§ 3.3* Source Coding Schemes : Arithmetic Code

- Arithmetic code is a lossless data compression method that encodes the message into a fractional number between 0 and 1.
- Given a real number $\mathcal{R} \in [0, 1]$, it can be determined by $\mathcal{R} = \sum_{i\in\mathbb{N}} V_i 2^{-i}$, where $V_i \in \{0, 1\}$. Therefore, $\mathcal{R}$ can be represented by $(V_0\ V_1\ V_2\ \ldots)$. A larger precision of $\mathcal{R}$ requires a longer binary vector for representation.
- A longer source symbol sequence will become a longer binary source codeword, which can in return represent a real number with a larger precision.
- The interval [0,1) is partitioned based on the source distribution and the symbol sequence of length n.
- A codebook based encoding and decoding manner can be replaced. Codebook is no longer needed.

§ 3.3* Source Coding Schemes : Arithmetic Code

- Given a source whose symbols are chosen from the alphabet $X=\{x_1, x_2, x_3, \ldots, x_U\}$, which exhibits a distribution of $p_1, p_2, p_3, \ldots, p_U$, respectively.
- The cumulative distribution function of source symbols is $F(x)$, i.e., let $F(x_0) = 0$, x_0 is an imaginary symbol

$$F(x_1) = p_1$$

$$F(x_2) = p_1 + p_2$$

$$\vdots$$

$$F(x_U) = p_1 + p_2 + \cdots + p_U = 1$$

Note : $F(x_i) - F(x_{i-1}) = p_i$.

§ 3.3* Source Coding Schemes : Arithmetic Code

- **Encoding** of a source symbol sequence $s_1, s_2, \cdots, s_n \in X$

  Initialization

  Let $A^{(0)} = 0$ and $B^{(0)} = 1$

  Encode symbol x_i at time instant t, i.e., $s_t = x_i$

  $$A^{(t)} = A^{(t-1)} + (B^{(t-1)} - A^{(t-1)})\, F(x_{i-1})$$

  $$B^{(t)} = A^{(t-1)} + (B^{(t-1)} - A^{(t-1)})\, F(x_i)$$

  Note : The coded interval of time instant t is $B^{(t)} - A^{(t)}$, and

  $$\begin{aligned} B^{(t)} - A^{(t)} &= \left(B^{(t-1)} - A^{(t-1)}\right)\left(F(x_i) - F(x_{i-1})\right) \\ &= \left(B^{(t-1)} - A^{(t-1)}\right)p_i \end{aligned}$$

§ 3.3* Source Coding Schemes : Arithmetic Code

Example 3.6 Given a source with symbols x_1, x_2, x_3 and their probabilities $p_1 = 0.6$, $p_2 = 0.3, p_3 = 0.1$.

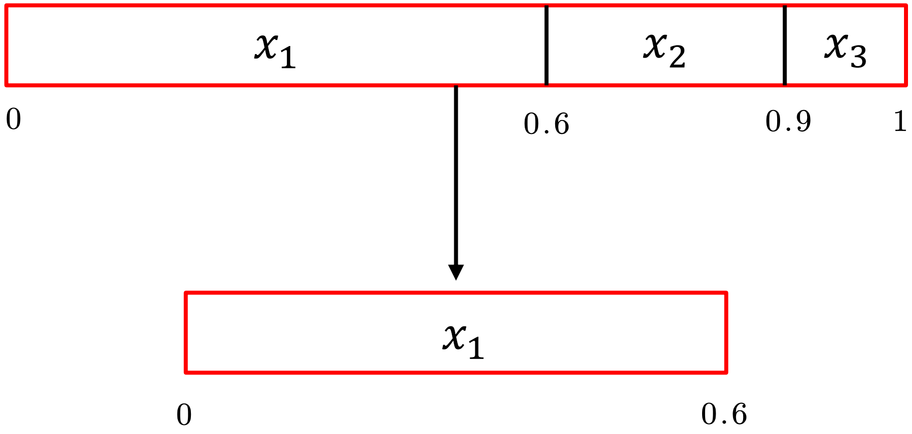
Given the sequence $(s_1\ s_2\ s_3\ s_4) = (x_1\ x_3\ x_2\ x_1)$

