Pre-Quantum Information Theory

Goutam Paul

Cryptology and Security Research Unit,
Indian Statistical Institute, Kolkata

February 9, 2016

Lecture at
International School and Conference on Quantum Information,
Institute of Physics (IOP), Bhubaneswar (Feb 9-18, 2016).

Outline

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption

2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding

3. Quantum Information

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

Information and Probability

Information and Probability

For an event with probability p, let $I(p)$ be the information contained in it.

Information and Probability

For an event with probability p, let $I(p)$ be the information contained in it.

- $p \downarrow \Rightarrow I(p) \uparrow$ and $p \uparrow \Rightarrow I(p) \downarrow$

Information and Probability

For an event with probability p, let $I(p)$ be the information contained in it.

- $p \downarrow \Rightarrow I(p) \uparrow$ and $p \uparrow \Rightarrow I(p) \downarrow$
- For two independent events with probabilities p_1 and p_2, $I(p_1 p_2) \propto I(p_1) + I(p_2)$.

Information and Probability

For an event with probability p, let $I(p)$ be the information contained in it.

- $p \downarrow \Rightarrow I(p) \uparrow$ and $p \uparrow \Rightarrow I(p) \downarrow$
- For two independent events with probabilities p_1 and p_2, $I(p_1 p_2) \propto I(p_1) + I(p_2)$.

Thus, a natural definition is

$$I(p) \triangleq \log\left(\frac{1}{p}\right) = -\log p.$$

Relation to Uncertainty / Surprise / Knowledge Gain

Amount of information contained in an event

Relation to Uncertainty / Surprise / Knowledge Gain

Amount of information contained in an event

= Amount of uncertainty *before the event happens*

Relation to Uncertainty / Surprise / Knowledge Gain

Amount of information contained in an event

= Amount of uncertainty *before the event happens*

= Amount of surprise *when the event happens*

Relation to Uncertainty / Surprise / Knowledge Gain

Amount of information contained in an event

= Amount of uncertainty *before the event happens*

= Amount of surprise *when the event happens*

= Amount of knowledge gain *after the event happens*

Average Information

Let X denote a random variable taking values from a discrete set (may denote a set of events or a source of symbols) with probabilities $p(x) = Prob(X = x)$.

Average Information

Let X denote a random variable taking values from a discrete set (may denote a set of events or a source of symbols) with probabilities $p(x) = Prob(X = x)$.

Average information in X (or of the corresponding set / source)

$$H(X) \triangleq E[I(p(X))]$$

Average Information

Let X denote a random variable taking values from a discrete set (may denote a set of events or a source of symbols) with probabilities $p(x) = Prob(X = x)$.

Average information in X (or of the corresponding set / source)

$$\begin{aligned} H(X) &\triangleq E[I(p(X))] \\ &= E[-\log p(X)] \end{aligned}$$

Average Information

Let X denote a random variable taking values from a discrete set (may denote a set of events or a source of symbols) with probabilities $p(x) = Prob(X = x)$.

Average information in X (or of the corresponding set / source)

$$
\begin{aligned}
H(X) &\triangleq E[I(p(X))] \\
&= E[-\log p(X)] \\
&= -\sum_{x \in X} p(x) \log p(x)
\end{aligned}
$$

Average Information

Let X denote a random variable taking values from a discrete set (may denote a set of events or a source of symbols) with probabilities $p(x) = Prob(X = x)$.

Average information in X (or of the corresponding set / source)

$$\begin{aligned} H(X) &\triangleq E[I(p(X))] \\ &= E[-\log p(X)] \\ &= -\sum_{x \in X} p(x) \log p(x) \end{aligned}$$

This is called the entropy of the variable X (or of the set / source).

