The role of entropy in classical and quantum communications

George Androulakis

University of South Carolina

Quantum Probability 40

Thursday August 15, 2019

Outline

1. Definition of Shannon entropy
2. Set-up of communication scheme
3. Shannon's noiseless coding Theorem
4. Shannon's noisy channel coding Theorem
5. Schumacher's coding Theorem
6. Give an upper bound for the accessible classical information of a q-q channel
7. The Holevo bound

Shannon entropy

Figure 1: By cartoonist Mark Heath

Definition of Shannon entropy

Entropy = measure of uncertainty (i.e. lack of information), measure of our surprise when an event happens.

First attempt to measure our surprise when an event happens: $\frac{1}{p}$.

Additivity of surprise for independent events.

Second attempt to measure our surprise when an event happens: $\log_2 \frac{1}{p}$

Average the surprises.

Shannon Entropy If P is a prob. distr. then $H(P) = -\sum_i p_i \log_2 p_i$. If X is a r.v. then $H(X) = \langle -\log_2 X \rangle$, (Claude Shannon, "A Mathematical Theory of Communication", 1948).

Set-up of communication scheme

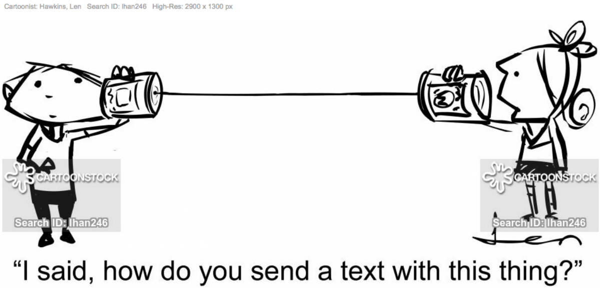


Figure 2: By cartoonist Len Hawkins

General communication scheme

$$\text{Alice or information source} \quad \mathcal{S}(\mathfrak{A}) \stackrel{\mathcal{C}}{\to} \mathcal{S}(\mathcal{A}) \stackrel{\Phi}{\to} \mathcal{S}(\mathcal{B}) \stackrel{\mathcal{D}}{\to} \mathcal{S}(\mathfrak{B}) \quad \text{Bob.}$$

where

$$\mathfrak{A}, \mathcal{A}, \mathcal{B}, \mathfrak{B} \text{ are von Neumann algebras,}$$

$$\mathcal{S}(\cdot) \text{ denotes the set of states (positive unital functionals) on the algebra,}$$

$$\mathcal{C} = \text{ coding}, \quad \mathcal{D} = \text{ decoding}, \quad \Phi = \text{ channel.}$$

Usually,

$$\mathfrak{A} = \mathfrak{B} = C(\{1, \ldots, M\}) \text{ or } C(\{0,1\}^m) : \quad \text{classical communication,}$$

$$\mathfrak{A} = \mathfrak{B} = \mathcal{B}(H) \text{ or } \mathcal{B}((\mathbb{C}^2)^{\otimes m}) : \quad \text{quantum communication.}$$

$0, 1, |0\rangle, |1\rangle =$ code characters. Sequences of code characters are called codewords.

(General) states and channels

Classical state spaces: $S=$ a finite set of symbols.
$\mathcal{A}=$ the Abelian algebra $C(S)$ of all functions on S.
$\mathcal{S}(\mathcal{A})=$ all states (positive unital functionals) (i.e. prob. distr.) on S.

Quantum state spaces: $\mathcal{A}=$ the non-Abelian algebra $\mathcal{B}(\mathcal{H})$.
$\mathcal{S}(\mathcal{A})=$ all states on $\mathcal{A}$.

