# Chapter 4

# *Entropy Rates of a Stochastic Process*

Peng-Hua Wang

Graduate Inst. of Comm. Engineering

National Taipei University

# Chapter Outline

**Chap. 4** Entropy Rates of a Stochastic Process

**4.1** Markov Chains

**4.2** Entropy Rate

**4.3** Example: Entropy Rate of a Random Walk on a Weighted Graph

**4.4** Second Law of Thermodynamics

**4.5** Functions of Markov Chains

# 4.1 Markov Chains

# Stationary

**Definition (Stationary)** A stochastic process is said to be stationary if

$$\begin{aligned}\Pr\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \\ = \Pr\{X_{1+\ell} = x_1, X_{2+\ell} = x_2, \dots, X_{n+\ell} = x_n\}\end{aligned}$$

for every $n$ and every shift $\ell$.

- the joint distribution of any subset of the sequence of random variables is invariant with respect to shifts in the time index.

# Markov chain

**Definition (Markov chain)** A discrete stochastic process $X_1, X_2, \ldots$ is said to be a Markov chain or a Markov process if for $n = 1, 2, \ldots,$

$$\begin{aligned} \Pr\{X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1\} \\ = \Pr\{X_{n+1} = x_{n+1} | X_n = x_n\}. \end{aligned}$$

- The joint pmf can be written as

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_2)\cdots p(x_n|x_{n-1}).$$

**Definition (Time invariant)** The Markov chain is said to be time invariant if the transition probability $p(x_{n+1}|x_n)$,

$$\Pr\{X_{n+1} = b | X_n = a\} = \Pr\{X_2 = b | X_1 = a\} \quad \text{for all } a, b \in \mathcal{X}.$$

# Markov chain

- We will assume that the Markov chain is time invariant.
- $X_n$ is called the **state** at time $n$.
- A time invariant Markov chain is characterized by its initial state and a probability transition matrix $P = [P_{ij}], i, j \in \{1, 2, \ldots, m\}$, where $P_{i,j} = \Pr\{X_{n+1} = j | X_n = i\}$.
- The pmf at time $n+1$ is

$$p(x_{n+1}) = \sum_{x_n} p(x_n) P_{x_n x_{n+1}}$$

- A distribution on the states such that the distribution at time $n+1$ is the same as the distribution at time $n$ is called a **stationary distribution**.

# Example 4.1.1

Consider a two-state Markov chain with a probability transition matrix

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

Find its stationary distribution and entropy.

**Solution.** Let $\mu_1, \mu_2$ be the stationary distribution.

$$\mu_1 = \mu_1(1-\alpha) + \mu_2\beta$$
$$\mu_2 = \mu_1\alpha + \mu_2(1-\beta)$$

and

$$\mu_1 + \mu_2 = 1.$$

# 4.2 Entropy Rate

# Entropy Rate

**Definition (Entropy Rate)** The entropy of a random process $\{X_i\}$ is defined by

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n).$$

**Definition (Conditional Entropy Rate)** The entropy of a random process $\{X_i\}$ is defined by

$$H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n | X_1, X_2, \ldots, X_{n-1}).$$

# Entropy Rate

- If $X_1, X_2, \ldots$ are i.i.d. random variables. Then

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = \lim \frac{nH(X_1)}{n} = H(X_1).$$

- If $X_1, X_2, \ldots$ are independent but not identical distributed

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i).$$

- We can choose a sequence of distributions on $X_1, X_2 \ldots$ such that the limit does not exist.

# Entropy Rate

**Theorem 4.2.2** For a stationary stochastic process, $H(X_n|X_{n-1},\ldots,X_1)$ is nonincreasing in $n$ and has a limit $H'(\mathcal{X})$.

**Proof.**

$$\begin{aligned} &H(X_{n+1}|X_1,X_2,\ldots,X_n)\\ \le &H(X_{n+1}|X_2,\ldots,X_n) \quad (\text{conditioning reduce entropy})\\ = &H(X_n|X_1,\ldots,X_{n-1}) \quad (\text{stationary}) \end{aligned}$$

Since $H(X_n|X_{n-1},\ldots,X_1)$ is nonnegative and decreasinging, it has a limit $H'(\mathcal{X})$.

# Entropy Rate

**Theorem 4.2.1** For a stationary stochastic process, both $H(\mathcal{X})$ and $H'(\mathcal{X})$ exist and are equal.

$$H(\mathcal{X}) = H'(\mathcal{X}).$$

**Proof.** By the chain rule,

$$\frac{1}{n}H(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} H(X_i|, X_{i-1}, \ldots, X_1),$$

that is, the entropy rate is the time average of the conditional entropies. Since the conditional entropies has a limit $H'(\mathcal{X})$. We conclude that the entropy rate has the same limit by Theorem of Cesáro mean. $\square$

# Cesáro mean

**Theorem (Cesáro mean)** If $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$.

**Proof.** Let $\epsilon > 0$. Since $a_n \to a$, there exists a number $N$ such that $|a_n - a| \leq \epsilon$ for $n > N$. Hence,

$$|b_n - a| = \left| \frac{1}{n}\sum_{i=1}^{n}(a_i - a) \right| \leq \frac{1}{n}\sum_{i=1}^{n}|(a_i - a)|$$

$$\leq \frac{1}{n}\sum_{i=1}^{N}|(a_i - a)| + \frac{n-N}{n}\epsilon \leq \frac{1}{n}\sum_{i=1}^{N}|(a_i - a)| + \epsilon \leq \epsilon$$

when $n$ is large enough. □