Determine values of cumulative distribution function $F(x)$ as

$$F(x_1) = 0.6, F(x_2) = 0.9, F(x_3) = 1.0, F(x_0) = 0$$

For $t = 1$, encode x_1:

$$A^{(1)} = A^{(0)} + \left(B^{(0)} - A^{(0)}\right) \cdot F(x_0) = 0,$$
$$B^{(1)} = A^{(0)} + \left(B^{(0)} - A^{(0)}\right) \cdot F(x_1) = 0.6$$



§ 3.3* Source Coding Schemes : Arithmetic Code

For $t = 2$, encode x_3:

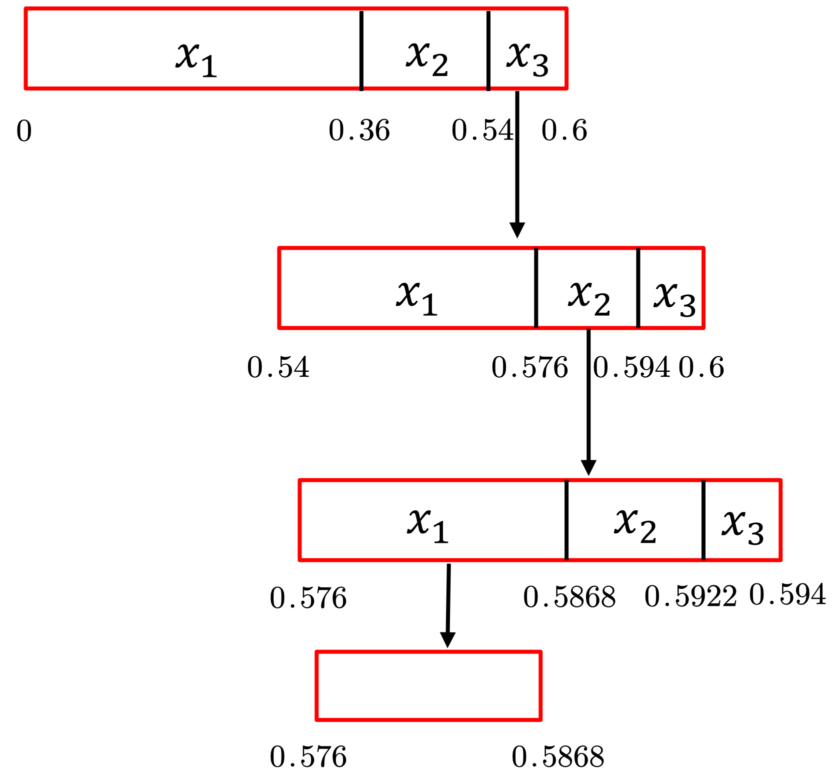
$$A^{(2)} = A^{(1)} + \left(B^{(1)} - A^{(1)}\right) \cdot F(x_2) = 0.54 \ ,$$
$$B^{(2)} = A^{(1)} + \left(B^{(1)} - A^{(1)}\right) \cdot F(x_3) = 0.6$$

For $t = 3$, encode x_2:

$$A^{(3)} = A^{(2)} + \left(B^{(2)} - A^{(2)}\right) \cdot F(x_1) = 0.576 \ ,$$
$$B^{(3)} = A^{(2)} + \left(B^{(2)} - A^{(2)}\right) \cdot F(x_2) = 0.594$$

For $t = 4$, encode x_1:

$$A^{(4)} = A^{(3)} + \left(B^{(3)} - A^{(3)}\right) \cdot F(x_0) = 0.576 \ ,$$
$$B^{(4)} = A^{(3)} + \left(B^{(3)} - A^{(3)}\right) \cdot F(x_1) = 0.5868$$



Finally, we get the internal $[0.576, 0.5868)$.