A **(noisy) channel** is an affine map $\Phi : \mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathcal{B})$ whose linear extension, still denoted by Φ, has completely positive adjoint $\Phi^\dagger$.

c-c (noisy) channels: $\Phi = (p(y|x))_{x\in S, y\in S'}$ a column stochastic matrix.

q-q (noisy) channels: $\Phi(\cdot) = \sum_j V_j \cdot V_j^*$ with $\sum_j V_j^* V_j = \mathbf{1}$.

q-q reversible channels: $\Phi(\cdot) = U \cdot U^*$, U is a unitary.

c-q (noisy) channels: $\Phi(P) = \sum_{s\in S} p_s \rho_s$

q-c (noisy) channels: $\Phi(\rho) = (\mathrm{tr}\,(\rho M_s))_{s\in S}$ where $M_s \geq 0$ for all $s \in S$ with $\sum_{s\in S} M_s = \mathbf{1}$ (POVM).

Codes

An example of a code $\mathcal{C} : S = \{a, b, \ldots, z\} \to \{0,1\}^+ := \cup_{n=0}^{\infty}\{0,1\}^n$,

$$\mathrm{a} \mapsto 0, \mathrm{b} \mapsto 1, \mathrm{c} \mapsto 00, \mathrm{d} \mapsto 01, \mathrm{e} \mapsto 10, \mathrm{f} \mapsto 11, \ldots$$

Definition

$\mathcal{C}$ is called **uniquely decodable** *if every finite sequence of code characters corresponds to at most one message.*
$\mathcal{C}$ is called **instantaneous** *if no codeword is a prefix of another codeword.*

Instantaneous codes $\subsetneq$ Uniquely decodable codes.
An example of an instantaneous code:

$$\mathrm{a} \mapsto 1, \mathrm{b} \mapsto 01, \mathrm{c} \mapsto 001, \mathrm{d} \mapsto 0001, \mathrm{e} \mapsto 00001, \ldots$$

An example of a uniquely decodable but not instantaneous code:

$$\mathrm{a} \mapsto 1, \mathrm{b} \mapsto 10, \mathrm{c} \mapsto 100, \mathrm{d} \mapsto 1000, \mathrm{e} \mapsto 10000, \ldots$$

Statistical ensembles

$S = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \mathrm{z}, \mathrm{space}\}$ An element of $\mathcal{S}(C(S))$ (i.e. a **statistical ensemble**): $(.0651, .0124, \ldots, .0007, .1918)^T \equiv$
((a, .0651), (b, .0124,), ... , (z, .0007), (.1918, ␣)) (freq. of English letters and space).
(obtained from http://www.data-compression.com/english.html)

Definition

Average codeword length of a code $\mathcal{C} = \sum_{i \in S} p_i \, length(\mathcal{C}(i))$.

For example, for the above statistical ensemble and for the instantaneous code

a $\mapsto$ 1 , b $\mapsto$ 01 , c $\mapsto$ 001 , d $\mapsto$ 0001 , ... ,

the average codeword length is equal to

$1 \times .0651 + 2 \times .0124 + \cdots + 26 \times .0007 + 27 \times .1918$

Loseless and asymptotically losseless classical data compression

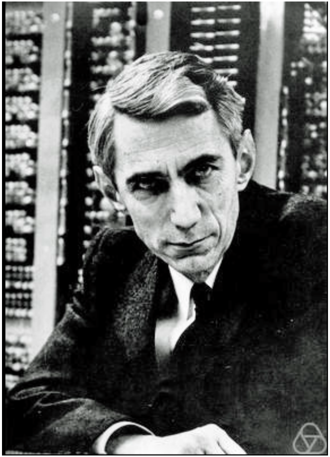


Figure 3: Claude Elwood Shannon (1916 – 2001)

Classical compression rates

Question

Assume that an information source emits symbols from a set S in an i.i.d. way according to the statistical ensemble $\mathcal{E} = (p_s)_{s \in S}$. What is the minimum numbers of bits per symbol needed for an asymptotically losseless data compression?

