Information Theory: Lecture Notes 2

zqy1018

June 9, 2020

Contents

1 Information Diagram

We can use the information diagram to find out the relationship between different information measures.

1.1 3 Random Variables

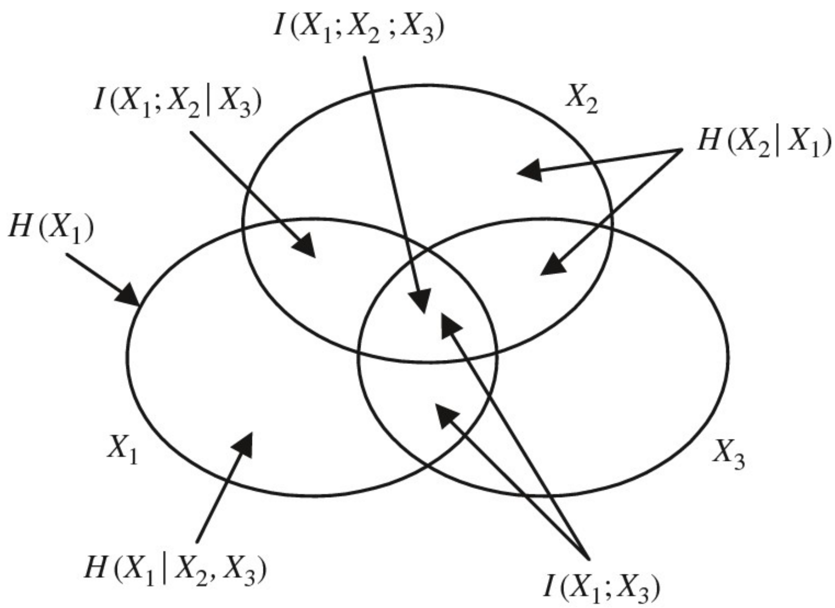

Figure 1: The information diagram of 3 random variables. Picture credit: *Information Theory and Network Coding* by Raymond W. Yeung.

We can see that (FILL IN HERE)

1.2 4 Random Variables

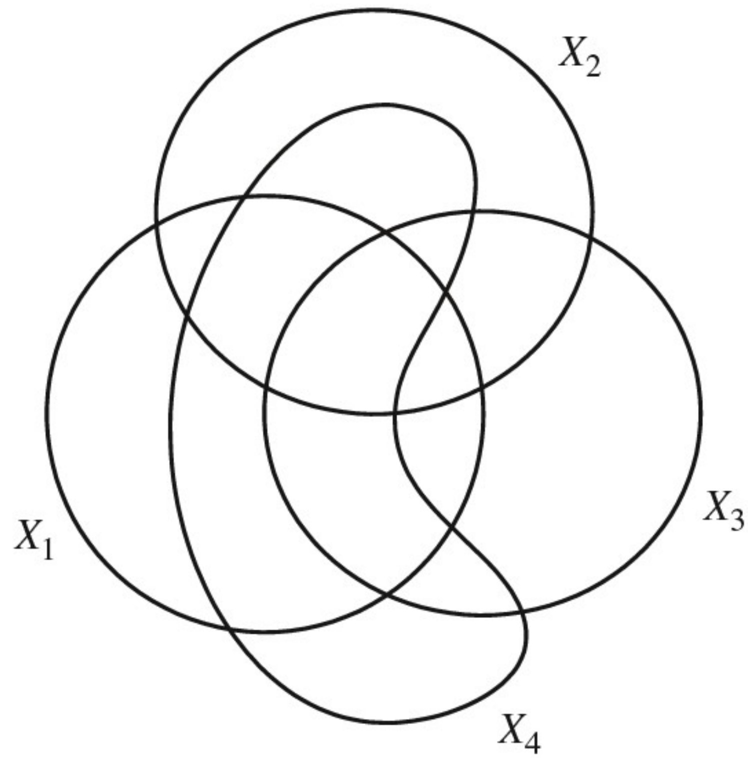

Figure 2: The information diagram of 4 random variables. Picture credit: *Information Theory and Network Coding* by Raymond W. Yeung.

1.3 Special Case: Markov Chain

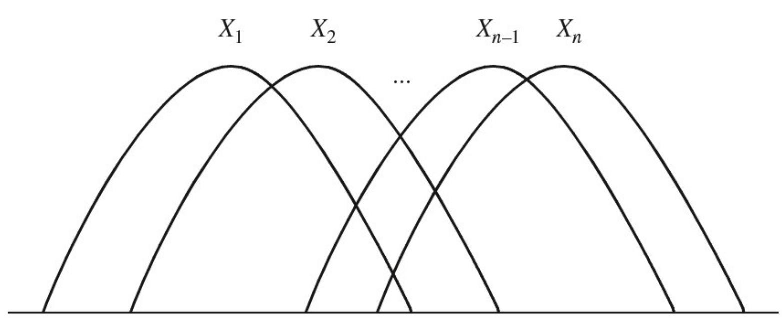

Figure 3: The information diagram of a Markov chain $X_1 \to X_2 \to \cdots \to X_n$. Picture credit: *Information Theory and Network Coding* by Raymond W. Yeung.

