Lecture 3: Shannon's Theory, Perfect Secrecy, and the One-Time Pad

–Cryptographic Algorithms and Protocols

Instructor: Xiujie Huang 黄秀姐

Office: Nanhai Building, Room 411

E-mail: t_xiujie@jnu.edu.cn

Department of Computer Science

School of Information Science and Technology

Jinan University

Outline

Outline

Elementary Probability Theory

Definition 3.1: Discrete Random Variable (on Page 63)

A discrete random variable, say $\mathbf{X}$, consists of a finite set $\mathcal{X}$ and a probability distribution defined on $\mathcal{X}$.

- $\mathbf{Pr}[x] \triangleq \mathbf{Pr}[\mathbf{X} = x]$, which is abbreviated to $p(x)$.
- $p(x) \geq 0$ for any $x \in \mathcal{X}$;
- $\sum_{x \in \mathcal{X}} p(x) = 1$;
- For the event/subset $E \subseteq \mathcal{X}$,

$$p(E) \triangleq \mathbf{Pr}[x \in E] = \sum_{x \in E} p(x). \tag{1.1}$$

Examples

1. Flipping a fair coin: $\mathbf{Pr}[Head] = \mathbf{Pr}[Tail] = 0.5$, i.e., $p(1) = p(0) = 0.5$
2. Throwing a pair of dice. (See Example 3.1 on Page 63)

Elementary Probability Theory

Definition 3.2: Joint and Conditional Probabilities (on Page 63)

Let $\mathbf{X}$ and $\mathbf{Y}$ be two discrete random variables defined on two finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively.

- The joint probability $\mathbf{Pr}[x, y] \stackrel{\Delta}{=} \mathbf{Pr}[\mathbf{X} = x, \mathbf{Y} = y]$, abbreviated to $p(x, y)$;
- The conditional probability $\mathbf{Pr}[x|y] \stackrel{\Delta}{=} \mathbf{Pr}[\mathbf{X} = x|\mathbf{Y} = y]$, abbreviated to $p(x|y)$;
- If $p(x, y) = p(x)p(y)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, then X and Y are independent.

Theorem 3.1 (Bayes' Theorem) (on Page 64)

If $p(y) > 0$, then

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)}. \tag{1.2}$$

Corollary 3.2 (on Page 64)

$\mathbf{X}$ and $\mathbf{Y}$ are independent $\Leftrightarrow$ $p(x|y) = p(x)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Outline

Probabilities in a Cryptosystem

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is specified, and a particular key $k \in \mathcal{K}$ is used for only one encryption.

Probabilities in a Cryptosystem

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is specified, and a particular key $k \in \mathcal{K}$ is used for only one encryption.

Probabilities in a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ (I)

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

1. the *a priori* probability of the plaintext: $p(x) = \mathbf{Pr}[\mathbf{X} = x]$;
2. the probability of the key: $p(K) = \mathbf{Pr}[\mathbf{K} = K]$;

Probabilities in a Cryptosystem

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is specified, and a particular key $k \in \mathcal{K}$ is used for only one encryption.

Probabilities in a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ (I)

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

1. the *a priori* probability of the plaintext: $p(x) = \mathbf{Pr}[\mathbf{X} = x]$;
2. the probability of the key: $p(K) = \mathbf{Pr}[\mathbf{K} = K]$;
3. the key and the plaintext are independent: $p(x, K) = p(x)p(K)$;

Probabilities in a Cryptosystem

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is specified, and a particular key $k \in \mathcal{K}$ is used for only one encryption.

Probabilities in a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ (I)

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

1. the *a priori* probability of the plaintext: $p(x) = \mathbf{Pr}[\mathbf{X} = x]$;
2. the probability of the key: $p(K) = \mathbf{Pr}[\mathbf{K} = K]$;
3. the key and the plaintext are independent: $p(x, K) = p(x)p(K)$;
4. Then, for a key $K \in \mathcal{K}$, define the set of all possible ciphertexts (if K is the key) as

$$C(K) = \{e_K(x) : x \in \mathcal{P}\}$$

For every ciphertext $y \in \mathcal{C}$,

$$p(y) = \sum_{\{K : y \in C(K)\}} p(K)p(x = d_k(y))$$

给定一个密文 y，这个密 文出现的概率等于所有可能产生这个密文的密钥和对应的明文的概率之和

Probabilities in a Cryptosystem

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is specified, and a particular key $k \in \mathcal{K}$ is used for only one encryption.