Reformulation: What is the infimum of positive numbers R such that for every $\delta > 0$ and for all n large enough there exist (**typical**) sets $T_{R,\delta,n} \subseteq S^n$ with $\frac{\log_2 |T_{R,\delta,n}|}{\log_2 |S|^n} < R$, and there exist coding and decoding maps $\mathcal{C}_n$ and $\mathcal{D}_n$ respectively as in the diagram

$$T_{R,\delta,n} \subseteq S^n \xrightarrow{\mathcal{C}_n} S^n \xrightarrow{\mathcal{D}_n} S^n$$

such that $\mathcal{D}_n \circ \mathcal{C}_n(t) = t \quad$ for all $t \in T_{R,\delta,n} \quad (\Leftrightarrow \mathcal{C}_n \text{ is } 1-1)$ and

$$\text{Prob}\,(T_{R,\delta,n}) = \sum_{s_1 s_2 \cdots s_n \in T_{R,\delta,n}} p_{s_1} p_{s_2} \cdots p_{s_n} > 1 - \delta.$$

Shannon's noiseless coding Theorem (asymptotically loseless data compression)

Theorem (Shannon's noiseless coding Theorem (asymptotically loseless data compression))

Assume that an information source emits symbols from a set S *according to the statistical ensemble* $\mathcal{E} = (p_s)_{s \in S}$ *and let* X *be the r.v. with values in* S *and p.m.f. equal to* $\mathcal{E}$*. Then for every* $R > H(X)$ *and for every* $\delta > 0$ *there exist sets* $T_{R,\delta,n} \subseteq S^n$ *for all* n *large enough, with* $\frac{\log_2 |T_{R,\delta,n}|}{\log_2 |S|^n} < R$ *such that* $Prob(T_{R,\delta,n}) \geq 1 - \delta$*.*
Moreover, for every $R < H(X)$ *and for every sequence of sets* $T_n \subseteq S^n$ *with* $\frac{\log_2 |T_n|}{\log_2 |S|^n} < R$ *we have that* $Prob(T_n) \to 0$*.*

$$T_{R,\delta,n} := \left\{ s_1 s_2 \cdots s_n \in S^n : \frac{1}{2^{nR}} < p_{s_1} p_{s_2} \cdots p_{s_n} < \frac{1}{2^{n(H(X)-\delta)}} \right\}.$$

Shannon's noiseless coding Thm, (loseless data compression)

Question

What is the minimum average codeword length among all uniquelly decodable codes

$$\{1,\ldots,M\} \overset{\mathcal{C}}{\to} \{0,1\}^+ := \cup_{n=0}^{\infty}\{0,1\}^n.$$

Theorem (Shannon's noiseless coding Thm, (loseless data compression))

Let $\mathcal{C} : \{1,\ldots,M\} \to \{0,1\}^+$ *be a uniquely decodable code. Assume that the symbols* $1,\ldots,M$ *are produced by i.i.d. copies of a r.v.* $X \sim (p_k)_{k=1}^M$ *and assume that the length* $(\mathcal{C}(k)) = n_k$ *for all* $1 \leq k \leq M$*. Then*
Average codeword length $= \sum_{k=1}^M p_k n_k \geq H(X)$.
Moreover, equality holds if and only if $p_k = \frac{1}{2^{n_k}}$ *for* $k = 1,\ldots,M$.

Optimal codes

An example:

X	Probabilities
x_1	$1/2$
x_2	$1/4$
x_3	$1/8$
x_4	$1/8$

Then

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - 2\times\frac{1}{8}\log_2\frac{1}{8} = \frac{7}{4}.$$

Consider the code $\mathcal{C}$ with $\mathcal{C}(x_1) = 0$, $\mathcal{C}(x_2) = 10$, $\mathcal{C}(x_3) = 110$ and $\mathcal{C}(x_4) = 111$. Then

$$\text{Average codeword length } = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = \frac{7}{4}.$$

In general such optimal code may not exist, (if $-\log_2 p_k \notin \mathbb{N}$ for some k), but always there exist a code (e.g. Huffman's or Shannon-Fano's code) s.t.

