ECEN 5022 Cryptography

Introduction to Information Theory

Peter Mathys

University of Colorado

Spring 2008

Entropy

- Definition: The **entropy** $H(X)$ of a discrete RV X with alphabet $\mathcal{A}$ is defined as

$$H(X) = -\sum_{x \in \mathcal{A}} p_X(x) \log p_X(x) .$$

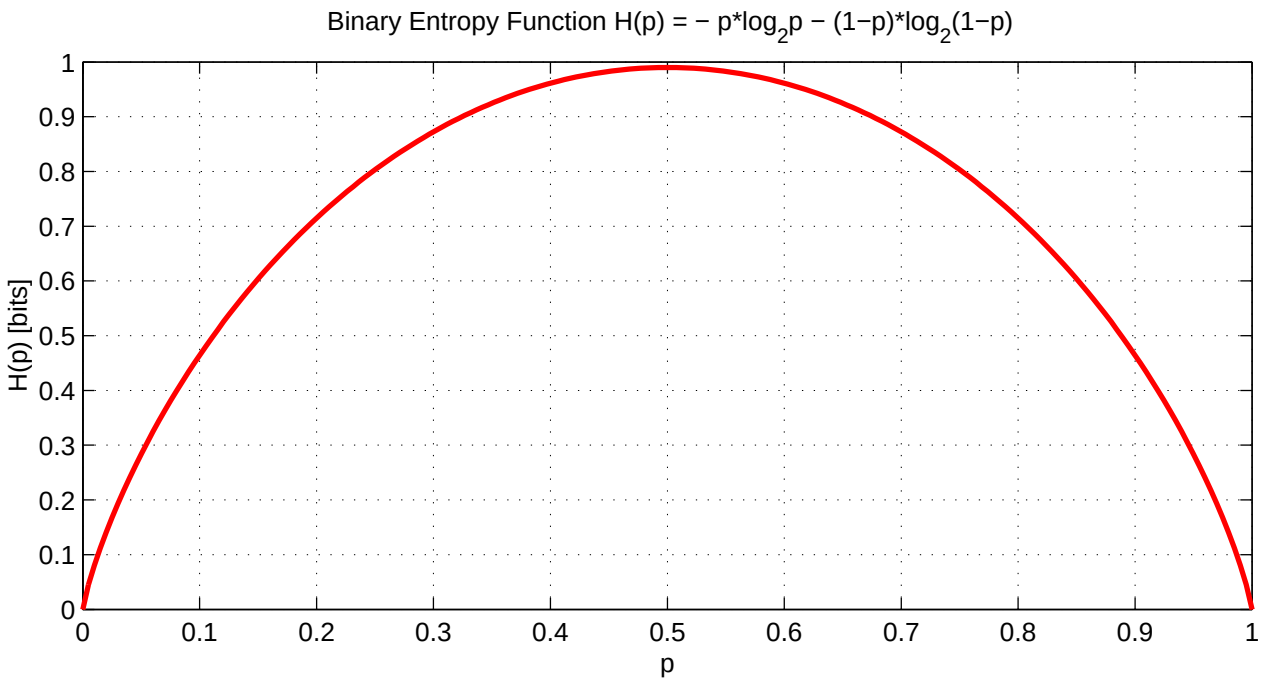
  $H(X)$ is bounded as $0 \leq H(X) \leq \log |\mathcal{A}|$. If $\log_2$ is used then the units of $H(X)$ are bits.

- Definition: The **joint entropy** $H(X, Y)$ of discrete RVs X and Y with alphabets $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$H(X, Y) = -\sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} p_{X,Y}(x, y) \log p_{X,Y}(x, y) .$$

  Bounds: $0 \leq H(X, Y) \leq H(X) + H(Y) \leq \log |\mathcal{A}| + \log |\mathcal{B}|$.

Binary Entropy Function



- Definition: The **binary** entropy function is defined as $H(p) = -p \log p - (1-p) \log(1-p)$. If $\log_2$ is used the result is in bits.