§ 3.3* Source Coding Schemes : Arithmetic Code

For $[0.576, 0.5868)$, determine the fractional number c that can be represented by a finite binary expression.

Initial $a = 0, b = 1, c = \frac{a+b}{2} = 0.5$.

While $c \notin [0.576, 0.5868)$ (binary search)

if $c < 0.576$

$a = c;$

else if $c \geq 0.5868$

$b = c;$

$c = \frac{a+b}{2};$

End

Output : $c = 0.578125$

§ 3.3* Source Coding Schemes : Arithmetic Code

Convert the fractional number 0.578125 to its binary representation.

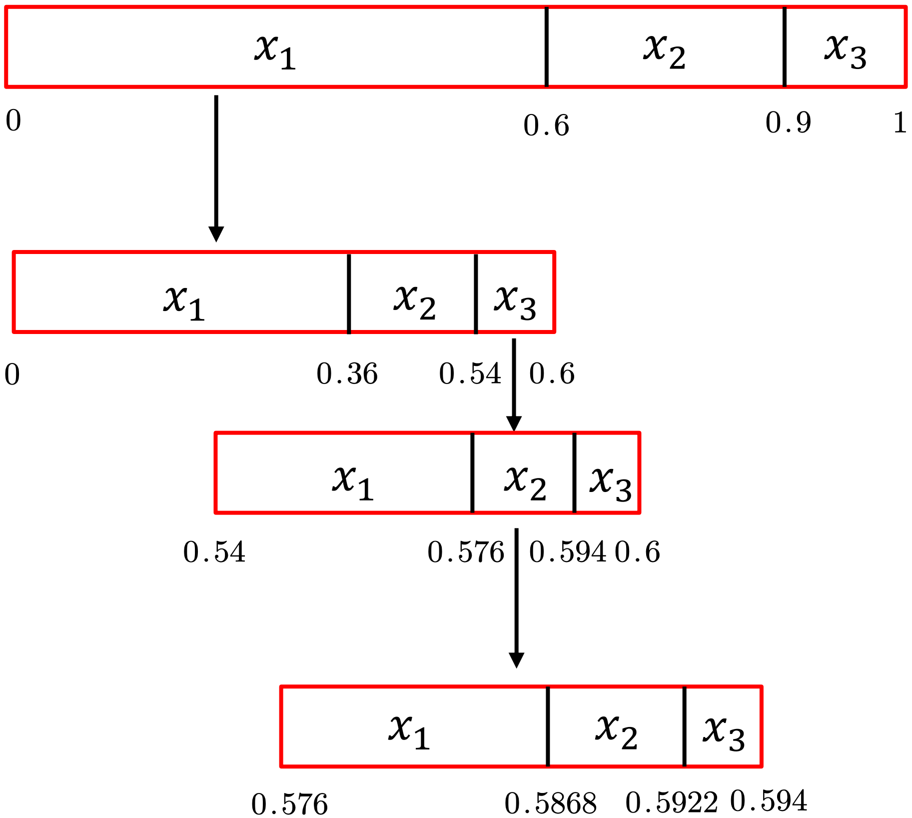
$$
\begin{aligned}
0.578125 \times 2 &= 1.15625 \\
0.15625 \times 2 &= 0.3125 \\
0.3125 \times 2 &= 0.625 \\
0.625 \times 2 &= 1.25 \\
0.25 \times 2 &= 0.5 \\
0.5 \times 2 &= 1
\end{aligned}
$$

$$0.100101_2 = 2^{-1} + 2^{-4} + 2^{-6} = 0.578125$$

The sequence $x_1 x_3 x_2 x_1$ is encoded to 0.100101_2.