$$H(X) \leq \text{ Average codeword length } \leq H(X) + 1.$$

The classical capacity of a c-c channel

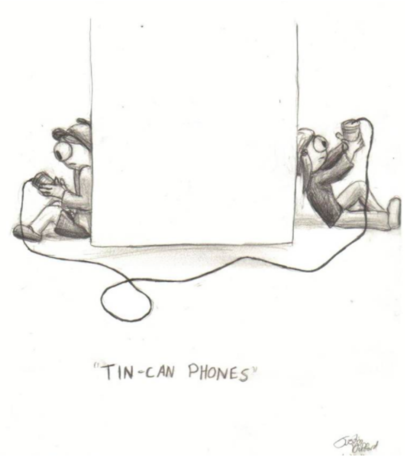


Figure 4: By cartoonist Justin Dufford.

The capacity of a noisy channel

Definition

Let $\Phi : \mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathcal{B})$ be a noisy channel. We define that the **classical capacity** $C_c(\Phi)$ *(resp.* **quantum capacity** $C_q(\Phi)$*) of Φ to be equal to the maximum asymptotic rate of (resp. cu)bits per repetition at which reliable classical (resp. quantum) communication is possible, i.e. the supremum of positive numbers R such that there exist two sequences $(n_k)_k$ and $(m_k)_k$ which tend to infinity such that $\frac{m_k}{n_k} \geq R$, and there exist sequences $(\mathcal{C}_{m_k,n_k})_k$ and $(\mathcal{D}_{n_k,m_k})_k$) of coding and decoding functions in the communication scheme*

$$\mathcal{S}(\mathcal{A}_{m_k}) \stackrel{\mathcal{C}_{m_k,n_k}}{\longrightarrow} \mathcal{S}(\mathcal{A}^{\otimes n_k}) \stackrel{\Phi^{\otimes n_k}}{\longrightarrow} \mathcal{S}(\mathcal{B}^{\otimes n_k}) \stackrel{\mathcal{D}_{n_k,m_k}}{\longrightarrow} \mathcal{S}(\mathcal{A}_{m_k})$$

such that $\|\mathbf{1}_{\mathcal{S}(\mathcal{A}_{m_k})} - \mathcal{D}_{n_k,m_k} \circ \Phi^{\otimes n_k} \circ \mathcal{C}_{m_k,n_k}\| \to 0$ as $k \to \infty$, where $\mathcal{A}_m := C(\{0,1\}^m)$, (resp. $\mathcal{A}_m := \mathcal{B}((\mathbb{C}^2)^{\otimes m})$).

Shannon's mutual information

Let X, Y be r.v. such that X takes the values $x_1, x_2, \ldots$ and Y takes the values $y_1, y_2, \ldots$.

Definition

The **Shannon's conditional entropy** $H(X|Y = y_j) = -\sum_i p(x_i|y_j) \log_2 p(x_i|y_j)$ *quantifies the uncertainty about the r.v.* X *conditionally that the r.v.* Y *takes the value* y_j.

Definition

The **Shannon's conditional entropy** $H(X|Y) = \sum_j p(y_j) H(X|Y = y_j)$ *the uncertainty about the r.v.* X *conditionally that the r.v.* Y *is known.*

Definition

The **Shannon's mutual information about** X **conveyed by** Y *is defined by* $I(X : Y) = H(X) - H(X|Y)$.

Easy Facts: $H(X|Y) = H(X, Y) - H(Y)$ and $I(X : Y) = I(Y : X)$.

Accessible information of a noisy c-c channel

Consider a c-c noisy channel Φ and ignore the coding and decoding schemes:

$$\mathcal{S}(\mathcal{A}) \stackrel{\Phi}{\to} \mathcal{S}(\mathcal{B}).$$

Let $\mathcal{E} = (x_i, p_i)_i$ be the statistical ensemble describing the input of the noisy channel Φ, i.e. the p.m.f. of the r.v. X. Let X be the r.v. that produces the inputs $x_1, x_2, \dots$ of the channel and Y be the r.v. that describes the outputs $y_1, y_2, \dots$ of the channel.