Conditional Entropy

- Definition: The **conditional entropy** $H(Y|X)$ of discrete RV Y given discrete RV X (with alphabets $\mathcal{B}$ and $\mathcal{A}$) is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{A}} p_X(x)\, H(Y|X=x) \\ &= -\sum_{x\in\mathcal{A}} p_X(x) \sum_{y\in\mathcal{B}} p_{Y|X}(y|x)\, \log p_{Y|X}(y|x) \\ &= -\sum_{x\in\mathcal{A}} \sum_{y\in\mathcal{B}} p_{X,Y}(x,y)\, \log p_{Y|X}(y|x) \end{aligned}$$

  Bounds: $0 \le H(Y|X) \le H(Y) \le \log|\mathcal{B}|$.
- $H(Y|X) = H(Y)$ (and $H(X|Y) = H(X)$) iff X and Y are statistically **independent**.
- **Chain rule:** $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$. This follows from Bayes rule.

Mutual Information

- Definition: The **mutual information** $I(X;Y)$ between two discrete RVs X and Y with alphabets $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathcal{A}}\sum_{y\in\mathcal{B}} p_{X,Y}(x,y)\,\log\frac{p_{X,Y}(x,y)}{p_X(x)\,p_Y(y)} \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

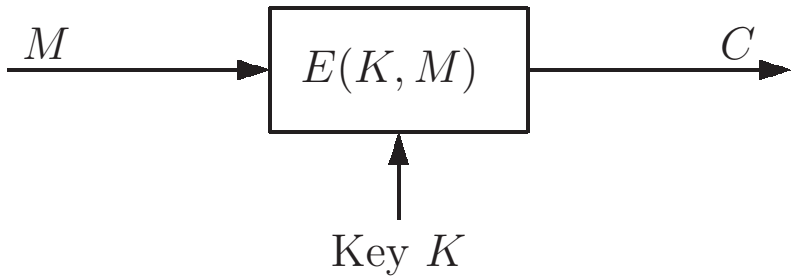
Bounds: $0 \le I(X;Y) \le \max\{\log|\mathcal{A}|, \log|\mathcal{B}|\}$.

Conditional Mutual Information

- Definition: The **conditional mutual information** $I(X;Y|Z)$ between two discrete RVs X and Y with alphabets $\mathcal{A}$ and $\mathcal{B}$, given discrete RV Z with alphabet $\mathcal{C}$ is defined as

$$\begin{aligned} I(X;Y|Z) &= \sum_{x\in\mathcal{A}}\sum_{y\in\mathcal{B}}\sum_{z\in\mathcal{C}} p_{X,Y,Z}(x,y,z)\,\log\frac{p_{X,Y|Z}(x,y|z)}{p_{X|Z}(x|z)\,p_{Y|Z}(y|z)} \\ &= H(X|Z) - H(X|YZ) \\ &= H(Y|Z) - H(Y|XZ) \\ &= H(X|Z) + H(Y|Z) - H(X,Y|Z) \end{aligned}$$

Perfect Cryptosystem



- A perfect cryptosystem should have the following properties

$$I(M; C) = 0\,, \qquad I(M; C|K) = H(M)\,.$$

- From the first statement $I(M; C) = H(M) - H(M|C) = 0$ and thus $H(M|C) = H(M)$, i.e., M, C must be statistically independent.
- From the second statement $I(M; C|K) = H(M|K) - H(M|CK) = H(M|K) = H(M)$ (2'nd equality follows since $H(M|CK) = 0$) and thus M and K must be statistically independent.

Key Equivocation

- Let the discrete RVs M, K, C denote the plaintext, key, ciphertext, respectively, of a cryptosystem.
- Definition. $H(K|C)$ is called the **key equivocation**.
- Theorem. If M, K are statistically independent then $H(K|C) = H(K) + H(M) - H(C)$.
- Proof: Express $H(M,K,C)$ in two ways as

$$H(M,K,C) = H(M,K) + \underbrace{H(C|M,K)}_{=0} = H(M,K)$$
$$= H(M) + \underbrace{H(K|M)}_{=H(K)} = H(M) + H(K)\,,$$

and as $H(M,K,C) = H(K,C) + \underbrace{H(M|K,C)}_{=0} = H(K,C)$.

