Chapter 2: Information Measures

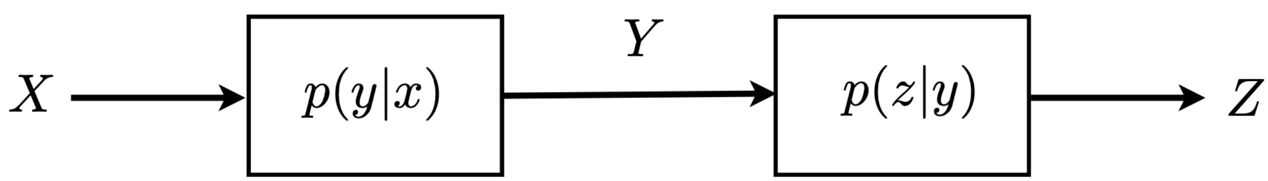
2.1. Independence and Markov Chain

April 18, 2020

Notations

- X: discrete rarndom variable taking values in $\mathcal{X}$.
- $\{p_X(x)\}$: probability distribution of X.
- $\mathcal{S}_X$: support of X, i.e., $\{x \in \mathcal{X} | p_X(x) > 0\}$
    - if $\mathcal{S}_X = \mathcal{X}$, we say that p is **strictly positive**.

Proposition 2.4: Conditional Independence



The above is the graph showing how X, Y and Z are related, for the case: $X \perp Z | Y$, in other words, X is independent of $Z \perp Y$.

- $p(x, y, z) = p(x)p(y|x)p(z|y)$
- $p(x, y, z) = p(x, y)p(z|y)$
- $p(x, y, z) = \frac{p(x,y)p(y,z)}{p(y)}$
- Think of passing X through the channel $p(y|x)$ to obtain Y, and pass Y through the channel $p(z|y)$ to oabtain Z.

Think it alternatively:

- if $X \perp Z | Y$:
    - $p(z|x, y) = p(z|y)$
    - $p(x, y, z) = p(x, y)p(z|y)$
- if not $X \perp Z | Y$:
    - $p(x, y, z) = p(x, y)p(z|x, y)$

Proposition 2.5: Condition of Condional Independence

$X \perp Z|Y$ if and only if $p(x,y,z) = f(x,y)g(z,y)$, for all x, y and z such that $p(y) > 0$.

Note:

- $f(\cdot)$ and $g(\cdot)$ are not necessarily probability functions. They just need to be functions that depends only on the specified variables.
- $0 \leq p(x,y,z) \leq p(x,y) \leq p(y)$
  - $p(y) = \sum_x p(x,y)$
  - $p(x,y) = \sum_z p(x,y,z)$

Proof of "If":

- Step 1:

$$\begin{aligned} p(x,y) &= \sum_z p(x,y,z) \\ &= \sum_z a(x,y)b(y,z) \\ &= a(x,y)\sum_z b(y,z) \end{aligned}$$

- Step 2:

$$\begin{aligned} p(y,z) &= \sum_x p(x,y,z) \\ &= \sum_x a(x,y)b(y,z) \\ &= b(y,z)\sum_x a(x,y) \end{aligned}$$

- Step 3:

  $p(y) = \sum_z p(y,z) = \left(\sum_x a(x,y)\right)\left(\sum_z b(y,z)\right) > 0$

- Step 4:

  $\frac{p(x,y)p(y,z)}{p(y)} = \frac{(a(x,y)\sum_z b(y,z))(b(y,z)\sum_x a(x,y))}{(\sum_x a(x,y))(\sum_z b(y,z))} = f(x,y)b(y,z) = p(x,y,z)$

- Step 5

  If $p(y) = 0$, given $0 \leq p(x,y,z) \leq p(x,y) \leq p(y) = 0$, we will have $p(x,y,z) = 0$. This is precisely the second case of proposition 2.4.

Definition 2.6: Markov Chain

For random variables $X_1, X_2, \cdots, X_n$, where $n \geq 3$, $X_1 \to X_2 \to \cdots \to X_n$ forms a Markov Chain if:

$$p(x_1, x_2, \cdots, x_n) = \\ \begin{cases} p(x_1, x_2)\, p(x_3|x_2) \cdots p(x_n|x_{n-1}) & \text{if } p(x_2), p(x_3), \cdots, p(x_{n-1}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Remark: $X_1 \to X_2 \to X_3$ is equivalent to $X_1 \perp X_2 | X_3$.

Proposition 2.7: Reversibe Markov Chain

$X_1 \to X_2 \to \cdots \to X_n$ forms a Markov Chain if and only if $X_n \to X_{n-1} \to \cdots \to X_1$ forms a Markov Chain.

Proposition 2.8: Clustered Markov Chain

$X_1 \to X_2 \to \cdots \to X_n$ forms a Markov Chain if and only if:

> $$X_1 \to X_2 \to X_3$$
> $$(X_1, X_2) \to X_3 \to X_4$$
> $$\vdots$$
> $$(X_1, X_2, \cdots, X_{n-2}) \to X_{n-1} \to X_n$$

forms a markov chain.

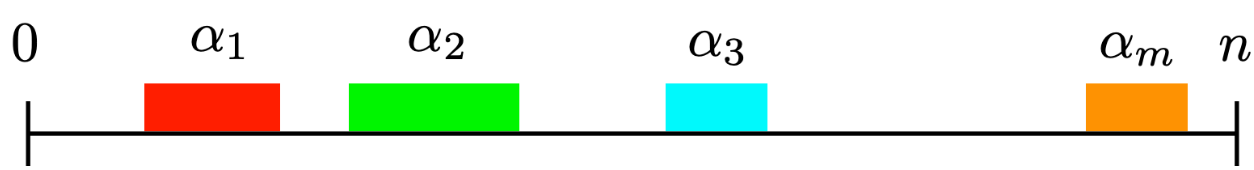
Proposition 2.9: Factorized Markov Chain

$X_1 \to X_2 \to \cdots \to X_n$ forms a Markov Chain if and only if:

> $$p(x_1, x_2, \cdots, x_n) = f_1(x_1, x_2) f_2(x_2, x_3) \cdots f_{n-1}(x_{n-1}, x_n)$$

for all $x_1, x_2, \cdots, x_n$ such that $p(x_1), p(x_2), \cdots, p(x_n) > 0$.

Proposition 2.10: Markov Subchain



Proposition 2.10 (Markov subchains) Let $\mathcal{N}_n = \{1, 2, \cdots, n\}$ and let $X_1 \to X_2 \to \cdots \to X_n$ form a Markov chain. For any subset α of $\mathcal{N}_n$, denote $(X_i, i \in \alpha)$ by X_α. Then for any disjoint subsets $\alpha_1, \alpha_2, \cdots, \alpha_m$ of $\mathcal{N}_n$ such that

$$k_1 < k_2 < \cdots < k_m$$

for all $k_j \in \alpha_j$, $j = 1, 2, \cdots, m$,

$$X_{\alpha_1} \to X_{\alpha_2} \to \cdots \to X_{\alpha_m}$$

forms a Markov chain. That is, a subchain of $X_1 \to X_2 \to \cdots \to X_n$ is also a Markov chain. (Exercise)