Joint and Conditional Entropy

Joint and Conditional Entropy

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

Joint and Conditional Entropy

$$H(X,Y) \triangleq -\sum_x \sum_y p(x,y) \log p(x,y).$$

$$H(Y \mid X) \triangleq \sum_x p(x) H(Y \mid X = x)$$

Joint and Conditional Entropy

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

$$\begin{aligned} H(Y \mid X) &\triangleq \sum_{x} p(x) H(Y \mid X = x) \\ &= \sum_{x} p(x)\left(-\sum_{y} p(y|x)\log p(y|x)\right) \end{aligned}$$

Joint and Conditional Entropy

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

$$\begin{aligned} H(Y \mid X) &\triangleq \sum_{x} p(x) H(Y \mid X = x) \\ &= \sum_{x} p(x) \left(-\sum_{y} p(y|x) \log p(y|x) \right) \\ &= -\sum_{x}\sum_{y} p(x, y) \log p(y|x) \end{aligned}$$

Important Results Related to Entropy

Important Results Related to Entropy

Chain Rule: $H(X, Y) = H(X) + H(Y|X)$

Important Results Related to Entropy

Chain Rule: $H(X, Y) = H(X) + H(Y|X)$

$$H(X, Y) \leq H(X) + H(Y)$$

Important Results Related to Entropy

Chain Rule: $H(X, Y) = H(X) + H(Y|X)$

$$H(X, Y) \leq H(X) + H(Y)$$

$$H(Y \mid X) \leq H(Y)$$

Mutual Information

$$I(X;Y) \triangleq \sum_{x}\sum_{y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

Mutual Information

$$
\begin{aligned}
I(X;Y) &\triangleq \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= H(X) - H(X|Y)
\end{aligned}
$$

Mutual Information

$$\begin{aligned} I(X;Y) &\triangleq \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \end{aligned}$$

STATISTICAL INDIAN INSTITUTE
UNITY IN DIVERSITY

Mutual Information

$$
\begin{aligned}
I(X;Y) &\triangleq \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

Roadmap

Information and Codeword Length

Kraft Inequality: The necessary and sufficient conditions for the existence of an instantaneous code over an r-ary alphabet with codeword lengths $\ell_1, \ell_2, \ldots, \ell_n$ satisfy

$$\sum_{i=1}^{n} r^{-\ell_i} \leq 1.$$

Information and Codeword Length

Kraft Inequality: The necessary and sufficient conditions for the existence of an instantaneous code over an r-ary alphabet with codeword lengths $\ell_1, \ell_2, \ldots, \ell_n$ satisfy

$$\sum_{i=1}^{n} r^{-\ell_i} \leq 1.$$

An Engineering Optimization:

Minimize $L = \sum_{i=1}^{n} p_i \ell_i$ s.t. $\sum_{i=1}^{n} r^{-\ell_i} \leq 1$ gives $\ell_i^* = -\log_r p_i$

and $L^* = \sum_{i=1}^{n} p_i \ell_i^* = H(X)$.

Entropy and Data Compression

For integer choice of codeword lengths,

$$H(X) \leq L^* < H(X) + 1.$$

Entropy and Data Compression

For integer choice of codeword lengths,

$$H(X) \leq L^* < H(X) + 1.$$

For supersymbols with n-symbols at a time,

$$H(X) \leq L_n^* < H(X) + \frac{1}{n}$$

and $L_n^* = H(X)$ is achievable for stationary distribution.

Entropy and Data Compression

For integer choice of codeword lengths,

$$H(X) \leq L^* < H(X) + 1.$$

For supersymbols with n-symbols at a time,

$$H(X) \leq L_n^* < H(X) + \frac{1}{n}$$

and $L_n^* = H(X)$ is achievable for stationary distribution.
This is Shannon's Source/Noiseless Coding Theorem.

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

Entropy as a Measure of Randomness

Entropy as a Measure of Randomness

Suppose $p_i \geq 0$, for $1 \leq i \leq n$.

Entropy as a Measure of Randomness

Suppose $p_i \geq 0$, for $1 \leq i \leq n$.

Maximize $\left(-\sum_i p_i log p_i\right)$ s.t. $\sum_i p_i = 1$ gives

Entropy as a Measure of Randomness

Suppose $p_i \geq 0$, for $1 \leq i \leq n$.

