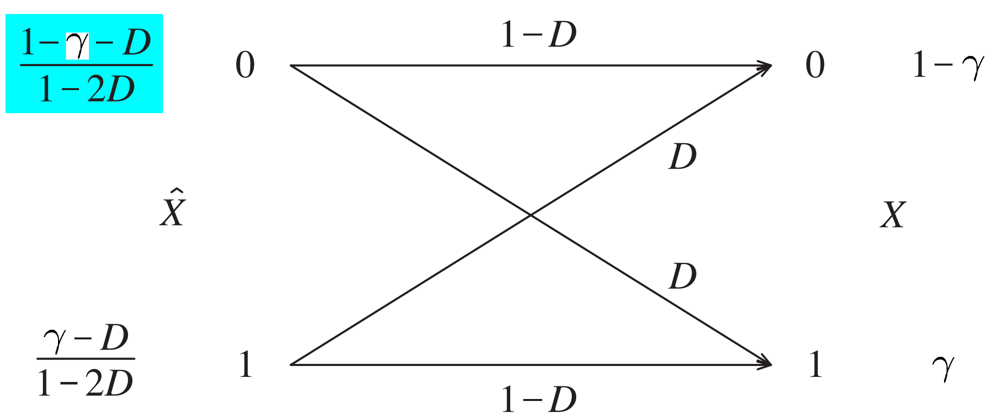
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

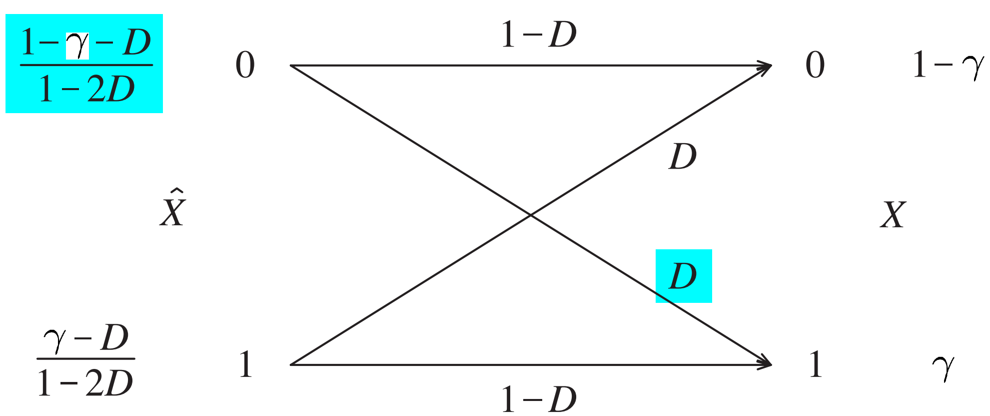
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X=0\} = 1-\gamma \quad \text{and} \quad \Pr\{X=1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0,1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X,0) = \Pr\{X=1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X,\hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X,\hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X,\hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

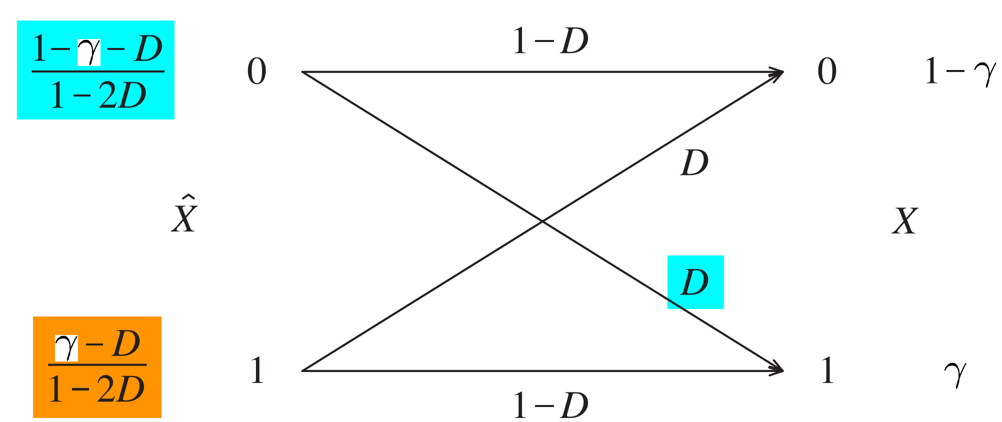
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \leq \gamma \leq \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

2. Since $\gamma \leq 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \leq D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\geq h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \neq \hat{X}\}) \\ &\geq h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \neq \hat{X}\} = Ed(X, \hat{X}) \leq D$ and $h_b(a)$ is increasing for $0 \leq a \leq \frac{1}{2}$.

5. Therefore,

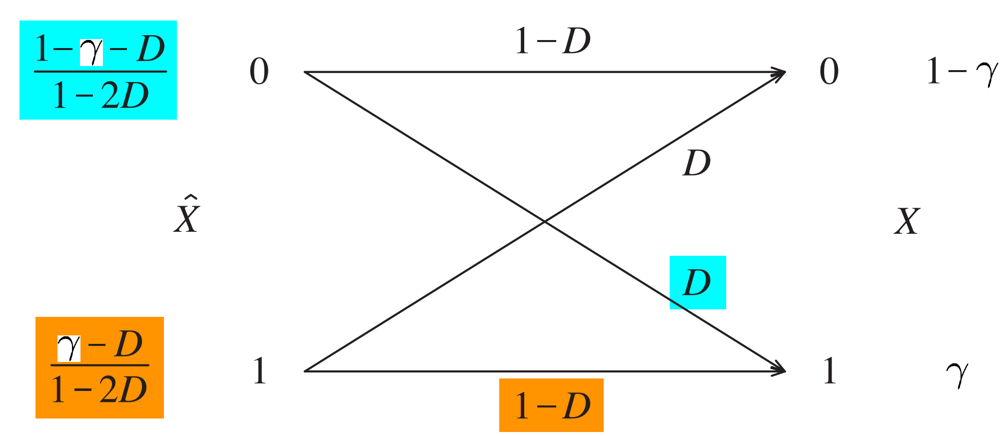
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}) \geq h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma - D}{1-2D}\right)(1-D) = \gamma$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

