Introduction to Quantum Information Processing

CS 667 / PH 767 / CO 681 / AM 871

Lecture 16 (2009)

Richard Cleve
DC 2117
cleve@cs.uwaterloo.ca

Distinguishing between two arbitrary quantum states

Holevo-Helstrom Theorem (1)

Theorem: for any two quantum states ρ and σ, the optimal measurement procedure for distinguishing between them succeeds with probability $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_{tr}$ (equal prior probs.)

Proof* (the attainability part):

Since $\rho - \sigma$ is Hermitian, its eigenvalues are real

Let Π_+ be the projector onto the positive eigenspaces

Let Π_- be the projector onto the non-positive eigenspaces

Take the POVM measurement specified by Π_+ and Π_- with the associations $+ \equiv \rho$ and $- \equiv \sigma$

* The other direction of the theorem (optimality) is omitted here

Holevo-Helstrom Theorem (2)

Claim: this succeeds with probability $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_{tr}$

Proof of Claim:

A key observation is $\mathrm{Tr}(\Pi_+ - \Pi_-)(\rho - \sigma) = \|\rho - \sigma\|_{tr}$

The success probability is $p_s = \frac{1}{2}\mathrm{Tr}(\Pi_+ \rho) + \frac{1}{2}\mathrm{Tr}(\Pi_- \sigma)$

& the failure probability is $p_f = \frac{1}{2}\mathrm{Tr}(\Pi_+ \sigma) + \frac{1}{2}\mathrm{Tr}(\Pi_- \rho)$

Therefore, $p_s - p_f = \frac{1}{2}\mathrm{Tr}(\Pi_+ - \Pi_-)(\rho - \sigma) = \frac{1}{2}\|\rho - \sigma\|_{tr}$

From this, the result follows ∎

Purifications & Ulhmann's Theorem

Any density matrix ρ, can be obtained by tracing out part of some larger ***pure*** state:

$$\rho = \sum_{j=1}^{d} \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right| = \mathrm{Tr}_2\left(\sum_{j=1}^{m} \sqrt{\lambda_j}\left|\varphi_j\right\rangle\left|j\right\rangle\right)\left(\sum_{j=1}^{m} \sqrt{\lambda_j}\left\langle\varphi_j\right|\left\langle j\right|\right)$$

a purification of ρ

Ulhmann's Theorem*: The ***fidelity*** between ρ and σ is the maximum of $\langle\varphi|\psi\rangle$ taken over all purifications $|\psi\rangle$ and $|\varphi\rangle$

* See [Nielsen & Chuang, pp. 410-411] for a proof of this

Recall our previous definition of fidelity as

$$F(\rho, \sigma) = \mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}} \equiv \|\rho^{1/2}\sigma^{1/2}\|_{tr}$$

Relationships between fidelity and trace distance

$$1 - F(\rho,\sigma) \leq \|\rho - \sigma\|_{tr} \leq \sqrt{1 - F(\rho,\sigma)^2}$$

See [Nielsen & Chuang, pp. 415-416] for more details

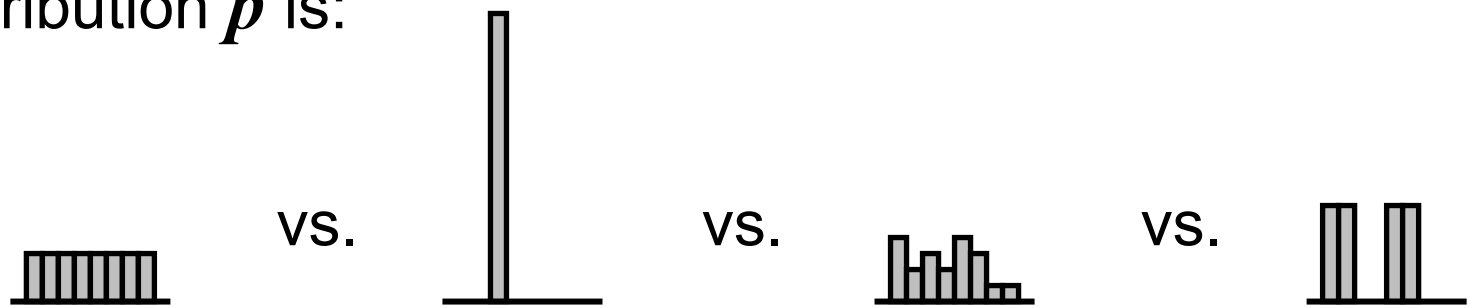
Entropy and compression

Shannon Entropy

Let $\boldsymbol{p} = (p_1,\ldots,p_d)$ be a probability distribution on a set $\{1,\ldots,d\}$