$\Phi = (p(y_j|x_i))_{i,j}$ (column stochastic matrix) where $p(y_j|x_i)$ is the probability that the output of the channel is y_j when the input of the channel is x_i then $\mathrm{Prob}\,(Y = y_j) = \sum_i \mathrm{Prob}\,(X = x_i) p(y_j|x_i)$.

Definition

The accessible information of the c-c channel Φ *is defined by* $Acc(\Phi) := \sup_{\mathcal{E}} I(X : Y)$, *(the maximum information about the input of the channel conveyed by the output of the channel).*

Shannon's noisy channel coding Theorem

Definition

A c-c channel $\Phi : \mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathcal{B})$ *is called* **memoryless** *if for every fixed* $n \in \mathbb{N}$, *if we apply the channel* $\Phi^{\otimes n}$ *repeatedly, the value of* $\Phi^{\otimes n}$ *at the* $(k+1)th$ *application only depends on the value of* $\Phi^{\otimes n}$ *on the* kth *application and not on the previous inputs/outputs.*

Theorem (Shannon's noisy channel Theorem)

Let $\Phi : \mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathcal{B})$ *be a c-c memoryless channel with* $C(\Phi) > 0$. *Then* $C_c(\Phi) = Acc(\Phi)$.

Quantum data compression

von Neumann entropy

Definition

If ρ is a density operators (states) on a Hilbert space $\mathcal{H}$, then the **von Neumann entropy** *$H(\rho)$ is defined as follows: $H(\rho) = -tr(\rho \log_2 \rho)$.*

Theorem

- *Unitary invar.:* $H(\rho) = H(U\rho U^*)$ $(\Rightarrow H(\rho) = H($ *eigenvalues of* $\rho))$.
- *Positivity:* $0 \leq H(\rho)(\leq dim(\mathcal{H}))$.
- *Concavity:* $H(\sum_k p_k \rho_k) \geq \sum_k p_k H(\rho_k)$ *for any prob. distr.* $(p_k)_k$ *and sequence of states* $(\rho_k)_k \subseteq \mathcal{S}(\mathcal{B}(\mathcal{H}))$.
- *Additivity: If* $\rho_i \in \mathcal{S}(\mathcal{H})_i)$ *then* $H(\rho_1 \otimes \rho_2) = H(\rho_1) + H(\rho_2)$.
- *Subadditivity: If* $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ *then* $H(\rho) \leq H(tr_{\mathcal{H}_1}(\rho)) + H(tr_{\mathcal{H}_2}(\rho))$.
- *Lower semicontinuity:* $\|\rho_n - \rho\|_1 \to 0 \Rightarrow H(\rho) \leq \liminf_n H(\rho_n)$.
- *Entropy increase:* $H(\rho) \leq H(\Phi(\rho))$.

Setting for quantum data compression

Let $\mathcal{H}$ be a d dimensional Hilbert space. Let a symbol set S of normalized vectors of a Hilbert space $\mathcal{H}$. WLOG assume that $\mathrm{Span}\, S = \mathcal{H}$. Each $s \in S$ is identified with the pure state $|s\rangle\langle s|$.

A quantum source emits symbols from S in an i.i.d. way according to the quantum statistical ensemble $(|s\rangle, p_s)_{s \in S}$. Thus the probability that the symbol $|s_1 s_2 \dots s_n\rangle$ (i.e. the pure state $|s_1 s_2 \dots s_n\rangle\langle s_1 s_2 \dots s_n| = |s_1\rangle\langle s_1| \otimes |s_2\rangle\langle s_2| \otimes \cdots \otimes |s_n\rangle\langle s_n|$), is emitted is equal to $p_{s_1} p_{s_2} \dots p_{s_n}$.