§ 3.3* Source Coding Schemes : Arithmetic Code

Decoding Arithmetic code 0.100101_2 (0.578125)



$0 < 0.578125 < 0.6$, output: x_1

$0.54 < 0.578125 < 0.6$, output: x_3

$0.576 < 0.578125 < 0.594$, output: x_2

$0.576 < 0.578125 < 0.5868$, output: x_1

Finally, the bits 0.100101_2 is decoded to the sequence $x_1 x_3 x_2 x_1$.

§ 3.3* Source Coding Schemes : Arithmetic Code

Theorem 3.6 Entropy Coding An entropy coding is any lossless source coding with an expected length greater than or equal to the source entropy.

- Note: Arithmetic code is an entropy coding.

Proof:

- Source : $X = \{x_1, x_2, \ldots, x_U\}$
- Symbol sequence : $s_1, s_2, \cdots, s_n$
- Average length per symbol: l
- Width of the final coding interval: $\beta = \prod_{t=1}^{n} p(s_t)$
- The minimum number of bits to represent the internal: $\lceil -\log_2 \beta \rceil = -\log_2 \beta + \sigma$

§ 3.3* Source Coding Schemes : Arithmetic Code

$$l = \frac{\lceil -\log_2 \beta \rceil}{n} = \frac{-\log_2 \beta + \sigma}{n}$$

$$= \frac{-\sum_{t=1}^{n} \log_2 p(s_t) + \sigma}{n}$$

$$L = \mathbb{E}[l]$$

$$= \mathbb{E}\left[-\frac{1}{n}\sum_{t=1}^{n} \log_2 p(s_t)\right] + \frac{\sigma}{n}$$

$$= -\frac{1}{n}\sum_{t=1}^{n} \log_2 p(s_t) + \frac{\sigma}{n}$$

$$\log_{n\to\infty} L = H(X)$$

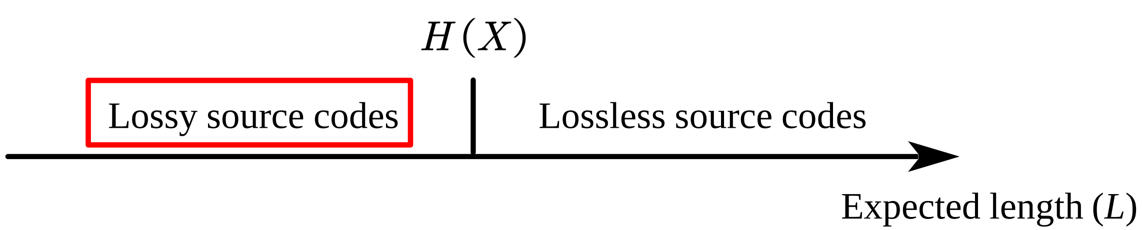
When $n \to \infty$, expected length of an arithmetic code can reach the entropy of the source.

§ 3.4* Rate-Distortion

- Consider a source coding with its rate (expected length) less than the source entropy.

- Under such a situation, source symbols are compressed with distortion.

- What is the best possible trade-off between rate and distortion?



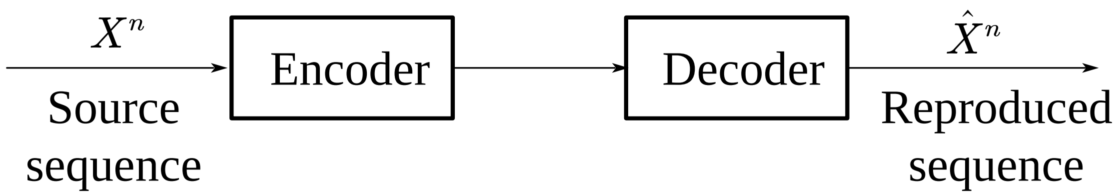
§ 3.4* Rate-Distortion

- Source alphabet: X; Reproduction (decoding output) alphabet: $\hat{X}$.
- Single-letter distortion measure: $d: X \times \hat{X} \to \mathcal{R}^+$.
- The value $d(x, \hat{x})$ denotes the distortion incurred when a source symbol $x \in X$ is reproduced as $\hat{x} \in \hat{X}$.
- Hamming distortion: $d(x, \hat{x}) = \begin{cases} 0, & \text{if } x = \hat{x}; \\ 1, & \text{if } x \neq \hat{x}. \end{cases}$
- The distortion between sequences x^n and $\hat{x}^n$: $d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{t=1}^{n} d(x_t, \hat{x}_t)$.

