

INFORMATION THEORY

BASICS

EE 6317 - Channel coding

Information Theory

- Source coding / data compression
- Channel coding / reliably communicating in the presence of noise.
- Hypothesis testing / classification with a prior.

Information Theory

- Source coding

$$\underline{H(X)} = - \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)}$$

- Channel coding

$$\underline{I(X; Y)} = \sum_{xy} p(xy) \log_2 \frac{p(xy)}{p(x)p(y)}$$

- Hypothesis testing

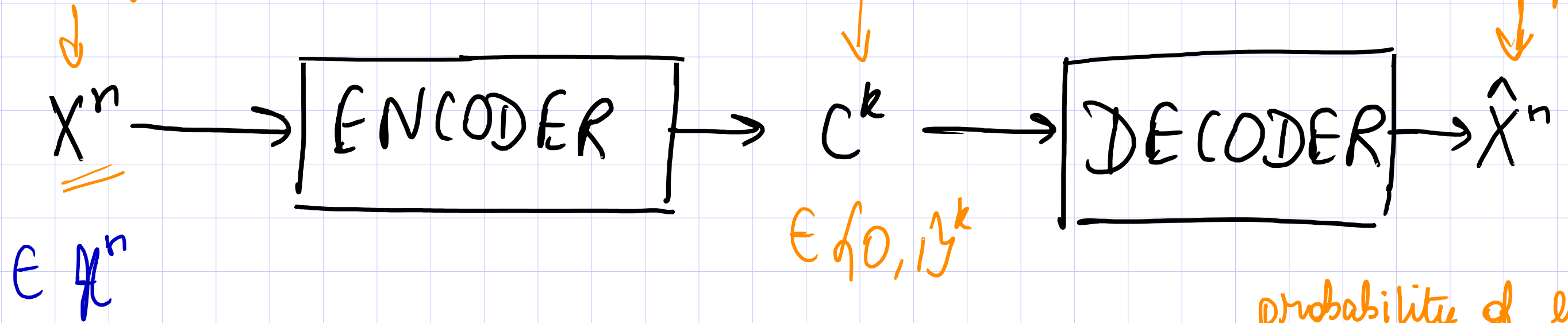
$$\underline{D(p||q)} = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

Source coding

raw file

compressed file

decompressed



$X^n \sim \text{iid}(p_x)$

A : input alphabet
(words, letters, ASCII chars)

$$R = \frac{k}{n}$$

- fixed-length

- variable-length

(zip, gzip, 7z)

probability of error

$$P_e = P_n[X^n \neq \hat{X}^n]$$

k is fixed (X^n)

k depend on X^n

The source coding theorem (Discrete memoryless source iid)