Find the smallest number of cubits per symbol for asymtotically lossless data recovery i.e. the smallest R such that for arbitrary $0 < \delta$ there exist arbitrarily large $n \in \mathbb{N}$, a subspace $T_{R,\delta,n}$ of $S^{\otimes n}$ with dimension at most 2^{Rn}, and a unitary map $U : \mathcal{H}^{\otimes n} \to \mathcal{H}^{\otimes n}$ which is identity when restricted to $T_{R,\delta,n}$ such that the average fidelity of any element of $T^{\perp}_{R,\delta,n}$ and its image via U is at most equal to δ: $T_{R,\delta,n} \subseteq S^{\otimes n} \subseteq \mathcal{H}^{\otimes n} \xrightarrow{U} \mathcal{H}^{\otimes n}$.

Schumacher's Theorem

Theorem (Schumacher's coding Theorem, 1995)

Let $\mathcal{H}$ be a d-dimensional Hilbert space and let S be a set of normalized vectors of $\mathcal{H}$. Assume that a quantum source emits elements of S in an i.i.d. way according to the statistical ensemble $\mathcal{E} = (|s\rangle, p_s)_{s\in S}$. Let $\rho = \sum_{s\in S} p_s |s\rangle\langle s|$ be the quantum state corresponding to this statistical ensemble. Since $\dim(\mathcal{H}) = d$ we have that $H(\rho) \leq d$. Given any $R > H(\rho)$ and $\delta > 0$ there exist arbitrarily large $n \in \mathbb{N}$, a subspace $T_{R,\delta,n}$ of $S^{\otimes n}$ of dimension at most 2^{Rn}, and a unitary operator $U : \mathcal{H}^{\otimes n} \to \mathcal{H}^{\otimes n}$ which is identity when restricted to $T_{R,\delta,n}$ and such that the average fidelity between the elements of $T^{\perp}_{R,\delta,n}$ and their images via U is at most equal to δ.

Idea of the proof of Schumacher's Theorem

$$\rho = \sum_{s \in S} p_s |s\rangle\langle s| \quad \text{(the average state).}$$

$$\text{Let } \rho = \sum_{t \in T} t|t\rangle\langle t| \quad \text{be the spectral decomposition of } \rho.$$

Thus

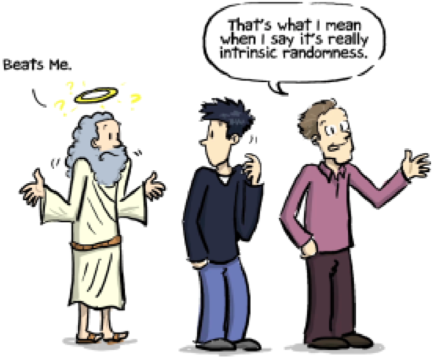
$$\rho^{\otimes n} = \sum_{t1,\ldots,t_n \in T} t_1 \cdots t_n |t_1 \cdots t_n\rangle\langle t_1 \cdots t_n| \quad \text{is the spectral decomp. of } \rho^{\otimes n}.$$

$$T := \left\{ |t_1 t_2 \cdots t_n\rangle : \frac{1}{2^{nR}} < t_1 t_2 \cdots t_n < \frac{1}{2^{n(H(X)-\delta)}} \right\} \text{ and } T_{R,\delta,n} := \mathrm{Span}(T)$$

$$\text{Then} \sum_{|t_1 t_2 \cdots t_n\rangle \in T^c_{R,\delta,n}} t_1 t_2 \cdots t_n F(|U(|t_1 t_2 \cdots t_n\rangle), |t_1 t_2 \cdots t_n\rangle) < \delta$$

where fidelity between ρ and $|v\rangle$, is $F(u, v) := \mathrm{tr}\,(u|v\rangle\langle v|) = \langle v|u|v\rangle$.

Compute the accessible information of a quantum channel



Define and give an upper bound for the accessible classical information of a q-q channel

Consider the classical communication scheme

$$\mathcal{S}(C(S)) \xrightarrow{\mathcal{C}} \mathcal{S}(\mathcal{B}(\mathcal{H})) \xrightarrow{\Phi} \mathcal{S}(\mathcal{B}(\mathcal{H})) \xrightarrow{\mathcal{D}} \mathcal{S}(C(S'))$$

via a noisy q-q channel Φ.