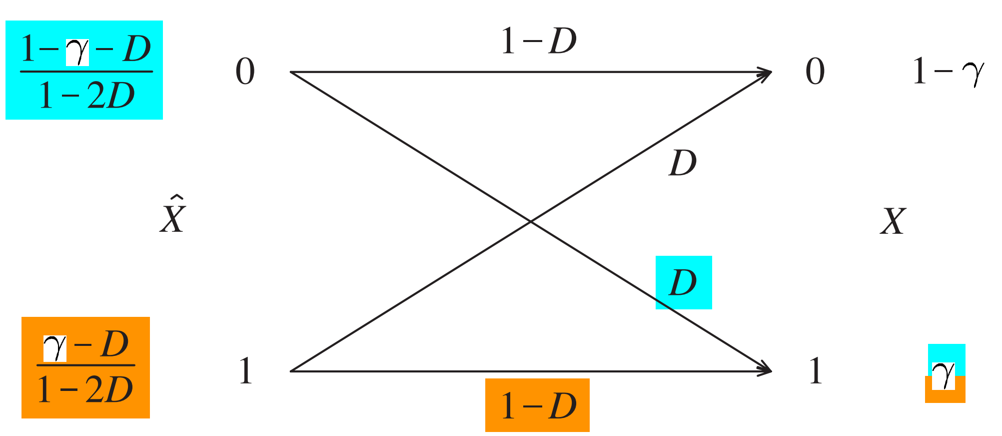
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1 - \gamma - D}{1 - 2D}\right) D + \left(\frac{\gamma - D}{1 - 2D}\right)(1 - D) = \gamma$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

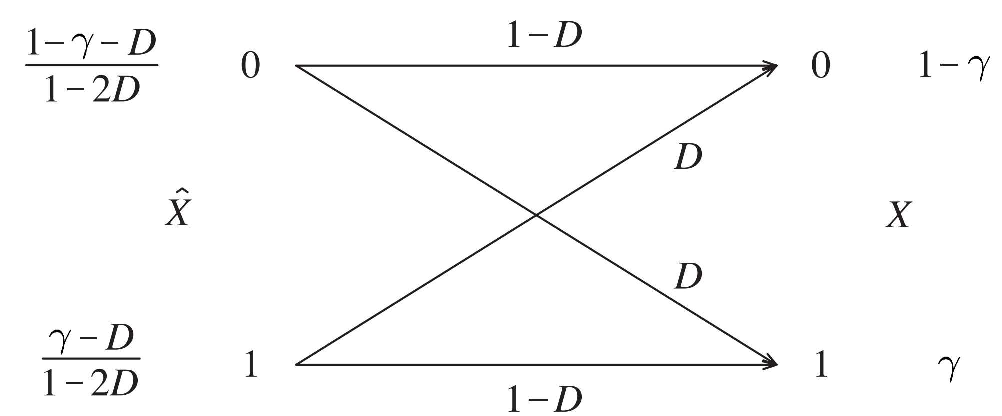
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1 - 2D} \le 1, \text{ because } D < \gamma \le 1/2$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

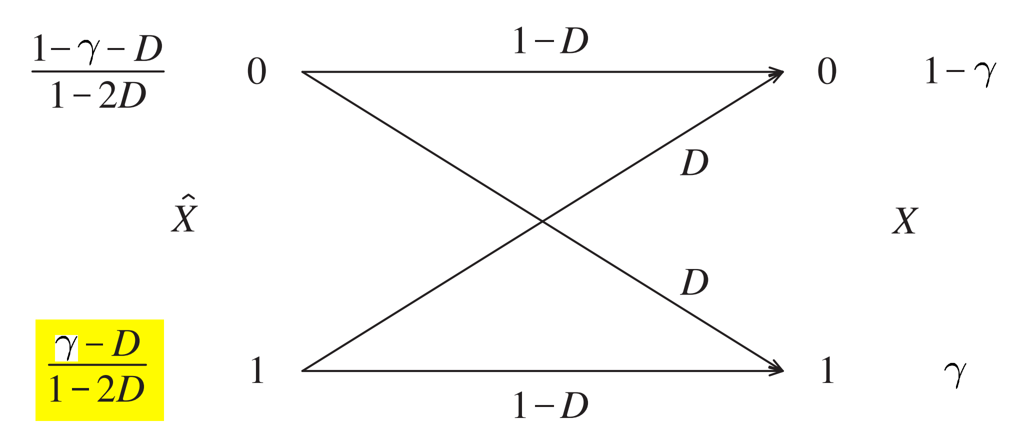
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1 - 2D} \le 1, \quad \text{because } D < \gamma \le 1/2$$

Example 8.20 (Binary Source) Let X be a binary random variable with

$$\Pr\{X = 0\} = 1 - \gamma \quad \text{and} \quad \Pr\{X = 1\} = \gamma.$$

Let $\hat{\mathcal{X}} = \{0, 1\}$ and d be the Hamming distortion measure.