Probabilities in a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ (I)

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

1. the *a priori* probability of the plaintext: $p(x) = \mathbf{Pr}[\mathbf{X} = x]$;
2. the probability of the key: $p(K) = \mathbf{Pr}[\mathbf{K} = K]$;
3. the key and the plaintext are independent: $p(x, K) = p(x)p(K)$;
4. Then, for a key $K \in \mathcal{K}$, define the set of all possible ciphertexts (if K is the key) as

$$C(K) = \{e_K(x) : x \in \mathcal{P}\}$$

给定一个密文 y，这个密 文出现的概率等于所有可能产生这个密文的密钥和对应的明文的概率之和

For every ciphertext $y \in \mathcal{C}$,

$$p(y) = \sum_{\{K : y \in C(K)\}} p(K)p(x = d_k(y)) = \sum_{\{x, K : e_K(x) = y\}} p(K)p(x)$$

Probabilities in a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ (II)

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

5. For any $x \in \mathcal{P}$ and $y \in \mathcal{C}$

$$p(y|x) = \sum_{\{K:x=d_K(y)\}} p(K) = \sum_{\{K:y=e_K(x)\}} p(K)$$

Probabilities in a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ (II)

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

5. For any $x \in \mathcal{P}$ and $y \in \mathcal{C}$

$$p(y|x) = \sum_{\{K:x=d_K(y)\}} p(K) = \sum_{\{K:y=e_K(x)\}} p(K)$$

6. The *a posteriori* probability of the plaintext x given the ciphertext y:

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)} = \frac{p(x) \times \sum_{\{K:x=d_K(y)\}} p(K)}{\sum_{\{K:y\in C(K)\}} p(K)p(x = d_K(y))}$$

Probabilities in a Cryptosystem

Example 3.3

Let $\mathcal{P} = \{a, b\}$ with $p(a) = 1/4$ and $p(b) = 3/4$. Let $\mathcal{K} = \{K_1, K_2, K_3\}$ with $p(K_1) = 1/2$, $p(K_2) = p(K_3) = 1/4$. Let $\mathcal{C} = \{1, 2, 3, 4\}$ with

$e_{K_1}(a) = 1$ and $e_{K_1}(b) = 2$;

$e_{K_2}(a) = 2$ and $e_{K_2}(b) = 3$;

$e_{K_3}(a) = 3$ and $e_{K_3}(b) = 4$.

Q: Compute the probabilities of the ciphertexts, and the a posteriori probabilities of the plaintexts.

Probabilities in a Cryptosystem

Example 3.3

Let $\mathcal{P} = \{a, b\}$ with $p(a) = 1/4$ and $p(b) = 3/4$. Let $\mathcal{K} = \{K_1, K_2, K_3\}$ with $p(K_1) = 1/2$, $p(K_2) = p(K_3) = 1/4$. Let $\mathcal{C} = \{1, 2, 3, 4\}$ with
$e_{K_1}(a) = 1$ and $e_{K_1}(b) = 2$;
$e_{K_2}(a) = 2$ and $e_{K_2}(b) = 3$;
$e_{K_3}(a) = 3$ and $e_{K_3}(b) = 4$.

Q: Compute the probabilities of the ciphertexts, and the a posteriori probabilities of the plaintexts.

A: $p(y = 1) = \frac{1}{8}$, $p(y = 2) = \frac{7}{16}$, $p(y = 3) = \frac{1}{4}$, $p(y = 4) = \frac{3}{16}$.

Probabilities in a Cryptosystem

Example 3.3

Let $\mathcal{P} = \{a, b\}$ with $p(a) = 1/4$ and $p(b) = 3/4$. Let $\mathcal{K} = \{K_1, K_2, K_3\}$ with $p(K_1) = 1/2$, $p(K_2) = p(K_3) = 1/4$. Let $\mathcal{C} = \{1, 2, 3, 4\}$ with
$e_{K_1}(a) = 1$ and $e_{K_1}(b) = 2$;
$e_{K_2}(a) = 2$ and $e_{K_2}(b) = 3$;
$e_{K_3}(a) = 3$ and $e_{K_3}(b) = 4$.

Q: Compute the probabilities of the ciphertexts, and the a posteriori probabilities of the plaintexts.

