CIE6020/MAT3350
Selected Topics in Information Theory

Lecture 1: Entropy

10 Janurary 2019

The Chinese University of Hong Kong, Shenzhen

Information

- Instructor: Shenghao Yang
  - office: Cheng Dao 606
  - email: shyang@cuhk.edu.cn
  - phone: 842 73827
  - office hour: Thursday/Friday, 11:20 - 12:00

Information

- Instructor: Shenghao Yang
  - office: Cheng Dao 606
  - email: shyang@cuhk.edu.cn
  - phone: 842 73827
  - office hour: Thursday/Friday, 11:20 - 12:00
- Lecture: Thursday/Friday, 10:00 - 11:20
- Classroom: 101 Zhi Xin

Recommended Books

- Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2nd. John Wiley & Sons, Inc, 2006
- David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003
- Raymond W. Yeung. *Information Theory and Network Coding*. Springer, 2008
- Abbas El Gammal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011
- F. J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 2007
- Tom Richardson and Ruediger Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006

Evaluation

- CIE6020
  - Homework (30%)
  - Course Project (30%)
  - Final Exam (40%)
- MAT3350
  - Homework (25%)
  - Course Project (25%)
  - Final Exam (50%)

Project Information

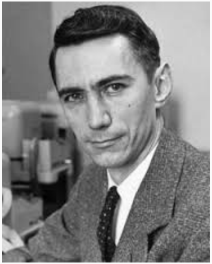
- A list of papers will be provided.
- Each student involves in one and only one project.
- Bi-weakly reports, midterm presentation, final report.

Why Learn Information Theory?

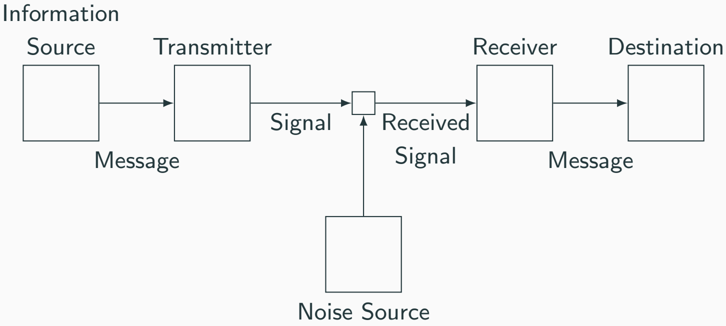
- IT provides high-level guidance about the information system design:
  - WiFi, 3G, 4G, 5G, ...
  - Distributed storage, content distribution network
  - Wireless ad hoc/mesh/sensor networks, Internet, IoT
  - Distributed/parallel computing
- It helps us to answer some common questions
  - What is information?
  - What does “entropy” mean?
  - How small can we compress a file, and how fast can we transmit information using LTE?
- IT finds applications in all major science and engineering sectors.

Claude E. Shannon (1916-2001)

- 1948, A Mathematical Theory of Communication (full article)
- 1937, founding digital circuit design theory
- cryptography
- artificial intelligence (see a demonstration)



Shannon's Diagram of a general Communication System



Background

Probability

- Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, also called *alphabets*.
- Let X and Y be discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively.
- Probability mass function: $p_X(x) = \Pr\{X = x\}$, $x \in \mathcal{X}$.
- We also denote the probability distribution by p rather than p_X when the random variable referred to is clear from context.
- Joint distribution: $p(x, y) = \Pr\{X = x, Y = y\}$.
- Conditional distribution: $p(x|y) = \frac{p(x,y)}{p(y)}$.
- If $(X, Y) \sim p(x, y)$ are independent, $p(x, y) = p(x)p(y)$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

Entropy

What is information

- Information is about uncertainty.
- Entropy is a measure of the uncertainty of a random variable.
- Entropy arises naturally as the fundamental limits of *source coding*.

Entropy

Definition
The *entropy* $H(X)$ of a discrete random variable X is defined by

$$H(X) = -\sum_{x} p(x) \log p(x).$$

Remark

1. The summation is over the support of X.
2. The log is to the base 2 and the unit of entropy is *bit*.
3. $H(X)$ depends only on $p(x)$, not on the actual values of x—entropy is independent of the alphabet $\mathcal{X}$. So we also write $H(X)$ as $H(p)$.

Other Forms

- Expectation form $H(X) = -\mathbb{E}\log(p(X))$
- Binary entropy function: $H(p) = -p\log p - (1-p)\log(1-p)$

Properties

- $H(X) \geq 0$ where equality holds iff X is a deterministic.
- $H(X) \leq \log |\mathcal{X}|$ where $\mathcal{X}$ is the alphabet of X. The equality holds iff X is uniformly distributed on $\mathcal{X}$.

Joint Entropy

- The entropy of a pair of random variables (X, Y) with alphabets $\mathcal{X}$ and $\mathcal{Y}$ is also defined by considering (X, Y) as a single random variable over $\mathcal{X} \times \mathcal{Y}$. For convenience, we write $H(X, Y) = H((X, Y))$.
- The joint entropy $H(X, Y)$ of a pair of discrete random variable (X, Y) with a joint distribution $p(x, y)$ is defined as

$$H(X, Y) = -\sum_{x} \sum_{y} p(x, y) \log p(x, y) = -\mathbb{E} \log p(X, Y).$$

Conditional Entropy and Mutual Information

Conditional Entropy

- For random variables X and Y, the *conditional entropy* $H(Y|X)$ is defined as

$$H(Y|X) = -\sum_{x,y} p(x,y) \log p(y|x) = -\mathbb{E} \log p(Y|X).$$

- Denote

$$H(Y|X=x) = H(p_{Y|X}(\cdot|x)) = -\sum_{y} p(y|x) \log p(y|x).$$

- We can write

$$H(Y|X) = \sum_{x} p(x) H(Y|X=x).$$

- In other words, the conditional entropy is the expectation of the entropy of the conditional distribution of Y given $X = x$.

Basic Properties

- $H(Y|X) \geq 0$ with equality iff Y is a function of X (over the support of X).
- (Chain rule) $H(X,Y) = H(X) + H(Y|X)$.
- $H(Y|X) \leq H(Y)$ with equality iff X and Y are independent. In other words, conditioning reduces entropy.

Mutual Information

Definition
The *mutual information* between random variables X and Y is defined as

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \mathbb{E} \log \frac{p(X,Y)}{p(X)p(Y)}.$$

Remark

1. $I(X;Y)$ is symmetrical in X and Y.
2. $I(X;X) = H(X)$: observing X can get all the information of X.
3. $I(X;Y) \geq 0$ (Log-sum inequality).
4. $I(X;Y)$ only depends on the joint distribution $p_{X,Y}$, so we also write $I(X;Y) = I(p_{X,Y})$.

Relations

- We have the following equalities:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

- If the alphabets are not finite, the above equalities hold provided that all the entropies and conditional entropies are finite.

Information Diagram of Two Random Variables