Alice transmits info according to a **classical statistical ensemble** $\mathcal{E} = (p_s)_{s \in S}$. Let X denote the r.v. with values in S and p.m.f. equal to $(p_s)_{s \in S}$.

Assume that Alice uses the **code** $\mathcal{C}(s) = \rho_s$ where $(\rho_s)_{s \in S}$ is a set of quantum states on some Hilbert space $\mathcal{H}$.

For every state $\Phi(\rho_s)$ that Bob receives, he performs a POVM $M = (M_{s'})_{s' \in S'}$, ($M_{s'} \geq 0$ for all $s' \in S'$ and $\sum\limits_{s' \in S'} M_{s'} = \mathbf{1}$) in order to obtain a p.m.f. (**decoding**) $\mathcal{D}(\Phi(\rho_s)) = (\operatorname{tr}(M_{s'}\Phi(\rho_s)))_{s' \in S'}$.

Give an upper bound on the accessible classical information that Bob receives.

Definition and upper bound for the accessible classical information of a q-q channel

Thus if Alice sends the symbol $s \in S$ then Bob receives the symbol $s' \in S'$ with transitional probability $p(s'|s) = \mathrm{tr}\left(M_{s'}\Phi(\rho_s)\right)$. Hence Bob receives the symbol $s' \in S'$ with probability $q_{s'} := \sum_{s \in S} p(s'|s) p_s$. Let Y be the r.v. with values in S' and p.m.f. $(q_{s'})_{s' \in S'}$.

Definition

Define the **accessible classical information** *that Bob receives to be* $Acc(\mathcal{C}, \Phi) = \sup_M I(X : Y)$.

If $(\Phi(\rho_s))_{s \in S}$ have pairwise orthogonal supports, then Bob can identify with certainty the states $\Phi(\rho_s)$ by choosing $S' = S$ and M_s to be the orthogonal projection to the support of $\Phi(\rho_s)$ for every $s \in S$. Hence $p(s'|s) = \mathrm{tr}\left(M_{s'}\Phi(\rho_s)\right) = \delta_{s,s'} \Rightarrow I(X : Y) = H(X) - H(X|Y) = H(X)$. On the other hand, if $(\Phi(\rho_s))_{s \in S}$ do not have pairwise orthogonal supports then no measurement will identify them perfectly, so $H(X|Y) > 0$ and $\mathrm{Acc}\left(\mathcal{C}, \Phi\right) < H(X)$.

Compute the classical capacity of a quantum channel

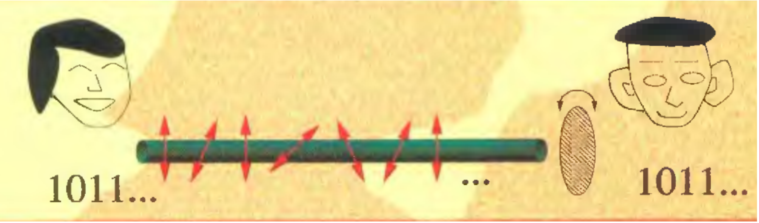


Figure 5: Detail from the cover page of “Introduction to quantum computation and information”, Lo, Popescu, Spiller (Editors), World Scientific 1998.

The Holevo Theorem

Question

Give an upper (tight!) bound on the classical capacity of a quantum channel.

Theorem (Holevo, 1973)

Consider the classical communication scheme

$$\mathcal{S}(C(\{0,1\}^{\otimes m})) \xrightarrow{\mathcal{C}_{m,n}} \mathcal{S}(\mathcal{B}(\mathcal{H})^{\otimes n}) \xrightarrow{\Phi^{\otimes n}} \mathcal{S}(\mathcal{B}(\mathcal{H})^{\otimes n}) \xrightarrow{\mathcal{D}_{m,n}} \mathcal{S}(C(\{0,1\}^{\otimes m}))$$

using repeated transmitions via a memoryless noisy q-q channel $\Phi : \mathcal{S}(\mathcal{B}(\mathcal{H})) \to \mathcal{S}(\mathcal{B}(\mathcal{H}))$*. Then* $C_c(\Phi) \leq \lim_{n\to\infty} \frac{1}{n}\chi(\Phi^{\otimes n})$.