A: $p(y = 1) = \frac{1}{8}$, $p(y = 2) = \frac{7}{16}$, $p(y = 3) = \frac{1}{4}$, $p(y = 4) = \frac{3}{16}$.

$p(a|y = 1) = 1$, $p(b|y = 1) = 0$, $p(a|y = 2) = \frac{1}{7}$, $p(b|y = 2) = \frac{6}{7}$,
$p(a|y = 3) = \frac{1}{4}$, $p(b|y = 3) = \frac{3}{4}$, $p(a|y = 4) = 0$, $p(b|y = 4) = 1$.

Perfect Secrecy

Definition 3.3: Perfect Secrecy (完全保密)(on Page 66)

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ has perfect secrecy if $p(x|y) = p(x)$ for all $x \in \mathcal{P}$ and $y \in \mathcal{C}$.

Perfect Secrecy

Definition 3.3: Perfect Secrecy (完全保密)(on Page 66)

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ has perfect secrecy if $p(x|y) = p(x)$ for all $x \in \mathcal{P}$ and $y \in \mathcal{C}$.

Cryptosystem 1.1 Shift Cipher

Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$. For $0 \leq k \leq 25$, define

1. Encryption rule e_k : $e_k(x) = (x + k) \mod 26$
2. Decryption rule d_k : $d_k(y) = (y - k) \mod 26$

Remark: A key $k \in \mathcal{K}$ is used for only one encryption.

Perfect Secrecy

Definition 3.3: Perfect Secrecy (完全保密)(on Page 66)

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ has perfect secrecy if $p(x|y) = p(x)$ for all $x \in \mathcal{P}$ and $y \in \mathcal{C}$.

Cryptosystem 1.1 Shift Cipher

Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$. For $0 \leq k \leq 25$, define

1. Encryption rule e_k : $e_k(x) = (x + k) \mod 26$
2. Decryption rule d_k : $d_k(y) = (y - k) \mod 26$

Remark: A key $k \in \mathcal{K}$ is used for only one encryption.

Theorem 3.3 (on Pages 66-67)

Suppose the 26 keys in the Shift Cipher are used with equal probability $1/26$. Then for any plaintext probability distribution, the Shift Cipher has perfect secrecy.

Perfect Secrecy

Definition 3.3: Perfect Secrecy (完全保密)(on Page 66)

A cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ has perfect secrecy if $p(x|y) = p(x)$ for all $x \in \mathcal{P}$ and $y \in \mathcal{C}$.

Cryptosystem 1.1 Shift Cipher

Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$. For $0 \leq k \leq 25$, define

1. Encryption rule e_k : $e_k(x) = (x + k) \mod 26$
2. Decryption rule d_k : $d_k(y) = (y - k) \mod 26$

Remark: A key $k \in \mathcal{K}$ is used for only one encryption.

Theorem 3.3 (on Pages 66-67)

Suppose the 26 keys in the Shift Cipher are used with equal probability $1/26$. Then for any plaintext probability distribution, the Shift Cipher has perfect secrecy. $\Leftrightarrow$ "unbreakable"

Perfect Secrecy of the Shift Cipher

Proof of Theorem 3.3

To prove the perfect secrecy of the Shift Cipher, we'd like to prove $p(y|x) = p(y)$ for all x and y since $p(x|y) = \frac{p(x)p(y|x)}{p(y)}$.

(1) $p(y|x) = p(k = (y - x) \mod 26) = 1/26.$

(2) $p(y) = \sum_{k \in \mathbb{Z}_{26}} p(k)p(x = d_k(y)) = \sum_{k \in \mathbb{Z}_{26}} \frac{1}{26} p(x = y - k)$
$= \frac{1}{26} \sum_{k \in \mathbb{Z}_{26}} p(x = y - k) = \frac{1}{26} \sum_{x \in \mathbb{Z}_{26}} p(x) = \frac{1}{26}.$

Perfect Secrecy of a Cryptosystem in general

In the cryptosystem, a particular key K is used for only one encryption.

Theorem 3.4 (on Pages 68-69)

Suppose $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is a cryptosystem where $|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|$. Then the cryptosystem provides perfect secrecy if and only if

1. every key is used with equal probability $1/|\mathcal{K}|$,
2. and for every $x \in \mathcal{P}$ and $y \in \mathcal{C}$, there is a unique key K such that $e_K(x) = y$.

Proof of Theorem 3.4 (Omitted)

See Pages 68-69 in the textbook.