Maximize $\left(-\sum_i p_i log p_i\right)$ s.t. $\sum_i p_i = 1$ gives

$$p_1 = p_2 = \cdots = p_n.$$

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

Encryption increases Entropy

Encryption increases Entropy

The goal of encryption is to make the transmitted message look random.

Encryption increases Entropy

The goal of encryption is to make the transmitted message look random.

Typically, $H(C) > H(P)$.

Encryption increases Entropy

The goal of encryption is to make the transmitted message look random.

Typically, $H(C) > H(P)$.

But, $H(P \mid C)$ may be $< H(P)$

Example: Plaintext Entropy

Given

Example: Plaintext Entropy

Given

Three possible plaintexts: a, b, c,
with probabilities 0.5, 0.3, 0.2.

Example: Plaintext Entropy

Given

Three possible plaintexts: a, b, c,
with probabilities 0.5, 0.3, 0.2.

Three possible ciphertexts: U, V, W.

Example: Plaintext Entropy

Given

Three possible plaintexts: a, b, c,
with probabilities 0.5, 0.3, 0.2.

Three possible ciphertexts: U, V, W.

Two possible keys: k_1, k_2, equally likely.

Example: Plaintext Entropy

Given

Three possible plaintexts: a, b, c,
with probabilities 0.5, 0.3, 0.2.

Three possible ciphertexts: U, V, W.

Two possible keys: k_1, k_2, equally likely.

Encryption under k_1: U, V, W.
Encryption under k_2: U, W, V.

Example: Plaintext Entropy (... contd)

One can calculate

Example: Plaintext Entropy (... contd)

One can calculate

$$p(U) = 0.5, p(V) = p(W) = 0.25.$$

Example: Plaintext Entropy (... contd)

One can calculate

$$p(U) = 0.5, p(V) = p(W) = 0.25.$$

$$p(a \mid V) = 0$$

$$p(b \mid V) = 0.6$$

$$p(c \mid V) = 0.4$$

Example: Plaintext Entropy (... contd)

One can calculate

$$p(U) = 0.5, p(V) = p(W) = 0.25.$$

$$p(a \mid V) = 0$$

$$p(b \mid V) = 0.6$$

$$p(c \mid V) = 0.4$$

Similarly, one can calculate probabilities of a, b, c given W.

Example: Plaintext Entropy (... contd)

Thus,

Example: Plaintext Entropy (... contd)

Thus,

$$\begin{aligned} H(P) &= -(0.5\log_2(0.5) + 0.3\log_2(0.3) + 0.2\log_2(0.2)) \\ &= 1.485 \end{aligned}$$

Example: Plaintext Entropy (... contd)

Thus,

$$\begin{aligned} H(P) &= -(0.5\log_2(0.5) + 0.3\log_2(0.3) + 0.2\log_2(0.2)) \\ &= 1.485 \end{aligned}$$

$$\begin{aligned} H(P \mid C) &= -\sum_{x\in\{U,V,W\}} \sum_{y\in\{a,b,c\}} p(x)p(y|x)\log_2 p(y|x) \\ &= 0.485 \end{aligned}$$

Perfect Secrecy

Perfect Secrecy

Information Theoretic Security:

Perfect Secrecy

Information Theoretic Security:

$$H(P \mid C) = H(P)$$

Perfect Secrecy

Information Theoretic Security:

$$H(P \mid C) = H(P)$$

Or, equivalently,

$$Prob(P \mid C) = Prob(P).$$

Perfect Secrecy

Information Theoretic Security:

$$H(P \mid C) = H(P)$$

Or, equivalently,

$$Prob(P \mid C) = Prob(P).$$

A necessary condition for this is

$$H(K) \geq H(P).$$

Roadmap

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

Discrete Channel

- Input alphabet X.
- Output alphabet Y.
- Probability Transition Matrix $p(y|x)$.

$$\text{Informational Channel Capacity } C = \max_{p(x)} I(X;Y).$$

Roadmap

An (M, n) Code

An (M, n) Code

- An index set $\{1, 2, \ldots, M\}$.