- The definition of χ is in the next page.
- Is this inequality saturated?
- Are there easily computable upper bounds?

Holevo's χ

Definition

If $\Psi : \mathcal{S}(\mathcal{B}(\mathcal{H})) \to \mathcal{S}(\mathcal{B}(\mathcal{H}))$ *is a q-q channel, then we define Holevo's* χ *as*

$$\chi(\Psi) := \sup \left\{ H\left(\sum_x p_x \Psi(\rho_x)\right) - \sum_x p_x H(\Psi(\rho_x)) \right\}$$

where the sup *is taken w.r.t. all prob. distr.* $(p_x)_x$ *and all collections* $(\rho_x)_x$ *of density operators on* $\mathcal{H}$*.*

- By the concavity of von Neumann entropy, $\chi(\Psi) \geq 0$.
- χ plays the role of the mutual information for q-q channels.
- A better understanding for why χ is a "natural" upper bound for the capacity for the q-q channel can be understood via the **quantum mutual information** which is presented next.

Quantum relative entropy

Definition

Given two states $\rho, \sigma \in \mathcal{S}(\mathcal{B}(\mathcal{H}))$ *the* **Umegaki relative entropy** *is defined by* $D(\rho||\sigma) = \begin{cases} tr(\rho(\log_2 \rho - \log_2 \sigma)) & \text{if } supp(\rho) \subseteq supp(\sigma) \\ \infty & \text{otherwise} \end{cases}$

Theorem

- *Positivity:* $D(\rho||\sigma) \geq 0$ *and if* $D(\rho||\sigma) = 0$ *then* $\rho = \sigma$*.*
- *Joint convexity:* $D(\lambda\rho_1 + (1-\lambda)\rho_2||\lambda\sigma_1 + (1-\lambda)\sigma_2) \leq \lambda D(\rho_1||\sigma_1) + (1-\lambda)D(\rho_2||\sigma_2)$*.*
- *Additivity:* $D(\rho_1 \otimes \sigma_1||\rho_2 \otimes \sigma_2) = D(\rho_1||\rho_2) + D(\rho_2||\sigma_2)$*.*
- *Unitary invariance:* $D(U\rho U^*||U\sigma U^*) = D(\rho||\sigma)$*.*
- *Monotonicity:* $D(\Phi(\rho_1)||\Phi(\rho_2)) \leq D(\rho_1||\rho_2)$*.*
- *Lower semicontinuity:* $\|\rho_n - \rho_1\|_1 \to 0$ *and* $\|\sigma_n - \sigma\|_1 \to 0$ *imply* $D(\rho||\sigma) \leq \liminf_n D(\rho_n||\sigma_n)$*.*

Quantum mutual information

Definition (Ohya 1983)

Given a q-q channel $\Phi : \mathcal{S}(\mathcal{B}(\mathcal{H})) \to \mathcal{S}(\mathcal{B}(\mathcal{H}))$ *and a state* $\rho \in \mathcal{S}(\mathcal{B}(\mathcal{H}))$ *the* **quantum mutual information** $I(\rho, \Phi)$ *is defined by the following expression:*

$$\sup \left\{ D\left(\sum_k \mu_k E_k \otimes \Phi(E_k) \middle\| \rho \otimes \Phi(\rho) \right) : \rho = \sum_k \mu_k E_k \text{ spectral decomp.} \right\}.$$

$I(\rho, \Phi)$ indicates how much quantum information about the specific input ρ of the q-q channel Φ is conveyed about its output. Thus $\sup_\rho I(\rho, \Phi)$ represents how much quantum information about the input of the channel Φ is conveyed by its output.

Theorem (Ohya, Watanabe, 2010)

$\chi(\Phi) = \sup_\rho I(\rho, \Phi)$.

Thank you for your attention!