Theorem (Shannon, 1948) For any $\epsilon > 0$,

① \exists compression algorithms that achieve

$$\lim_{n \rightarrow \infty} p_e = 0 \quad \&$$

$$\lim_{n \rightarrow \infty} R = H(X) + \epsilon$$

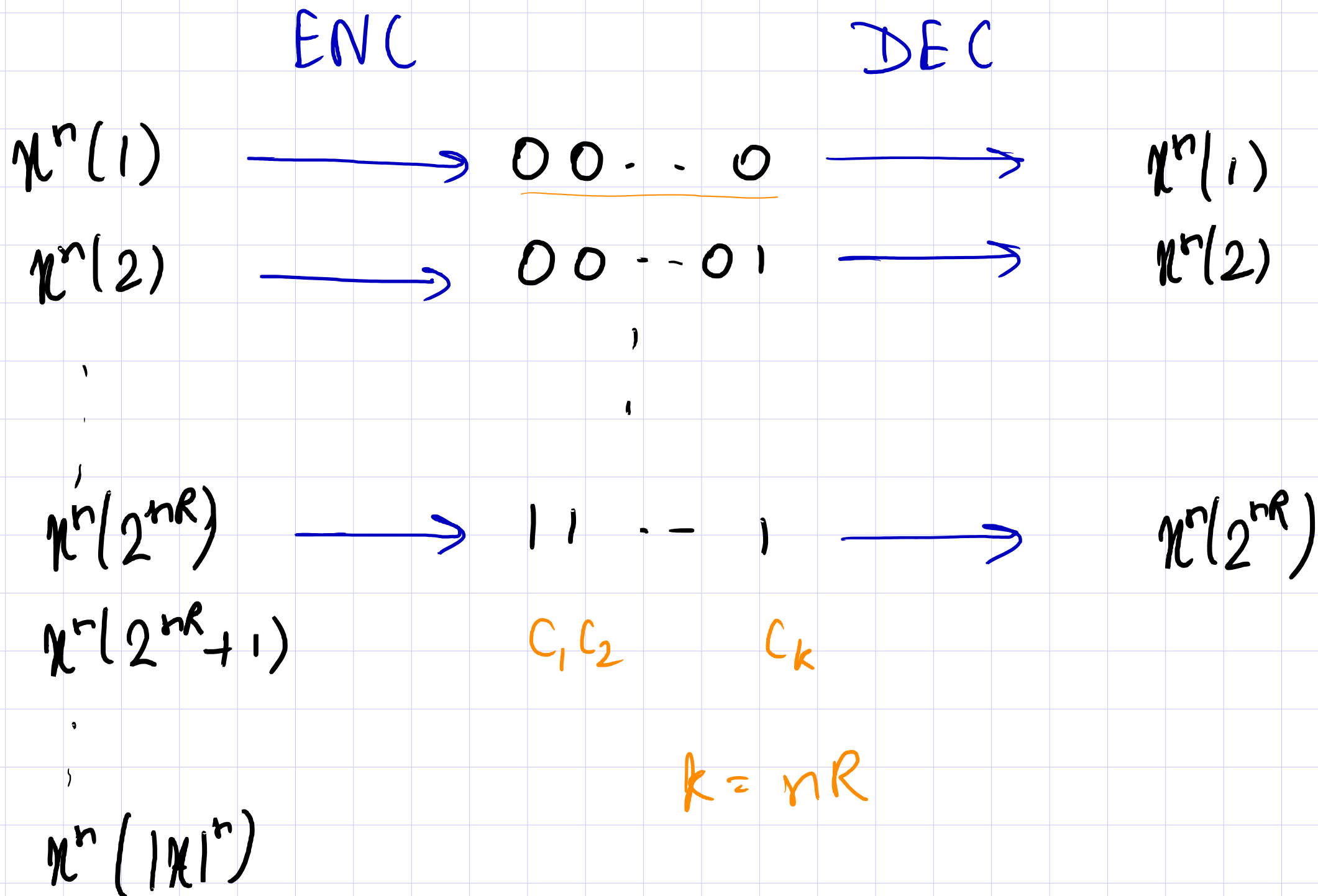
② For any compression algorithm with

$$\lim_{n \rightarrow \infty} R \leq H(X) - \epsilon,$$

$$\lim_{n \rightarrow \infty} p_e > 0.$$

} not possible
to compress
below $H(X)$

General fixed-length compressor



reproducible

\mathcal{R}

always in error

Analysis

Rate

$$R = \frac{1}{n} \log_2 |\mathcal{R}| = \frac{k}{n}$$

Probability of
error

$$P_e = \sum_{x^n \in \mathcal{R}^c} P_{X^n}(x^n) \\ = P_{X^n}[X^n \notin \mathcal{R}]$$

High probability set \mathcal{A} :

$$P_{X^n}[X^n \notin \mathcal{A}] \rightarrow 0 \\ \text{as } n \rightarrow \infty$$

Source compression: find \mathcal{A} of min. size

(Weakly) Typical set (Cover & Thomas)

$$T_{\epsilon, w}^{(n)}(p_X) = \left\{ x^n \in \mathcal{X}^n : H(X) - \epsilon \leq \frac{1}{n} \log_2 \frac{1}{p_{X^n}(x^n)} \leq H(X) + \epsilon \right\}$$

OR

$$2^{-n(H(X) + \epsilon)} \leq p_{X^n}(x^n) \leq 2^{-n(H(X) - \epsilon)}$$

$\mathcal{R} = T_{\epsilon, w}^{(n)}(p_X) \rightarrow$ one compressor.