Determination of $R_I(D)$

1. First consider $0 \le \gamma \le \frac{1}{2}$. We will show that

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

2. Since $\gamma \le 1/2$, $\hat{x}^* = 0$ and $D_{max} = Ed(X, 0) = \Pr\{X = 1\} = \gamma$.

3. Consider any $\hat{X}$ and let $Y = d(X, \hat{X})$.

4. Conditioning on $\hat{X}$, X and Y determine each other, and so, $H(X|\hat{X}) = H(Y|\hat{X})$.

Then for any $\hat{X}$ such that $Ed(X, \hat{X}) \le D$, where $D < \gamma = D_{max}$,

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &= h_b(\gamma) - H(Y|\hat{X}) \\ &\ge h_b(\gamma) - H(Y) \qquad (1) \\ &= h_b(\gamma) - h_b(\Pr\{X \ne \hat{X}\}) \\ &\ge h_b(\gamma) - h_b(D), \qquad (2) \end{aligned}$$

because $\Pr\{X \ne \hat{X}\} = Ed(X, \hat{X}) \le D$ and $h_b(a)$ is increasing for $0 \le a \le \frac{1}{2}$.

5. Therefore,

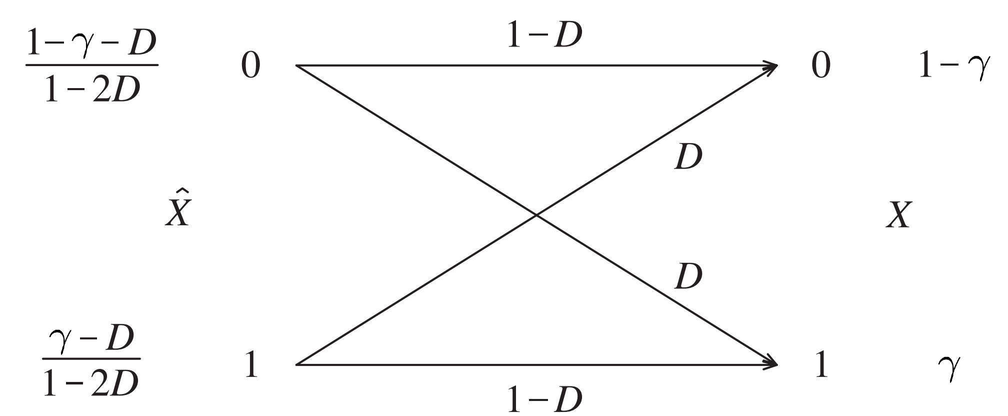
$$R_I(D) = \min_{\hat{X}: Ed(X, \hat{X}) \le D} I(X; \hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1 - \gamma - D}{1 - 2D}\right) D + \left(\frac{\gamma - D}{1 - 2D}\right) (1 - D) = \gamma$$

$$0 \le \frac{\gamma - D}{1 - 2D} \le 1, \text{ because } D < \gamma \le 1/2$$

5. Therefore,

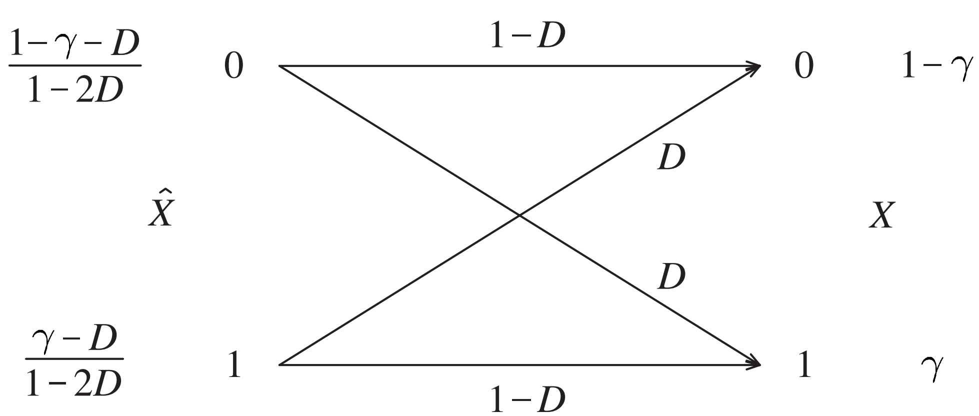
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

5. Therefore,

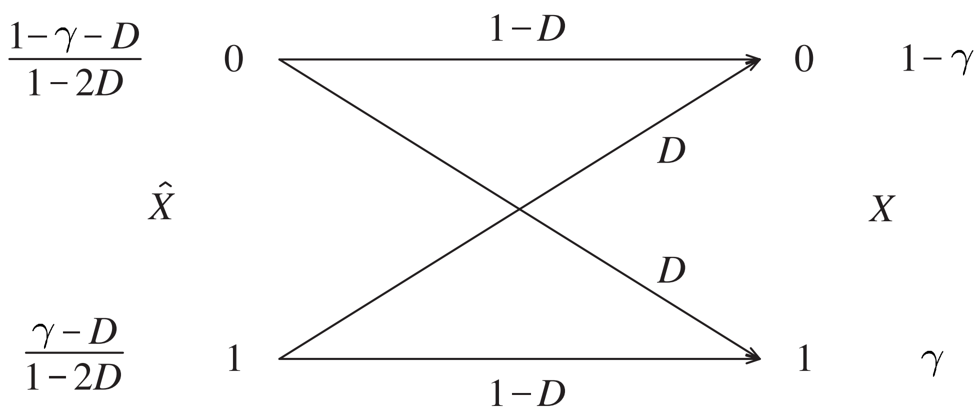
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \leq D} I(X;\hat{X}) \geq h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \leq \frac{\gamma - D}{1-2D} \leq 1, \text{ because } D < \gamma \leq 1/2$$