In light of $x^n = (x_1, x_2, \ldots, x_n)$ and $\hat{x}^n = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$

§ 3.4* Rate-Distortion

- Let $(X_t, 1 \leq t \leq n)$ (also denoted as X^n) be an i.i.d. information source with random variable $X \sim p(x), x \in X$.



A $(2^{nR}, n)$ rate-distortion code:

n: length of source symbol sequence

nR: length of (binary) codeword sequence

2^{nR}: cardinality of source codebook

Rate of this code: $R = \frac{\log_2 2^{nR}}{n}$

§ 3.4* Rate-Distortion

A $(2^{nR}, n)$ rate-distortion code is defined by two functions :

Encoding function : $f_n: X^n \to \{1, 2, \cdots, 2^{nR}\}$

Decoding function : $g_n: \{1, 2, \cdots, 2^{nR}\} \to \hat{X}^n$

For simplicity, codewords are denoted by their indices.

Distortion of this source code: $D = \mathbb{E}[d\left(X^n, g_n\left(f_n(X^n)\right)\right)]$

$$= \sum_{x^n} p(x^n) d\left(x^n, g_n\left(f_n(x^n)\right)\right)$$

On average, the amount of distortion (disagreement) between a source symbol x and a reproduced symbol $\hat{x}$.

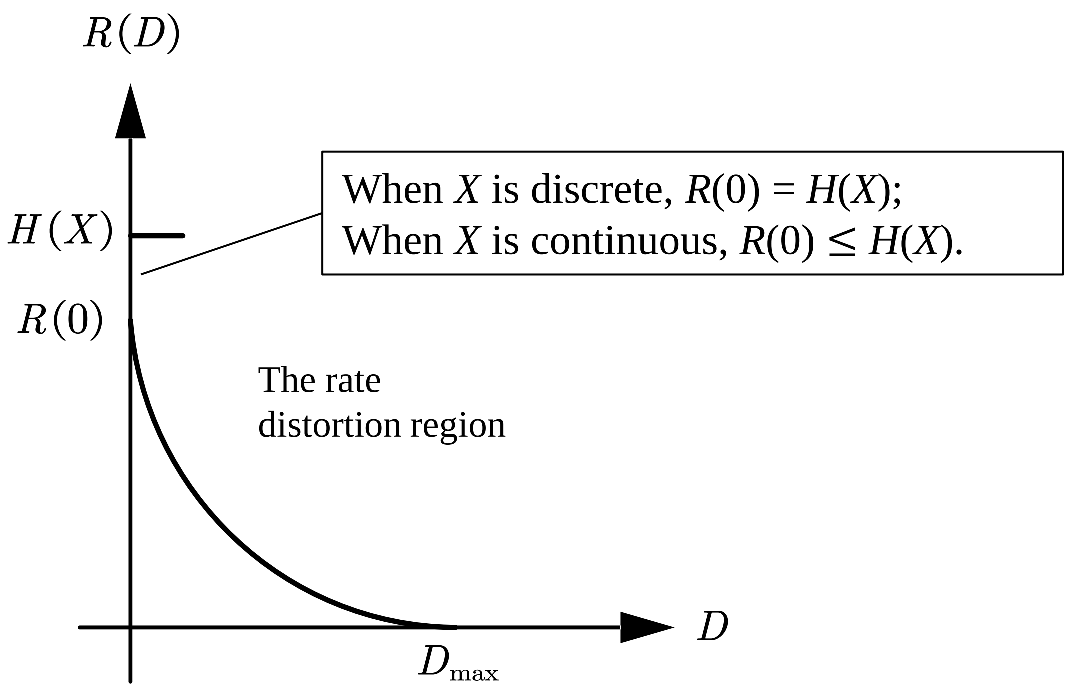
$D \in [0, 1]$ (under Hamming metric)
$R \in [0, H(X)]$