Perfect Secrecy of the One-time Pad

Cryptosystem 3.1 One-time Pad (OTP, 一次一密) (on Page 69)

Let $n \geq 1$. Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_2)^n$. For a key $K = (k_1, k_2, \ldots, k_n)$,

1. Encryption: $e_K(x_1, x_2, \ldots, x_n) = (x_1 \oplus k_1, x_2 \oplus k_2, \ldots, x_n \oplus k_n)$
2. Decryption: $d_K(y_1, y_2, \ldots, y_n) = (y_1 \oplus k_1, y_2 \oplus k_2, \ldots, y_n \oplus k_n)$

Perfect Secrecy of the One-time Pad

Cryptosystem 3.1 One-time Pad (OTP, 一次一密) (on Page 69)

Let $n \geq 1$. Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_2)^n$. For a key $K = (k_1, k_2, \ldots, k_n)$,

1. Encryption: $e_K(x_1, x_2, \ldots, x_n) = (x_1 \oplus k_1, x_2 \oplus k_2, \ldots, x_n \oplus k_n)$
2. Decryption: $d_K(y_1, y_2, \ldots, y_n) = (y_1 \oplus k_1, y_2 \oplus k_2, \ldots, y_n \oplus k_n)$

Characteristics of OTP

1. Each key is used for only one encryption.
2. a well-known realization of perfect secrecy.
3. Ease of encryption and decryption.
4. first described by Gilbert Vernam in 1917 for use in automatic encryption and decryption of telegraph messages.
5. an "unbreakable" cryptosystem for 30+ years with no proof; in 1949, Shannon gave a proof by "perfect secrecy".

Sufficient Conditions for Perfect Secrecy

Characteristics of OTP

6. The One-time Pad is vulnerable (易被攻击的) to a known-plaintext attack since the key K can be computed from $x \oplus y$.
7. the amount of key is as big as the amount of plaintext, which creates sever key communication and management problems.
8. has been employed in military and diplomatic contexts.

Sufficient Conditions for Perfect Secrecy

Characteristics of OTP

6. The One-time Pad is vulnerable (易被攻击的) to a known-plaintext attack since the key K can be computed from $x \oplus y$.
7. the amount of key is as big as the amount of plaintext, which creates sever key communication and management problems.
8. has been employed in military and diplomatic contexts.

Requirements of An Unbreakable System

1949, in "Communication Theory of Secrecy System", Shannon proved:

- any theoretically unbreakable (or unconditionally secure) system must have essentially the same following characteristics as the One-time Pad:
  1. the key must be chosen truly random;
  2. the key space must be as large as the plaintext set;
  3. the key must be never reused in whole or part;从未全部或部分重复使用;
  4. the key must be kept secret.

Outline

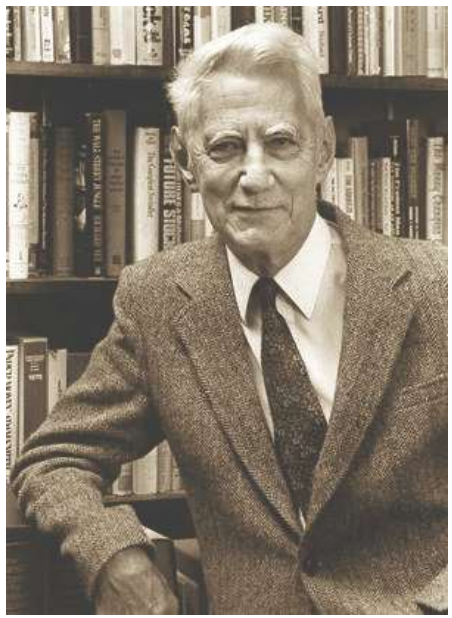
Claude E. Shannon (1916-2001)

the father of Information Theory (Information Age)

The landmark paper: 1948, "A Mathematical Theory of Communication"

the Godfather of Public-Key Cryptography (Art $\longrightarrow$ Science)

The famous work on Security:
1949, "Communication Theory of Secrecy System"



Claude E. Shannon (1916-2001)

mathematician, electrical engineer, and cryptographer

- 20 years old, two bachelor's degrees from University of Michigan (electrical engineering and mathematics)
- 22 years old, master's degree, MIT; 24 years old, Ph.D. degree, MIT
- His childhood hero was Thomas Edison (1847-1931) (a distant cousin).
- He ever worked for the cryptanalysis on German's rockets during the Word War II.
- He was a player and an inventor.
- He won many famous awards.