7. Therefore, we conclude that for $0 \leq \gamma \leq 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

5. Therefore,

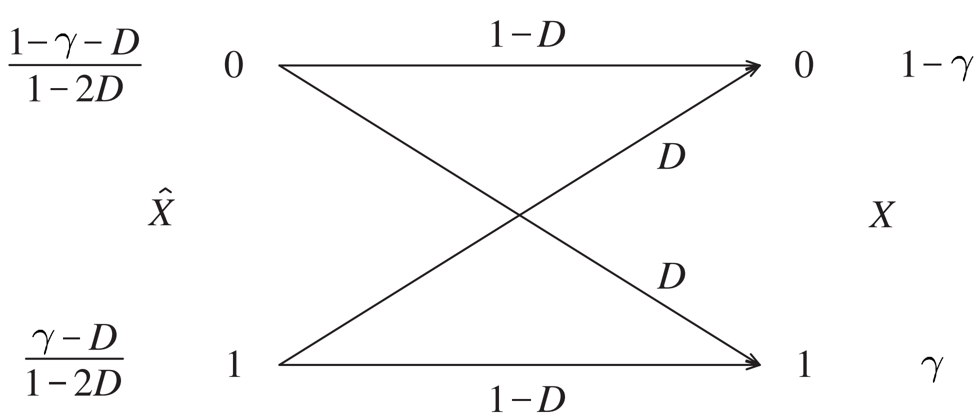
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\leq D} I(X;\hat{X}) \geq h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \leq \frac{\gamma - D}{1-2D} \leq 1, \text{ because } D < \gamma \leq 1/2$$

7. Therefore, we conclude that for $0 \leq \gamma \leq 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

5. Therefore,

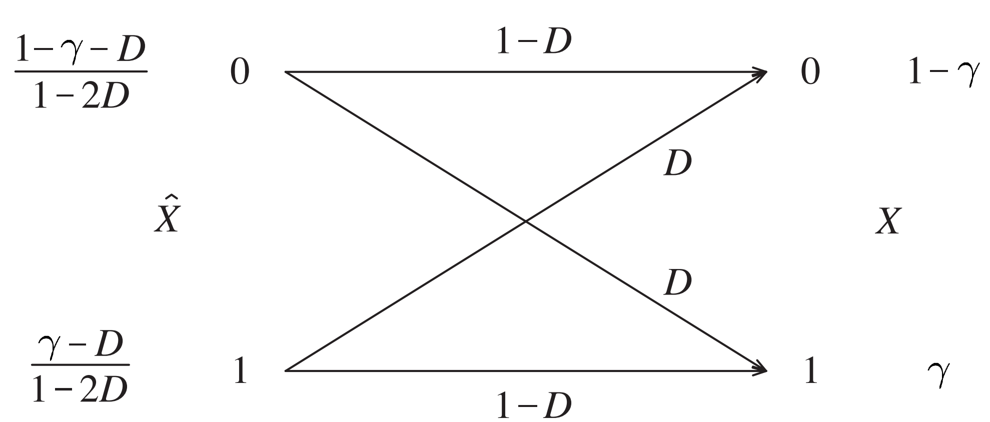
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

5. Therefore,

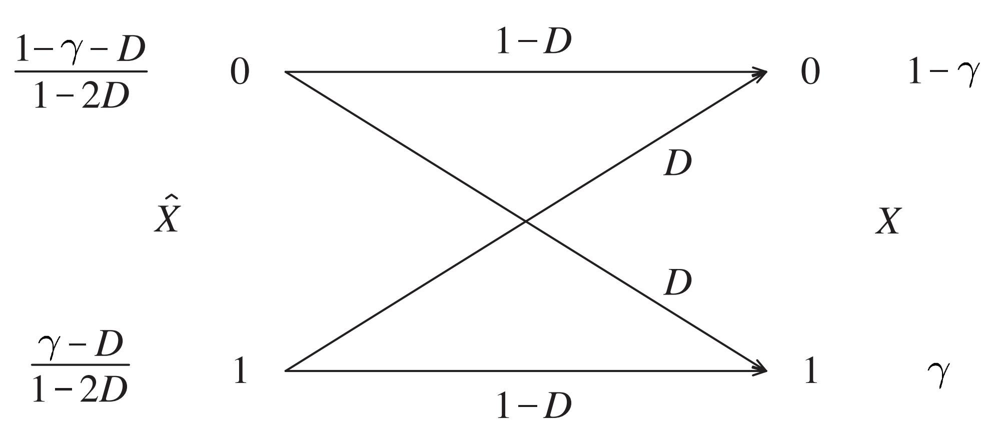
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma - D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

5. Therefore,

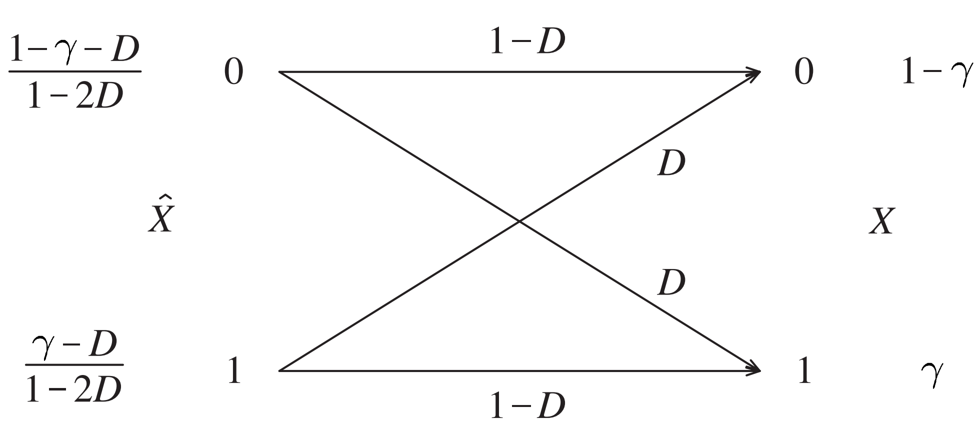
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right)D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