Definition 3.1: A rate-distortion pair (R, D) is said to be achievable if there exists a sequence of $\left(2^{nR}, n\right)$ rate-distortion codes (f_n, g_n) with

$$\lim_{n\to\infty} \mathbb{E}[d\left(X^n, g_n\left(f_n(X^n)\right)\right)] \leq D.$$

§ 3.4* Rate-Distortion

- **Definition 3.2 :** Rate-distortion function $R(D)$ is the minimum of all rates R for a given distortion D such that (R, D) is achievable.



§ 3.4* Rate-Distortion

Properties of rate-distortion function $R(D)$:

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

Let $D' \geq D$, $(R(D), D)$ is achievable $\Rightarrow$ $(R(D), D')$ is achievable.

(2) $R(D)$ is convex.

$$R(\mathbb{E}[D]) \leq \mathbb{E}[R(D)]$$

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

$(0, D_{\max})$ is achievable $\Rightarrow R(D) = 0$ for $D \geq D_{\max}$.

(4) $R(D) \leq H(X)$.

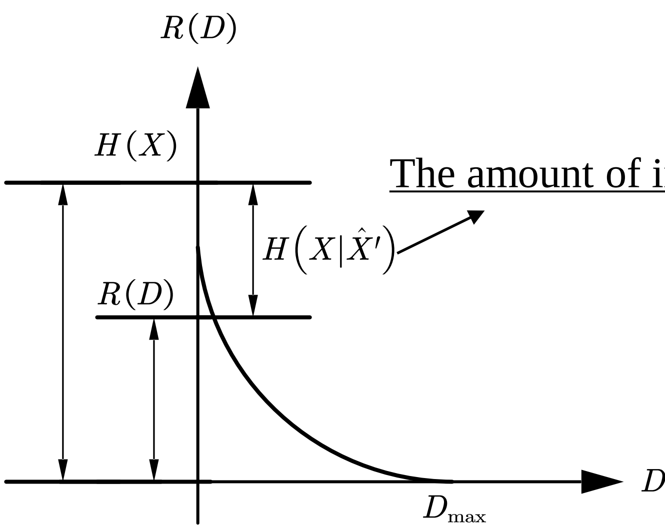
The rate-distortion code is a lossy source code.

§ 3.4* Rate-Distortion

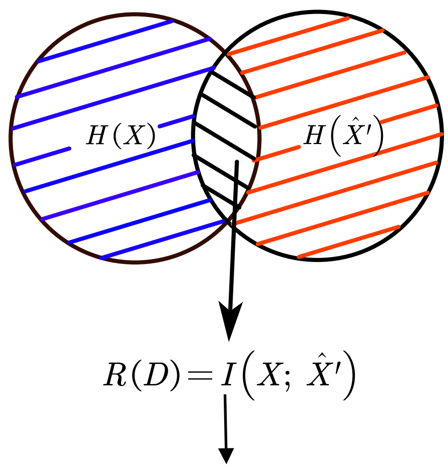
Definition 3.3 : Information rate-distortion function : $R^{(I)}(D) = \min_{\hat{X}:\mathbb{E}[d(X,\hat{X})]\le D} I(X;\hat{X})$

Theorem 3.7 Rate-Distortion Theorem $R^{(I)}(D) = R(D)$.

Assume that $\min_{\hat{X}:\mathbb{E}d(X,\hat{X})\le D} I(X;\hat{X}) = I(X;\hat{X}')$



The amount of information we loose



The amount of information we know about X through source code of rate $R(D)$

$R(D)$ is the minimum achievable rate for lossless compression of the residual information $I(X;\hat{X}')$.

