4. Markov Chains

Markov chain: A discrete-time stochastic process $\{X_1, X_2, ...\}$
if for all $n = 1, 2, \cdots,$

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, .., X_1 = x_1] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n]$$

for all $x^{n+1} \in \mathcal{X}^{n+1}$.

- A Markov chain is time invariant if the conditional probability $p(x_{n+1}|x_n)$ does not depend on n

$$\mathbb{P}[X_{n+1} = b | X_n = a] = \mathbb{P}[X_2 = b | X_1 = a]$$

- A time invariant Markov Chain with $\mathcal{X} = \{1, 2, ..., M\}$ is characterized by an $M \times M$
probability transition matrix P where

$$P_{i,j} = \mathbb{P}[X_{n+1} = j | X_n = i] \qquad i, j \in \{1, 2, .., M\}$$

The vector of probabilities at time n is denote by

$$\mu_n = (\mathbb{P}[X_n = 1], \mathbb{P}[X_n = 2], ..., \mathbb{P}[X_n = M])$$

and is updated according to

$$\mu_{n+1} = \mu_n P$$

A time-invariant Markov process is completely specified by three items:

1. the set of all states $\mathcal{X}$
2. the $|\mathcal{X}| \times |\mathcal{X}|$ transition probability matrix P
3. the starting distribution μ_1.

A distribution μ is a **stationary distribution** of a time-invariant Markov process if

$$\mu = \mu P$$

Theorem: Consider a time-invariant Markov process that is irreducible and aperiodic. Then

1. The stationary distribution μ is unique
2. Independently of the starting distribution μ_1, the distribution μ_n will converge to the
stationary distribution as μ as $n \to \infty$.
3. The Markov process is stationary if, and only if, the starting distribution μ_1 is chosen
to be the steady state distribution.

Theorem: For a stationary time-invariant Markov process. the entropy rate is given by

$$H(\mathcal{X}) = H(X_2|X_1)$$

where the conditional entropy is calculated using the stationary distribution.