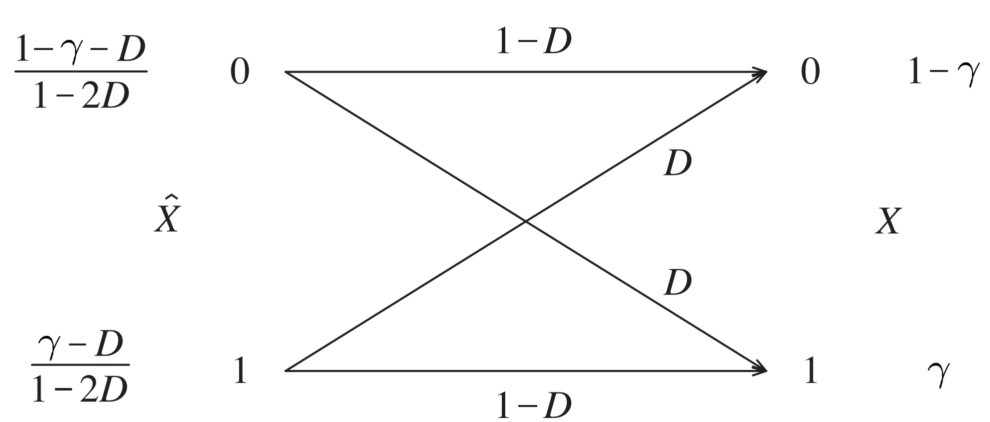
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma - D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

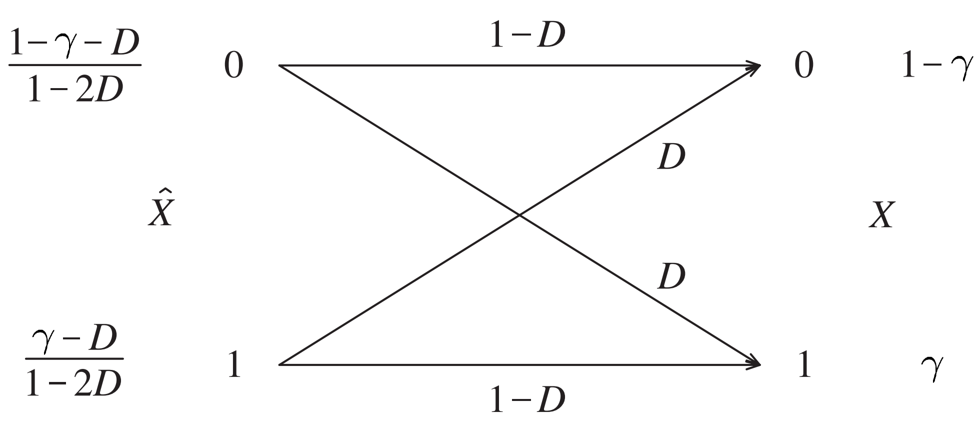
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

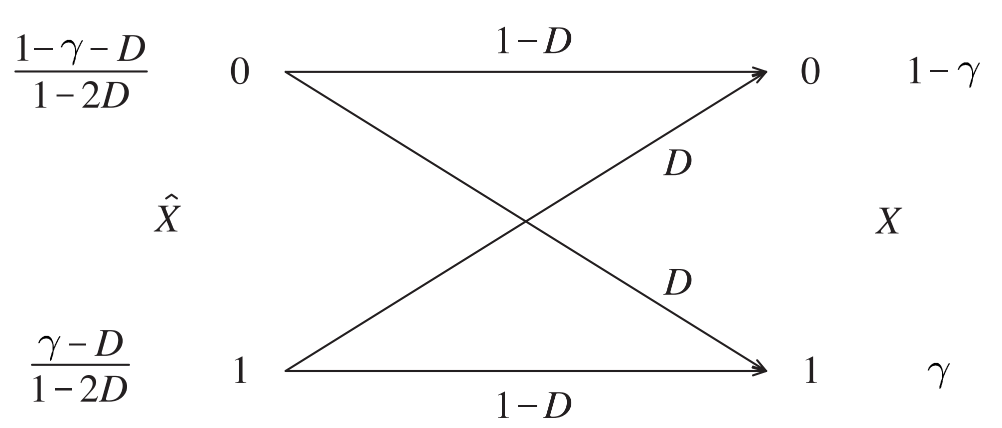
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \ne \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

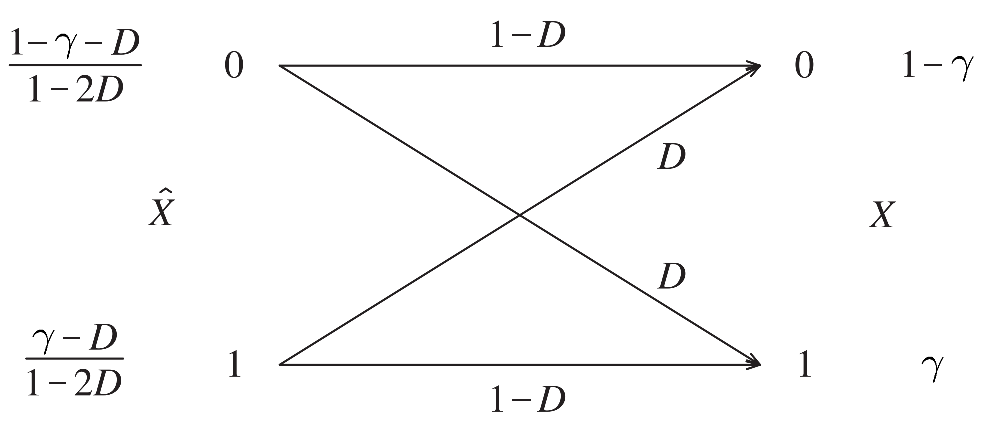
$$R_I(D) = \min_{\hat{X}: Ed(X,\hat{X}) \le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