Claude E. Shannon (1916-2001)

mathematician, electrical engineer, and cryptographer

- 20 years old, two bachelor's degrees from University of Michigan (electrical engineering and mathematics)
- 22 years old, master's degree, MIT; 24 years old, Ph.D. degree, MIT
- His childhood hero was Thomas Edison (1847-1931) (a distant cousin).
- He ever worked for the cryptanalysis on German's rockets during the Word War II.
- He was a player and an inventor.
- He won many famous awards.
- Shannon-style research:
  1. he followed his nose
  2. engineering problem $\rightarrow$ elegant mathematical theory

Outline

Criteria of Security Evaluating

Three most useful security criteria

1. Computational Security: computational effort
2. Provable Security: by means of reduction
3. Unconditional Security: infinite computational resources
   Perfect Secrecy

Criteria of Security Evaluating

Three most useful security criteria

1. Computational Security: computational effort
2. Provable Security: by means of reduction
3. Unconditional Security: infinite computational resources
   Perfect Secrecy

The attack model should be specified at the same time.

- A cryptosystem is *** secure against *** attack. For example:
  the Shift/Substitution/Vigenere Cipher is not computationally secure against a ciphertext-only attack;
  the Hill Cipher is computationally secure against a ciphertext-only attack;
  the Hill Cipher is not computationally secure against a known-plaintext attack;
  the Shift/Substitution Cipher is unconditionally secure against a ciphertext -only attack, if a single element of plaintext is encrypted with a given key.

Outline

Why we introduce Entropy?

- Perfect Secrecy is studied for the special situation where a key is used for only one encryption.
- Q: What happens when more and more plaintexts are encrypted by the same key, and how likely a cryptanalyst is able to carry out a successful ciphertext-only attack, given sufficient time?
  A: The basic tool is the idea of entropy, given by Shannon in 1948.

Shannon's Entropy

Definition 3.4 (on Page 71)

Let X be a discrete RV on a finite set $\mathcal{X}$. Then, the *entropy* of X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2(p(x)). \tag{5.1}$$

Remarks: 1) $0 \cdot \log_2 0 = 0$; 2) the base of the logarithms can be changed.

Shannon's Entropy

Definition 3.4 (on Page 71)

Let X be a discrete RV on a finite set $\mathcal{X}$. Then, the *entropy* of X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2(p(x)). \geq 0 \tag{5.1}$$

Remarks: 1) $0 \cdot \log_2 0 = 0$; 2) the base of the logarithms can be changed.

Meaning of Entropy

- Entropy is a mathematical measure of information or uncertainty.

Shannon's Entropy

Definition 3.4 (on Page 71)

Let X be a discrete RV on a finite set $\mathcal{X}$. Then, the *entropy* of X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2(p(x)). \tag{5.1}$$

Remarks: 1) $0 \cdot \log_2 0 = 0$; 2) the base of the logarithms can be changed.

Meaning of Entropy

- Entropy is a mathematical measure of information or uncertainty.

An Example:

Let $X \sim p(x)$ with $p(x) = \{1/2, 1/4, 1/4\}$; What is $H(X)$?

Shannon's Entropy

Definition 3.4 (on Page 71)

Let X be a discrete RV on a finite set $\mathcal{X}$. Then, the *entropy* of X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2(p(x)). \tag{5.1}$$

Remarks: 1) $0 \cdot \log_2 0 = 0$; 2) the base of the logarithms can be changed.

Meaning of Entropy

- Entropy is a mathematical measure of information or uncertainty.

An Example:

Let $X \sim p(x)$ with $p(x) = \{1/2, 1/4, 1/4\}$; What is $H(X)$?
$H(X) = 1/2 \log_2 2 + 1/4 \log_2 4 + 1/4 \log_2 4 = 1.5$ bits.

Definitions 3.6: (on Page 74)

Let X and Y be two discrete RVs on two finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively.