An (M, n) Code

- An index set $\{1, 2, \ldots, M\}$.
- An encoding function $C : \{1, 2, \ldots, M\} \rightarrow X^n$.

An (M, n) Code

- An index set $\{1, 2, \ldots, M\}$.
- An encoding function $C : \{1, 2, \ldots, M\} \rightarrow X^n$.
- A decoding function $D : Y^n \rightarrow \{1, 2, \ldots, M\}$.

Error probability

Error probability

- Conditional error probability given index i was sent:
$\epsilon_i = \Pr(D(Y^n) \neq i | X^n = C(i)) = \sum_{D(y^n)\neq i} p(y^n | c(i))$.

Error probability

- Conditional error probability given index i was sent:
  $\epsilon_i = \Pr(D(Y^n) \neq i | X^n = C(i)) = \sum_{D(y^n) \neq i} p(y^n | c(i))$.
- Maximum error probability $\epsilon_{max} = \max_{i \in \{1,2,\ldots,M\}} \epsilon_i$.

Error probability

- Conditional error probability given index i was sent:
  $\epsilon_i = \Pr(D(Y^n) \neq i | X^n = C(i)) = \sum_{D(y^n) \neq i} p(y^n | c(i))$.
- Maximum error probability $\epsilon_{max} = \max_{i \in \{1,2,\ldots,M\}} \epsilon_i$.
- Average error probability $\epsilon_{avg} = \frac{1}{M} \sum_{i=1}^{M} \epsilon_i$.

Rate

- $R = \frac{\log_2 M}{n}$ bits per transmission.

Rate

- $R = \frac{\log_2 M}{n}$ bits per transmission.
- A rate R is said to be achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes such that $\epsilon_{max} \to 0$ as $n \to \infty$.

Rate

- $R = \frac{\log_2 M}{n}$ bits per transmission.
- A rate R is said to be achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes such that $\epsilon_{max} \to 0$ as $n \to \infty$.
- Operational channel capacity is the supremum of all achievable rates.

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

Shannon's Noisy Channel Coding Theorem

Shannon's Noisy Channel Coding Theorem

- All rates below capacity are achievable.

Shannon's Noisy Channel Coding Theorem

- All rates below capacity are achievable.
- $\forall R < C$, $\exists$ a sequence of codes such that $\epsilon_{max} \to 0$ as $n \to \infty$.

Shannon's Noisy Channel Coding Theorem

- All rates below capacity are achievable.
- $\forall R < C$, $\exists$ a sequence of codes such that $\epsilon_{max} \to 0$ as $n \to \infty$.
- Informational capacity = operational capacity.

Band Limited Gaussian Channel

$$C = W \log \left(1 + \frac{P}{N_0 W} \right)$$

bits per second, where $\frac{N_0}{2}$ watts/Hz is the noise spectral density and P is the signal power.

Roadmap

1. Measures of Information
   - Uncertainty
   - Compressibility
   - Randomness
   - Encryption
2. Measures of Information Flow
   - Channel Capacity
   - Code
   - Noisy Coding
3. Quantum Information

From Pre-Quantum to Quantum

From Pre-Quantum to Quantum

- von Neumann entropy.

From Pre-Quantum to Quantum

- von Neumann entropy.
- Schumacher's quantum noiseless coding theorem.

From Pre-Quantum to Quantum

- von Neumann entropy.
- Schumacher's quantum noiseless coding theorem.
- Holevo bound: upper bound of accessible information.

From Pre-Quantum to Quantum

- von Neumann entropy.
- Schumacher's quantum noiseless coding theorem.
- Holevo bound: upper bound of accessible information.
- Classical capacity and quantum capacity of quantum channels.

THANK YOU

Questions / Comments ?

Homepage: http://www.goutampaul.com
Email: goutam.k.paul@gmail.com