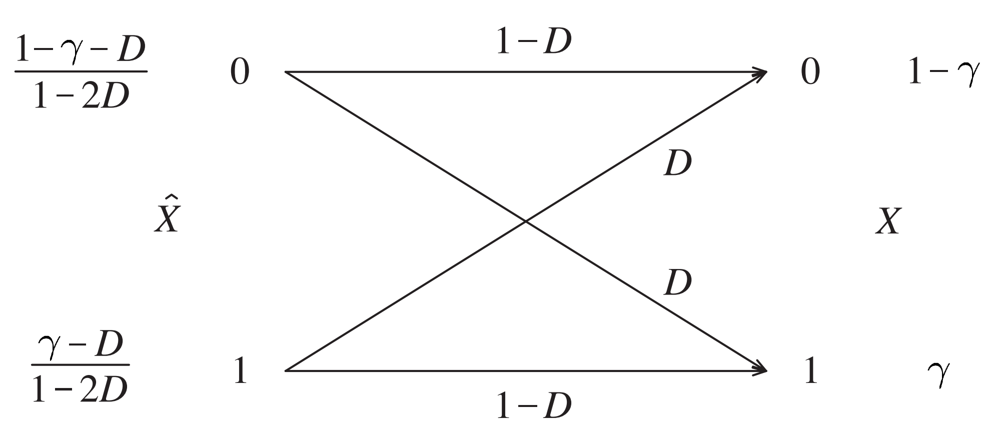
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\leq D} I(X;\hat{X}) \geq h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \leq \frac{\gamma - D}{1-2D} \leq 1, \text{ because } D < \gamma \leq 1/2$$

7. Therefore, we conclude that for $0 \leq \gamma \leq 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \gamma \\ 0 & \text{if } D \geq \gamma. \end{cases}$$

8. For $1/2 \leq \gamma \leq 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \leq D < 1-\gamma \\ 0 & \text{if } D \geq 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \leq D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \geq \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \leq \gamma \leq 1$.

5. Therefore,

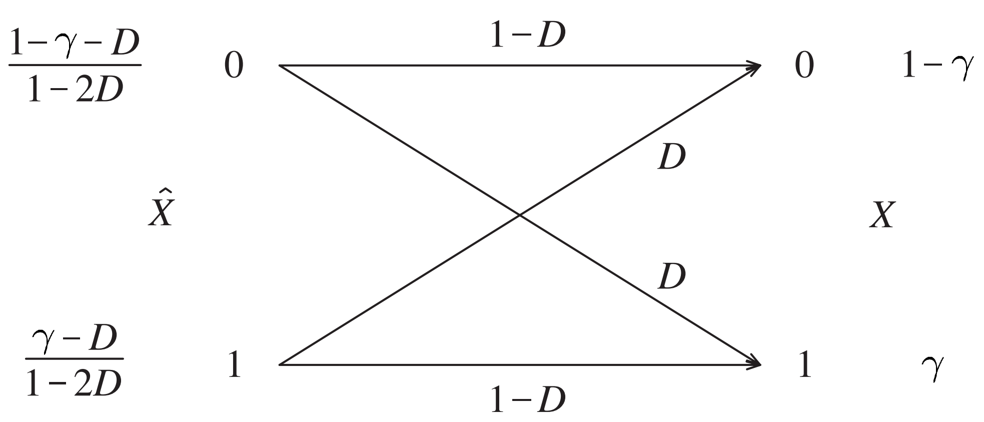
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

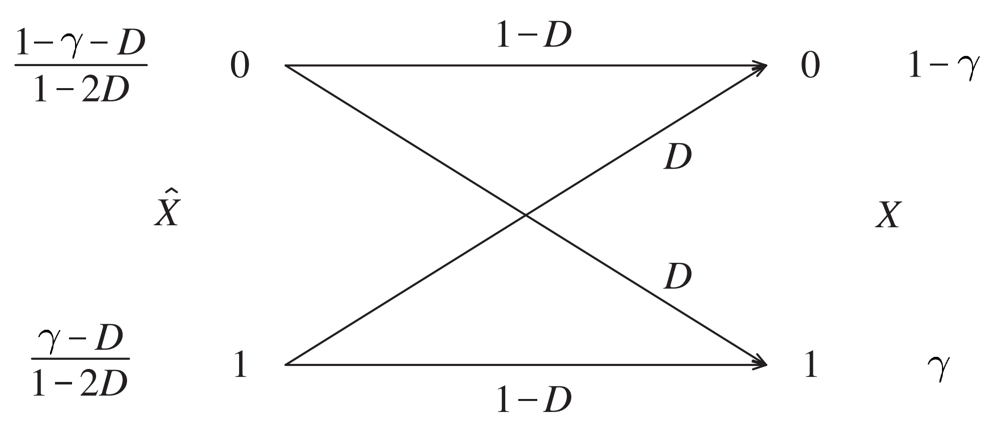
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