1. The *conditional entropy* of X for any fixed $y \in \mathcal{Y}$ is
$$H(X|y) = -\sum_{x \in \mathcal{X}} p(x|y) \log_2(p(x|y)). \tag{5.2}$$
2. The *conditional entropy* of X given Y is
$$H(X|Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log_2(p(x|y)). \tag{5.3}$$
3. The *(joint) entropy* of (X,Y) is
$$H(X,Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log_2(p(x,y)). \tag{5.4}$$

Properties of Entropy

Theorems and Corollaries

1. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

p_i 代表概率质量函数中某个特定事件发生的概率。那么p_i可以是$P(X=i)$

2. $H(X) \leq \log_2 n$ with equality iff $p_i = 1/n$, where $|\mathcal{X}| = n$.

3. $H(X,Y) \leq H(X) + H(Y)$ with equality iff X and Y are independent.

4. $H(X|Y) \leq H(X)$ with equality iff X and Y are independent.

By the concavity of log function and Jensen's inequality

Properties of Entropy

Definition 3.5: Concavity $\bigcap$ (Omitted)

A real-valued function f is *concave* on an interval I if for all $x, y \in I$,

$$f(\frac{x+y}{2}) \geq \frac{f(x)+f(y)}{2}. \tag{5.5}$$

A real-valued function f is *strictly concave* on an interval I if for all $x \neq y \in I$,

$$f(\frac{x+y}{2}) > \frac{f(x)+f(y)}{2}. \tag{5.6}$$

Examples

$\log_2 x$, $-x^2$ are concave.

Properties of Entropy

Theorem 3.5: Jensen's Inequality (Omitted)

Suppose f is continuous and strictly concave on the interval I, and

$$\sum_{i=1}^{n} a_i = 1,$$

for $a_i > 0$, $1 \leq i \leq n$. Then,

$$\sum_i a_i f(x_i) \leq f(\sum_i a_i x_i).$$

Further, equality occurs if and only if $x_1 = x_2 = \ldots = x_n$.

by inductive method

Outline

Entropy in a Cryptosystem

Key Equivocation

Suppose that $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is a cryptosystem. Let $\mathbf{P}$, $\mathbf{C}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

$H(\mathbf{P})$, $H(\mathbf{C})$, $H(\mathbf{K})$, $H(\mathbf{P}|\mathbf{C})$, $H(\mathbf{C}|\mathbf{P})$, $H(\mathbf{K}|\mathbf{C})$, ...

$H(\mathbf{K}|\mathbf{C})$ is called the key equivocation (密钥含糊度).

$H(\mathbf{K}|\mathbf{C})$ measures the amount of uncertainty of the key remaining when the ciphertext is known.

Entropy in a Cryptosystem

Key Equivocation

Suppose that $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is a cryptosystem. Let $\mathbf{P}$, $\mathbf{C}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively.

$H(\mathbf{P})$, $H(\mathbf{C})$, $H(\mathbf{K})$, $H(\mathbf{P}|\mathbf{C})$, $H(\mathbf{C}|\mathbf{P})$, $H(\mathbf{K}|\mathbf{C})$, ...

$H(\mathbf{K}|\mathbf{C})$ is called the key equivocation (密钥含糊度).

$H(\mathbf{K}|\mathbf{C})$ measures the amount of uncertainty of the key remaining when the ciphertext is known.

Theorem 3.10 (on Page 75)

Suppose that $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is a cryptosystem. Then

$$H(\mathbf{K}|\mathbf{C}) = H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C}).$$

(See Page 75 for the proof.)

Spurious Keys and Unicity Distance (伪密钥和唯一解距离)

Assumptions

Suppose $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is the cryptosystem being used. Let $\mathbf{P}$, $\mathbf{C}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively

- a ciphertext-only attack: Oscar knows the a string of ciphertext $y_1 y_2 \cdots y_n$, which is the encryption of $x_1 x_2 \cdots x_n$ using one key K, i.e., $y_i = e_K(x_i)$.
- Oscar has infinite computational resources.
- Oscar knows that the plaintext is "natural" language, called L.
- The basic goal of the cryptanalyst is to determine the key.

Spurious Keys and Unicity Distance (伪密钥和唯一解距离)

Assumptions

Suppose $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ is the cryptosystem being used. Let $\mathbf{P}$, $\mathbf{C}$ and $\mathbf{K}$ be the RVs defined on $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{K}$, respectively

- a ciphertext-only attack: Oscar knows the a string of ciphertext $y_1 y_2 \cdots y_n$, which is the encryption of $x_1 x_2 \cdots x_n$ using one key K, i.e., $y_i = e_K(x_i)$.
- Oscar has infinite computational resources.
- Oscar knows that the plaintext is "natural" language, called L.
- The basic goal of the cryptanalyst is to determine the key.