§ 3.4* Rate-Distortion

Example 3.7 Given a binary source that satisfies $\Pr\{X=0\}=1-\gamma$ and $\Pr\{X=1\}=\gamma$ and Hamming distortion D which implies $\mathbb{E}[d(X,\hat{X})]=\Pr\{X\neq\hat{X}\}\leq D$. Determine $R(D)$.

Without loss of generality, let us assume $\gamma \leq \frac{1}{2}$.

$$I(X;\hat{X}) = H(X) - H(X|\hat{X})$$
$$= H_b(\gamma) - H(X \oplus \hat{X}|\hat{X})$$
$$\geq H_b(\gamma) - H(X \oplus \hat{X})$$
$$= H_b(\gamma) - H_b\left(\Pr\{X \neq \hat{X}\}\right)$$

Note : $H_b(\gamma) = -\gamma\log_2\gamma - (1-\gamma)\log_2(1-\gamma)$

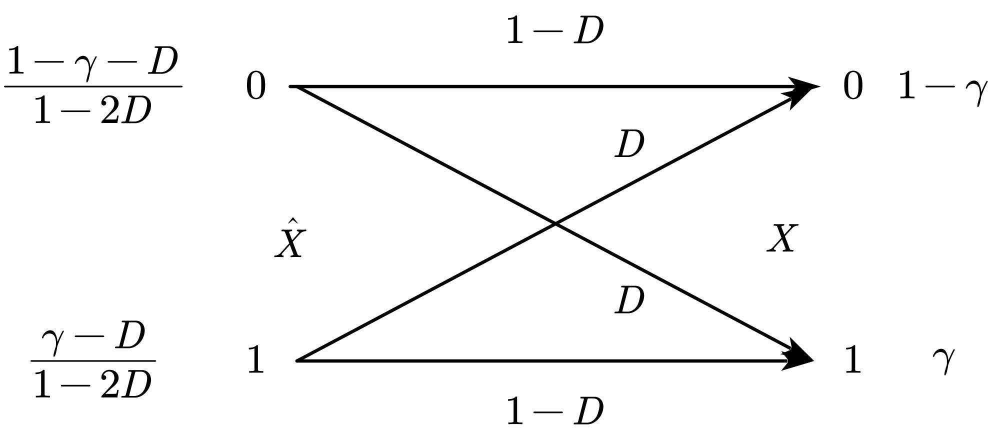
$H_b(\gamma)$ increases with γ for $\gamma \leq \frac{1}{2}$ and assume $D \leq \frac{1}{2}$

$$I(X;\hat{X}) \geq H_b(\gamma) - H_b(D)$$

§ 3.4* Rate-Distortion

$$R^{(I)}(D) = R(D) = \min_{\hat{X}:\mathbb{E}[d(X,\hat{X})]\le D} I\left(X;\hat{X}\right) \ge H_b(\gamma) - H_b(D).$$

When $\gamma \ge D$



For this BSC, $I\left(X;\hat{X}\right) = H(X) - H\left(X\middle|\hat{X}\right)$

$$= H_b(\gamma) - (-D\log_2 D - (1-D)\log_2(1-D))$$

$$= H_b(\gamma) - H_b(D).$$

§ 3.4* Rate-Distortion

Information rate distortion function:

$$R^{(I)}(D) = R(D) = \min_{\hat{X}:\mathbb{E}[d(X,\hat{X})]\le D} I\left(X;\hat{X}\right)$$

$$= H_b(\gamma) - H_b(D), \text{ where } \gamma < \frac{1}{2} \text{ and } \gamma \ge D.$$

$$\downarrow$$

$$R(D) = \begin{cases} H_b(\gamma) - H_b(D), & \text{if } 0 \le D \le \min(\gamma, 1-\gamma)\,; \\ 0, & \text{if } D \ge \min(\gamma, 1-\gamma)\,. \end{cases}$$