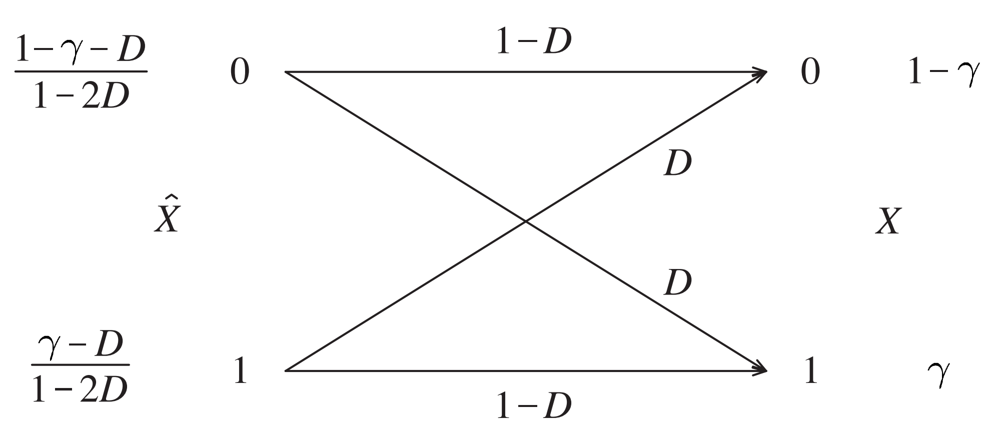
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} \overset{h_b(1-\gamma)}{\cancel{h_b(\gamma)}} - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

for $0 \le \gamma \le 1$.

5. Therefore,

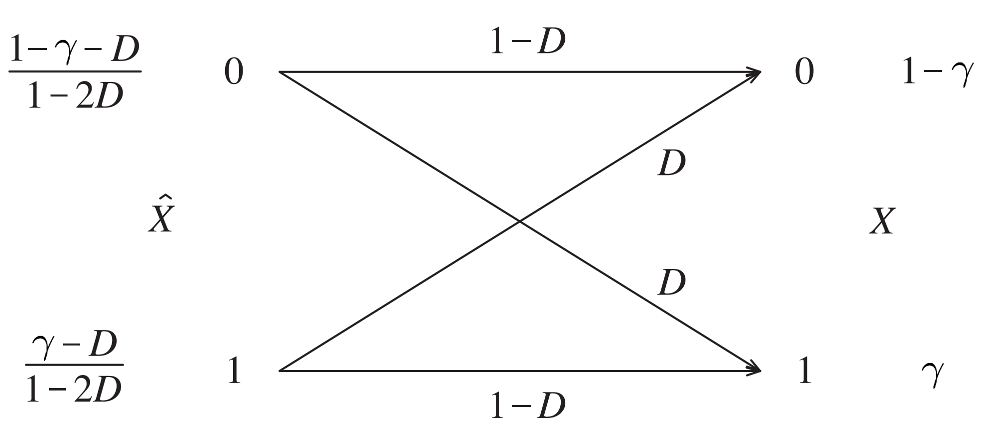
$$R_I(D) = \min_{\hat{X}:Ed(X,\hat{X})\le D} I(X;\hat{X}) \ge h_b(\gamma) - h_b(D).$$

Now need to construct $\hat{X}$ which is tight for (1) and (2), so that the above bound is achieved.

6. Observe that

- (1) tight $\Leftrightarrow$ Y independent of $\hat{X}$
- (2) tight $\Leftrightarrow$ $\Pr\{X \neq \hat{X}\} = D$

The required $\hat{X}$ can be specified by the following reverse BSC:



Note that

$$\left(\frac{1-\gamma-D}{1-2D}\right) D + \left(\frac{\gamma-D}{1-2D}\right)(1-D) = \gamma$$

$$0 \le \frac{\gamma - D}{1-2D} \le 1, \text{ because } D < \gamma \le 1/2$$

7. Therefore, we conclude that for $0 \le \gamma \le 1/2$,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \gamma \\ 0 & \text{if } D \ge \gamma. \end{cases}$$

8. For $1/2 \le \gamma \le 1$, by exchanging the roles of the symbols 0 and 1 and applying the same argument, we obtain $R_I(D)$ as above except that γ is replaced by $1-\gamma$, i.e.,

$$R_I(D) = \begin{cases} h_b(1-\gamma) - h_b(D) & \text{if } 0 \le D < 1-\gamma \\ 0 & \text{if } D \ge 1-\gamma. \end{cases}$$