Useful Definitions and Results

1. **Spurious keys:** the remaining possible but incorrect keys after some analysis done by Oscar.
   (Oscar knows ciphertexts "WNAJW" using a shift cipher. It has two "meaningful" plaintexts, 'river' and 'arena' with respective keys F and W. One key will be correct, the other will be spurious.)

Spurious Keys and Unicity Distance (伪密钥和唯一解距离)

Useful Definitions and Results

2. **The average number of the spurious keys:** $\bar{s}_n = \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| - 1$,
   where $K(\mathbf{y}) = \{K \in \mathcal{K} : \exists\, \mathbf{x} \in \mathcal{P}^n \text{ such that } p(\mathbf{x}) > 0 \text{ and } e_K(\mathbf{x}) = \mathbf{y}\}$.

Spurious Keys and Unicity Distance (伪密钥和唯一解距离)

Useful Definitions and Results

2. **The average number of the spurious keys:** $\bar{s}_n = \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| - 1$, where $K(\mathbf{y}) = \{K \in \mathcal{K} : \exists\, \mathbf{x} \in \mathcal{P}^n \text{ such that } p(\mathbf{x}) > 0 \text{ and } e_K(\mathbf{x}) = \mathbf{y}\}$.
3. Theorem 3.11 (on page 79) says $\bar{s}_n \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1$ ($\to 0$ as $n \to \infty$).

Spurious Keys and Unicity Distance (伪密钥和唯一解距离)

Useful Definitions and Results

2. **The average number of the spurious keys:** $\bar{s}_n = \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| - 1$, where $K(\mathbf{y}) = \{K \in \mathcal{K} : \exists\, \mathbf{x} \in \mathcal{P}^n \text{ such that } p(\mathbf{x}) > 0 \text{ and } e_K(\mathbf{x}) = \mathbf{y}\}$.
3. Theorem 3.11 (on page 79) says $\bar{s}_n \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1$ ($\to 0$ as $n \to \infty$).
4. **The entropy per letter** of a natural language L: $H_L = \lim_{n \to \infty} \frac{H(\mathbf{P}^n)}{n}$, where $\mathbf{P}^n$ is the RV that has as its probability distribution that of all n-grams of plaintext.
   The rudundancy of L: $R_L = 1 - \frac{H_L}{\log_2 |\mathcal{P}|}$.
   (Definition 3.7 on page 77)

Spurious Keys and Unicity Distance (伪密钥和唯一解距离)

Useful Definitions and Results

2. **The average number of the spurious keys:** $\bar{s}_n = \sum_{\mathbf{y} \in \mathcal{C}^n} p(\mathbf{y})|K(\mathbf{y})| - 1$, where $K(\mathbf{y}) = \{K \in \mathcal{K} : \exists\, \mathbf{x} \in \mathcal{P}^n \text{ such that } p(\mathbf{x}) > 0 \text{ and } e_K(\mathbf{x}) = \mathbf{y}\}$.
3. Theorem 3.11 (on page 79) says $\bar{s}_n \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1$ ($\to 0$ as $n \to \infty$).
4. **The entropy per letter** of a natural language L: $H_L = \lim_{n \to \infty} \frac{H(\mathbf{P}^n)}{n}$, where $\mathbf{P}^n$ is the RV that has as its probability distribution that of all n-grams of plaintext.
   The rudundancy of L: $R_L = 1 - \frac{H_L}{\log_2 |\mathcal{P}|}$.
   (Definition 3.7 on page 77)
5. **Unicity distance:** n_0 such that $\bar{s}_{n_0} = 0$; i.e., the average amount n_0 of ciphertexts $y_1, y_2, \cdots, y_{n_0}$ required for an opponent to be able to uniquely compute the key, given enough computing time. (Definition 3.8 on page 79)
   $n_0 \approx \frac{\log_2 |\mathcal{K}|}{R_L \log_2 |\mathcal{P}|}$ by Theorem 3.11.
   For the Substitution Cipher, if $R_L = 0.75$, then $n_0 \approx \frac{\log_2 |26!|}{R_L \log_2 |26|} \approx 25$.

Summary

Homework

Problem Set 2

Exercises 3.5, 3.8, 3.9(a), 3.15, 3.17.

Thanks for your attention!

Questions?

暨南大學
JINAN UNIVERSITY