Here, the area of every atomic piece is non-negative.

2 Data Processing Inequality

Theorem 1. (Data Processing Inequality) If $X \to Y \to Z$, then $I(X;Y) \geq I(X;Z)$.

Proof. We have

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$
$$I(X;Y,Z) = I(X;Y) + I(X;Z|Y)$$

Since $I(X;Z|Y) = 0$, we have $I(X;Y) - I(X;Z) = I(X;Y|Z) \geq 0$. □

Corollary. If $X \to Y \to Z$, then $I(X;Y) \geq I(X;Y|Z)$.

Corollary. If $X \to Y \to Z$, then $H(X|Z) \geq H(X|Y)$.

Note. The conditional mutual information is not necessarily less than or equal to the mutual information. For example, assume X, Y are independent and uniformly distributed on $\{0,1\}$. Let $Z = X \operatorname{xor} Y$. Then $I(X;Y|Z) > I(X;Y)$.

3 Fano's Inequality

3.1 Background

Here we denote X as a cause and Y as a result. We want to estimate X by observing Y. Thus we have a estimator function $\hat{X}(Y)$.

Remark. This function can be nondeterministic, i.e. it can be a random variable. And we do not restrict the alphabet of $\hat{X}$ to be equal to $\mathcal{X}$.

We can see that $X \to Y \to \hat{X}$. Let $P_e = p(X \neq \hat{X})$. Now the problem is: how to bound the discrepancy between X and $\hat{X}$?

3.2 Fano's Inequality

Theorem 2. (Fano's Inequality) For any estimator $\hat{X}$ such that $X \to Y \to \hat{X}$, with $P_e = p(X \neq \hat{X})$, we have

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y)$$

Proof. For the latter inequality, we have

$$I(X;Y) \geq I(X;\hat{X}) \implies H(X|\hat{X}) \geq H(X|Y)$$

For the former inequality, use an error indicator E such that if $X = \hat{X}$, then $E = 0$; otherwise $E = 1$. Then

$$H(E,X|\hat{X}) = H(X|\hat{X}) + H(E|X,\hat{X}) = H(E|\hat{X}) + H(X|E,\hat{X})$$

Since E is determined by X and $\hat{X}$, $H(E|X,\hat{X}) = 0$. Then

$$\begin{aligned}
H(X|\hat{X}) &= H(E|\hat{X}) + H(X|E,\hat{X}) \\
&\leq H(E) + H(X|E,\hat{X}) \\
&= H(P_e) + P_e H(X|E=1,\hat{X}) + (1-P_e)H(X|E=0,\hat{X}) \\
&= H(P_e) + P_e H(X|E=1,\hat{X}) \\
&\leq H(P_e) + P_e H(X) \\
&\leq H(P_e) + P_e \log|\mathcal{X}|
\end{aligned}$$

□

Corollary. If $\hat{X}$ has the alphabet $\mathcal{X}$, then the inequality can be strengthened:

$$H(P_e) + P_e \log(|\mathcal{X}|-1) \geq H(X|\hat{X})$$

Proof. Just notice that when $E = 1$, then X will have only $|\mathcal{X}| - 1$ possibilities. □

4 Applications of Information Measures

Information measures are useful.

Example 1. (Causality)

In information theory, we can use random variables to denote the conditions given in the problem, and apply the techniques in information measures to check whether a given condition is satisfied.

For example, given $X \perp Y|Z$ and $X \perp Z$, we can prove $X \perp Y$ by showing that $I(X;Y|Z) = I(X;Z) = 0 \implies I(X;Y) = 0$.

Example 2. (Information-theoretic security)

We can use information measures to find the condition for good security.

For example, in a simple cryptosystem, let X be the plain text, Y be the cipher text, and Z be the key in a secret.

Since Y is generated from X and Z, we have $H(Y|X,Z) = 0$. And since we can restore X with Y and Z, $H(X|Y,Z) = 0$. Then they implies $I(X;Y) \geq H(X) - H(Z)$.

If we want to achieve perfect security, i.e. X is independent of Y, then $I(X;Y) = 0 \implies H(X) \leq H(Z)$.

5 Proof via Convexity and Concavity

5.1 Log-sum Inequality

Theorem 3. (Log-sum Inequality) For non-negative numbers $a_1, \cdots, a_n, b_1, \cdots b_n$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

with equality iff $\frac{a_i}{b_i}$ is a constant for any i.

5.2 Convexity and Concavity of Information Measures

(FILL IN HERE)

Acknowledgment

The contents are mainly based on the course materials of CS258, 2020 Spring, Shanghai Jiao Tong University and *Elements of Information Theory* by Thomas M. Cover and Joy A. Thomas. Some contents are also based on *Information Theory and Network Coding* by Raymond W. Yeung.