Then the (Shannon) ***entropy*** of $\boldsymbol{p}$ is $H(p_1,\ldots,p_d) = -\sum_{j=1}^{d} p_j \log p_j$

Intuitively, this turns out to be a good measure of "how random" the distribution $\boldsymbol{p}$ is:



$H(\boldsymbol{p}) = \log d$ $\quad$ $H(\boldsymbol{p}) = 0$

Operationally, $H(\boldsymbol{p})$ is the number of bits needed to store the outcome (in a sense that will be made formal shortly)

Von Neumann Entropy

For a density matrix ρ, it turns out that $S(\rho) = -\mathrm{Tr}\rho\log\rho$ is a good quantum analog of entropy

Note: $S(\rho) = H(p_1,\ldots,p_d)$, where $p_1,\ldots,p_d$ are the eigenvalues of ρ (with multiplicity)

Operationally, $S(\rho)$ is the number of ***qubits*** needed to store ρ (in a sense that will be made formal later on)

Both the classical and quantum compression results pertain to the case of large blocks of n independent instances of data:

- probability distribution $\boldsymbol{p}^{\otimes n}$ in the classical case, and
- quantum state $\rho^{\otimes n}$ in the quantum case

Classical compression (1)

Let $\boldsymbol{p} = (p_1,\ldots,p_d)$ be a probability distribution on a set $\{1,\ldots,d\}$ where n independent instances are sampled:

$(j_1,\ldots,j_n) \in \{1,\ldots,d\}^n$ (d^n possibilities, $n \log d$ bits to specify one)

Theorem*: for all $\varepsilon > 0$, for sufficiently large n, there is a scheme that compresses the specification to $n(H(\boldsymbol{p}) + \varepsilon)$ bits while introducing an error with probability at most ε

Intuitively, there is a subset of $\{1,\ldots,d\}^n$, called the "typical sequences", that has size $2^{n(H(\boldsymbol{p}) + \varepsilon)}$ and probability $1 - \varepsilon$

A nice way to prove the theorem, is based on two cleverly defined random variables …

* "Plain vanilla" version that ignores, for example, the tradeoffs between n and ε

Classical compression (2)

Define the random variable $f:\{1,\ldots,d\} \to \mathbf{R}$ as $f(j) = -\log p_j$

Note that $E[f] = \sum_{j=1}^{d} p_j f(j) = -\sum_{j=1}^{d} p_j \log p_j = H(p_1,\ldots,p_d)$

Define $g:\{1,\ldots,d\}^n \to \mathbf{R}$ as $g(j_1,\ldots,j_n) = \dfrac{f(j_1)+\cdots+f(j_n)}{n}$

Thus $E[g] = H(p_1,\ldots,p_d)$

Also, $g(j_1,\ldots,j_n) = -\dfrac{1}{n}\log(p_{j_1}\cdots p_{j_n})$

Classical compression (3)

By standard results in statistics, as $n \to \infty$, the observed value of $g(j_1,\ldots,j_n)$ approaches its expected value, $H(\boldsymbol{p})$

More formally, call $(j_1,\ldots,j_n) \in \{1,\ldots,d\}^n$ ***ε-typical*** if

$$\left|g(j_1,\ldots,j_n) - H(p)\right| \le \varepsilon$$

Then, the result is that, for all $\varepsilon > 0$, for sufficiently large n,

$$\Pr[(j_1,\ldots,j_n) \text{ is } \varepsilon\text{-typical}] \ge 1 - \varepsilon$$

We can also bound the ***number of*** these ε-typical sequences:

- By definition, each such sequence has probability $\ge 2^{-n(H(\boldsymbol{p}) + \varepsilon)}$
- Therefore, there can be at most $2^{n(H(\boldsymbol{p}) + \varepsilon)}$ such sequences

Classical compression (4)

In summary, the compression procedure is as follows:

The input data is $(j_1,\ldots,j_n) \in \{1,\ldots,d\}^n$, each independently sampled according the probability distribution $\boldsymbol{p} = (p_1,\ldots,p_d)$