Properties

$$\textcircled{1} \quad \Pr[X^n \notin T_{\epsilon, w}^{(n)}(p_X)] \rightarrow 0$$

$\text{as } n \rightarrow \infty$

High prob set!

$$\textcircled{2} \quad |T_{\epsilon, w}^{(n)}(p_X)| \leq 2^{n(H(X) + \epsilon)}$$

$$R \leq H(X) + \epsilon$$

$$\textcircled{3} \quad |T_{\epsilon, w}^{(n)}(p_X)| \geq 2^{n(H(X) - \epsilon)} (1 - \epsilon)$$

for all suff. large n .

$$Y(x^n) = \frac{1}{n} \log \frac{1}{P_{x^n}(x^n)} = \frac{1}{n} \log \frac{1}{\prod_{i=1}^n P_X(x_i)}$$

$$= \sum_{i=1}^n \frac{\log \frac{1}{P_X(x_i)}}{n}$$

$$Y(x^n) = \sum_{i=1}^n \frac{\log \frac{1}{P_X(x_i)}}{n} \rightarrow \text{Sum of iid RVs!}$$

$$\mathbb{E} Y(x^n) = H(X)$$

$$P_n \left[|Y(x^n) - H(X)| > \epsilon \right] \rightarrow 0$$

as $n \rightarrow \infty$

$$P_n[X^n \in T_{\epsilon, w}^{(n)}(p_x)]$$

$\geq 1/2$

for large n

≤ 1

$$\sum_{x^n \in T_{\epsilon}^{(n)}(p_x)} P_{X^n}(x^n)$$

$\geq 1/2$
 ≤ 1

$$P_{X^n}(x^n) \geq 2^{-n(H(x) + \epsilon)}$$

$$\sum_{x^n \in T_{\epsilon, w}^{(n)}(p_x)}$$

$$2^{-n(H(x) + \epsilon)}$$

$\geq 1/2$
 ≤ 1

$$2^{-n(H(x) + \epsilon)}$$

$$|T_{\epsilon, w}^{(n)}(p_x)| \leq 1$$

$\geq 1/2$

$$|T_{\epsilon, w}^{(n)}(p_x)| \leq 2^{n(H(p_x) + \epsilon)}$$

$$|T_{\epsilon, w}^{(n)}(p_x)| \geq 2^{n(H(x) - \epsilon)}$$

for large n

Chapter 3 (C4T)

(Strongly) typical set

Type of x^n

$$\mu_{x^n}(a) = \frac{\# \text{ occurrences of } a \text{ in } x^n}{n}$$

$$T_{\epsilon, \delta}^{(n)}(p_X) = \left\{ x^n \in \mathcal{X}^n : \begin{array}{l} | \mu_{x^n}(a) - p_X(a) | \leq \epsilon p_X(a) \\ \forall a \in \mathcal{X} \end{array} \right\}$$

Eg

$$\underline{a} \quad \underline{b} \quad \underline{a} \quad \underline{a} \quad c \quad \underline{b} \quad \underline{b} \quad \underline{a} \quad \underline{a} = x^9$$

$$M_{x^n}(a) = \frac{5}{9}$$

$$M_{x^n}(b) = \frac{3}{9}$$

$$M_{x^n}(c) = \frac{1}{9}$$

$$M_{x^n} = \begin{bmatrix} 5/9 \\ 3/9 \\ 1/9 \end{bmatrix}$$

$T_{\epsilon, S}^{(n)}(p_X)$ is a high probability set

$$M_{X^n}(a) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i = a\}}$$

$\begin{cases} 1 & \text{if } X_i = a \\ 0 & \text{if } X_i \neq a \end{cases}$

$$M_{X^n}(a) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i = a\}}$$

$$\mathbb{E} M_{X^n}(a) = \frac{1}{n} \sum_{i=1}^n (1 \times p_X(a) + 0 \times (1 - p_X(a)))$$

$$= p_X(a)$$

By WLLN, $P_{X^n} [X^n \notin T_{\epsilon, S}^{(n)}(p_X)] \rightarrow 0$

as $n \rightarrow \infty$

Properties of entropy

$$\textcircled{1} \quad H(X) \geq 0 \quad \forall p_X$$
$$H(X) = 0 \quad \text{iff } X \text{ is deterministic}$$

$\textcircled{2}$ $H(X)$ is a concave function of p_X .

$$\textcircled{3} \quad H(X_1, X_2) = H(X_1) + H(X_2|X_1)$$

Mutual information

$$\textcircled{1} \quad I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \\ = I(Y; X)$$

$$\textcircled{2} \quad I(X; Y) \geq 0 \quad \& \quad p_{XY} \text{ equality iff } X \perp Y$$

$$\textcircled{3} \quad I(X; Y) \text{ - convex in } p_{Y|X} \text{ for fixed } p_X \\ \text{ f } (p_X, p_{Y|X}) \text{ concave in } p_X \text{ for fixed } p_{Y|X}$$