9. Combining the two cases, we have

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

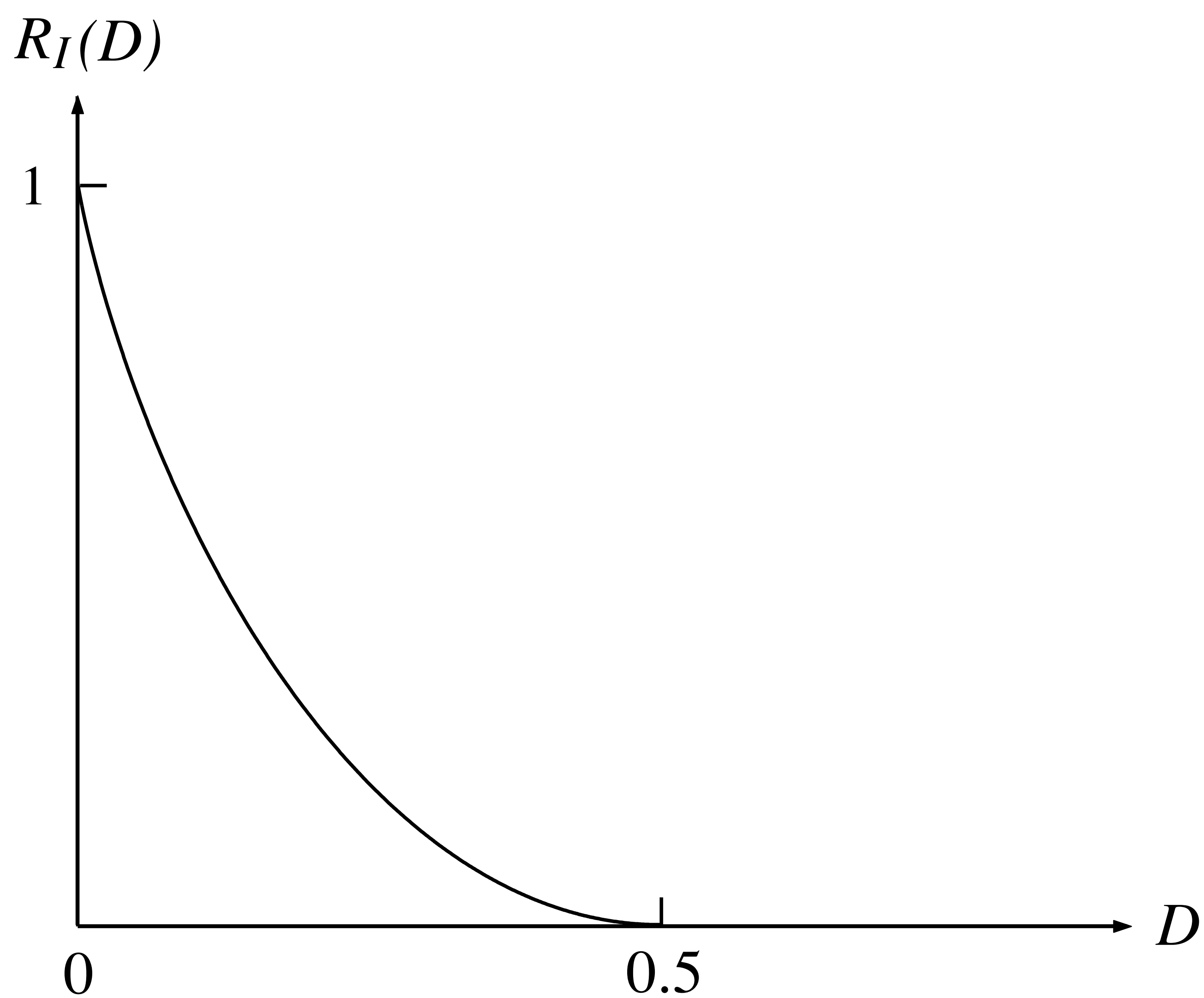
for $0 \le \gamma \le 1$.

For the uniform binary source, i.e., $\gamma = 1/2$, with the Hamming distortion measure,

$$R_I(D) = \begin{cases} 1 - h_b(D) & \text{if } \; 0 \leq D < 1/2 \\ 0 & \text{if } D \geq 1/2. \end{cases}$$

For the uniform binary source, i.e., $\gamma = 1/2$, with the Hamming distortion measure,

$$R_I(D) = \begin{cases} 1 - h_b(D) & \text{if } \ 0 \leq D < 1/2 \\ 0 & \text{if } D \geq 1/2. \end{cases}$$



A Remark

The rate-distortion theorem does not include the source coding theorem as a special case:

A Remark

The rate-distortion theorem does not include the source coding theorem as a special case:

- In Example 8.20,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \le D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \ge \min(\gamma, 1-\gamma) \end{cases}$$

  for $0 \le \gamma \le 1$, where $\gamma = \Pr\{X = 1\}$.

A Remark

The rate-distortion theorem does not include the source coding theorem as a special case:

- In Example 8.20,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \leq D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \geq \min(\gamma, 1-\gamma) \end{cases}$$

  for $0 \leq \gamma \leq 1$, where $\gamma = \Pr\{X = 1\}$.

- Therefore, $R_I(0) = h_b(\gamma) = H(X)$.

A Remark

The rate-distortion theorem does not include the source coding theorem as a special case:

- In Example 8.20,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \leq D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \geq \min(\gamma, 1-\gamma) \end{cases}$$

  for $0 \leq \gamma \leq 1$, where $\gamma = \Pr\{X = 1\}$.

- Therefore, $R_I(0) = h_b(\gamma) = H(X)$.

A Remark

The rate-distortion theorem does not include the source coding theorem as a special case:

- In Example 8.20,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \leq D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \geq \min(\gamma, 1-\gamma) \end{cases}$$

  for $0 \leq \gamma \leq 1$, where $\gamma = \Pr\{X = 1\}$.

- Therefore, $R_I(0) = h_b(\gamma) = H(X)$.

- By the rate-distortion theorem, if $R > H(X)$, the average Hamming distortion, i.e., the error probability per symbol, can be made arbitrarily small.

A Remark

The rate-distortion theorem does not include the source coding theorem as a special case:

- In Example 8.20,

$$R_I(D) = \begin{cases} h_b(\gamma) - h_b(D) & \text{if } \ 0 \leq D < \min(\gamma, 1-\gamma) \\ 0 & \text{if } D \geq \min(\gamma, 1-\gamma) \end{cases}$$

  for $0 \leq \gamma \leq 1$, where $\gamma = \Pr\{X = 1\}$.

- Therefore, $R_I(0) = h_b(\gamma) = H(X)$.

- By the rate-distortion theorem, if $R > H(X)$, the average Hamming distortion, i.e., the error probability per symbol, can be made arbitrarily small.

- However, by the source coding theorem, if $R > H(X)$, the message error probability can be made arbitrarily small, which is much stronger.