The compression procedure is to leave $(j_1,\ldots,j_n)$ intact if it is ε-typical and otherwise change it to some fixed ε-typical sequence, say, $(j,\ldots,j)$ (which will result in an error)

Since there are at most $2^{n(H(\boldsymbol{p})+\varepsilon)}$ ε-typical sequences, the data can then be converted into $n(H(\boldsymbol{p})+\varepsilon)$ bits

The error probability is at most ε, the probability of an atypical input arising

Quantum compression (1)

The scenario: n independent instances of a d-dimensional state are randomly generated according some distribution:

$$\begin{cases} |\varphi_1\rangle & \text{prob. } p_1 \\ \vdots & \vdots \quad\; \vdots \\ |\varphi_r\rangle & \text{prob. } p_r \end{cases}$$

Example: $\begin{cases} |0\rangle & \text{prob. } \frac{1}{2} \\ |+\rangle & \text{prob. } \frac{1}{2} \end{cases}$

Goal: to "compress" this into as few qubits as possible so that the original state can be reconstructed with small error in the following sense …

The expected* trace distance between the reconstructed state and the state that was actually generated should be small

* Defined as the expected value of the trace distance, taken with respect to the randomness of the generation procedure

Quantum compression (2)

Define $\rho = \sum_{i=1}^{r} p_i \left|\varphi_i\right\rangle\left\langle\varphi_i\right|$

Theorem: for all $\varepsilon > 0$, for sufficiently large n, there is a scheme that compresses the data to $n(S(\rho) + \varepsilon)$ qubits, with expected trace distance $\leq \sqrt{2\varepsilon}$

For the aforementioned example, $\approx 0.6n$ qubits suffices

The compression method:

Express ρ in its eigenbasis as $\rho = \sum_{j=1}^{d} q_j \left|\psi_j\right\rangle\left\langle\psi_j\right|$

With respect to this basis, we will define an ε-typical subspace of dimension $2^{n(S(\rho)+\varepsilon)} = 2^{n(H(\boldsymbol{q})+\varepsilon)}$

Quantum compression (3)

The **ε-*typical subspace*** is that spanned by $|\psi_{j_1},\ldots,\psi_{j_n}\rangle$ where $(j_1,\ldots,j_n)$ is ε-typical with respect to $(q_1,\ldots,q_d)$

Define Π_{typ} as the projector into the ε-typical subspace

By the same argument as in the classical case, the subspace has dimension $\leq 2^{n(S(\rho)+\varepsilon)}$ and $\text{Tr}(\Pi_{\text{typ}}\rho^{\otimes n}) \geq 1-\varepsilon$

This is because ρ is the density matrix of $\left\{ \begin{array}{lll} |\psi_1\rangle & \text{prob.} & q_1 \\ \vdots & \vdots & \vdots \\ |\psi_d\rangle & \text{prob.} & q_d \end{array} \right.$

Quantum compression (4)

Calculation of the expected fidelity:

$$\sum_{i_1\ldots i_n} p_{i_1\ldots i_n} \left\langle \varphi_{i_1\ldots i_n} \left| \Pi_{\text{typ}} \right| \varphi_{i_1\ldots i_n} \right\rangle = \sum_{i_1\ldots i_n} p_{i_1\ldots i_n} \operatorname{Tr}\left(\Pi_{\text{typ}} \left| \varphi_{i_1\ldots i_n} \right\rangle\left\langle \varphi_{i_1\ldots i_n} \right| \right)$$

$$= \operatorname{Tr}\left(\Pi_{\text{typ}} \sum_{i_1\ldots i_n} p_{i_1\ldots i_n} \left| \varphi_{i_1\ldots i_n} \right\rangle\left\langle \varphi_{i_1\ldots i_n} \right| \right)$$

$$= \operatorname{Tr}\left(\Pi_{\text{typ}} \rho^{\otimes n} \right)$$

$$\geq 1 - \varepsilon$$

Abbreviations:

$p_{i_1\ldots i_n} = p_{i_1}\ldots p_{i_n}$

$|\varphi_{i_1\ldots i_n}\rangle = |\varphi_{i_1}\rangle\ldots|\varphi_{i_n}\rangle$

Using $\|\rho - \sigma\|_{\text{tr}} \leq \sqrt{1 - \mathrm{F}(\rho,\sigma)^2}$, we can upper bound the expected trace distance by $\sqrt{2\varepsilon}$