$$\Longrightarrow \qquad H(K|C) = H(K,C) - H(C) = H(M,K,C) - H(C)$$
$$= H(M) + H(K) - H(C)\,.$$

Entropy of Language

- Plaintext RV: $M^n = (M_1, M_2, \ldots, M_n)$.
- Ciphertext RV: $C^n = (C_1, C_2, \ldots, C_n)$.
- Question: How much ciphertext is needed to dermine key uniquely?
- Definition: The rate or entropy of a language $\mathcal{L}$ is defined as

$$H_{\mathcal{L}} = \lim_{n \to \infty} \frac{H(M^n)}{n} .$$

  For English $1.0 \leq H_{\mathcal{L}} \leq 1.5$.
- Definition: The redundancy of a language $\mathcal{L}$ with alphabet $\mathcal{M}$ is defined as

$$R_{\mathcal{L}} = 1 - \frac{H_{\mathcal{L}}}{\log |\mathcal{M}|} .$$

  English has redundancy ≈ 0.75.

Spurious Keys

- Definition. The set of possible keys k for encryption function E_k given some ciphertext $\mathbf{c} \in \mathcal{C}^n$ of length n is

$$K(\mathbf{c}) = \left\{ k \in \mathcal{K} \,\middle|\, \exists\, \mathbf{m} \in \mathcal{M}^n \text{ with } Pr(\mathbf{m}) > 0 \text{ and } \mathbf{c} = E_k(\mathbf{m}) \right\}$$

  "Set of keys that yield plausible plaintexts given ciphertext $\mathbf{c}$."
- Definition. Only one member of $K(\mathbf{c})$ is correct. The rest are called **spurious keys** and denoted S_n.
- The expected number of spurious keys, averaged over all ciphertexts $\mathbf{c} \in \mathcal{C}^n$ of length n, is

$$E[S_n] = \sum_{\mathbf{c} \in \mathcal{C}^n} \left(|K(\mathbf{c})| - 1 \right) Pr(\mathbf{c}) \,.$$

Unicity Distance

- Theorem. If $|\mathcal{C}| = |\mathcal{M}|$ and keys are chosen equiprobably from the keyspace $\mathcal{K}$ then, given a ciphertext of sufficiently large length n, the expected number of spurious keys satisfies

$$E[S_n] \geq \frac{|\mathcal{K}|}{|\mathcal{M}|^{nR_{\mathcal{L}}}} - 1\,,$$

  where $|\mathcal{K}|$ is the size of the keyspace, $|\mathcal{M}|$ is the size of the plaintext alphabet, and $R_{\mathcal{L}}$ is the redundancy of the plaintext language.

- Setting $E[S_n] = 0$ yields $\quad nR_{\mathcal{L}} \log |\mathcal{M}| \geq \log |\mathcal{K}| \quad$ and thus

$$n \geq n_0 = \frac{\log |\mathcal{K}|}{R_{\mathcal{L}} \log |\mathcal{M}|}\,.$$

  The quantity n_0 is called the **unicity distance** of a cryptosystem.

Examples

- For all examples $|\mathcal{M}| = 26 \Longrightarrow \log_2 |\mathcal{M}| = 4.70$ and $R_{\mathcal{L}} = 0.75$.
- **Shift (Caesar) cipher**. $|\mathcal{K}| = 26 \Longrightarrow \log_2 |\mathcal{K}| = 4.70$

  $$\text{Unicity distance:} \qquad n_0 = \frac{4.70}{0.75 \times 4.70} = 1.333\,.$$

- **Simple substitution cipher**. $|\mathcal{K}| = 26! \Longrightarrow \log_2 |\mathcal{K}| = 88.38$

  $$\text{Unicity distance:} \qquad n_0 = \frac{88.38}{0.75 \times 4.70} = 25.07\,.$$

- **One-time pad**. $|\mathcal{K}| = 26^n \Longrightarrow \log_2 |\mathcal{K}| = 4.70\,n$

  $$\text{Unicity distance:} \qquad n_0 = \frac{4.70\,n}{0.75 \times 4.70} = 1.333\,n\,.$$

  The only solution to $n \geq n_0 = 1.333\,n$ is $n = 0$.