$$\textcircled{4} \quad H(X) \geq H(X|Y)$$

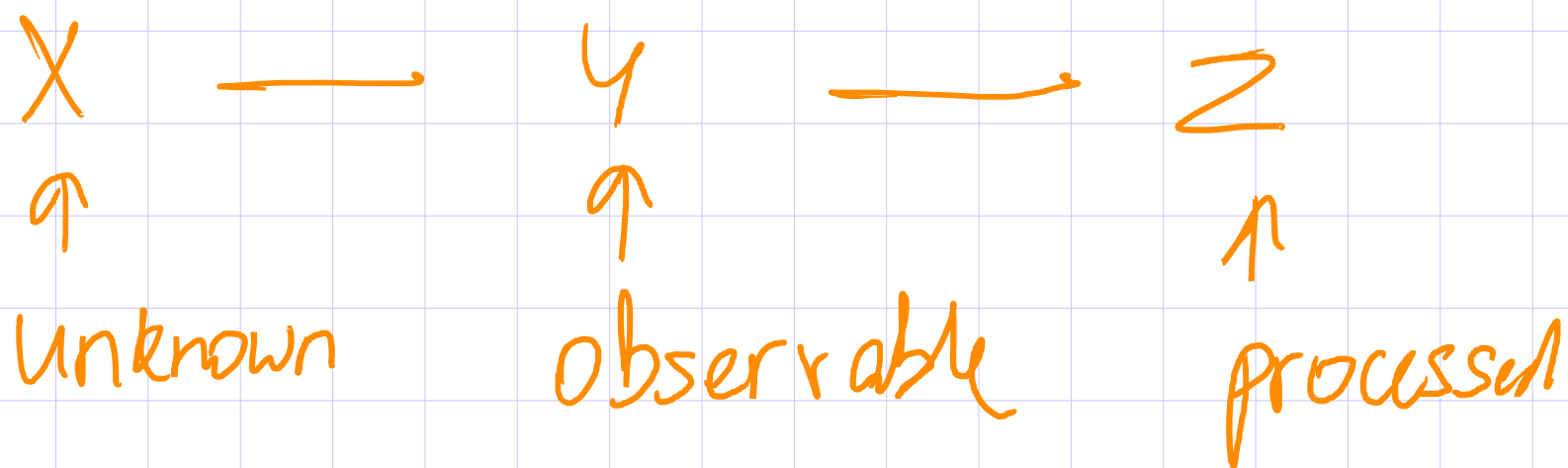
$$\textcircled{5} \quad I(X; YZ) = I(X; Y) + I(X; Z|Y)$$

Data processing inequality

$$X \leftrightarrow Y \leftrightarrow Z$$

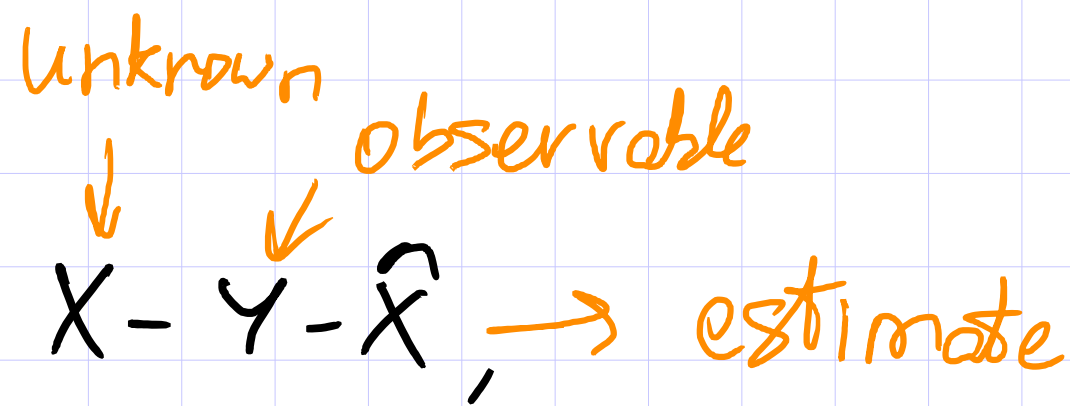
Markov chain

$$\Rightarrow I(X; Y) \geq I(X; Z)$$



Fano's inequality

For any Markov chain



$$H(X|Y) \leq H(X|\hat{X}) \leq H_2(p_e) + p_e \log_2 |\mathcal{X}|$$

where

$$p_e = \Pr[X \neq \hat{X}]$$

$$H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

(Chapter 2 of C&T)