Introduction to Quantum Information Processing

QIC 710 / CS 678 / PH 767 / CO 681 / AM 871

Lectures 17–18 (2013)

Richard Cleve
DC 2117 / QNC 3129
cleve@cs.uwaterloo.ca

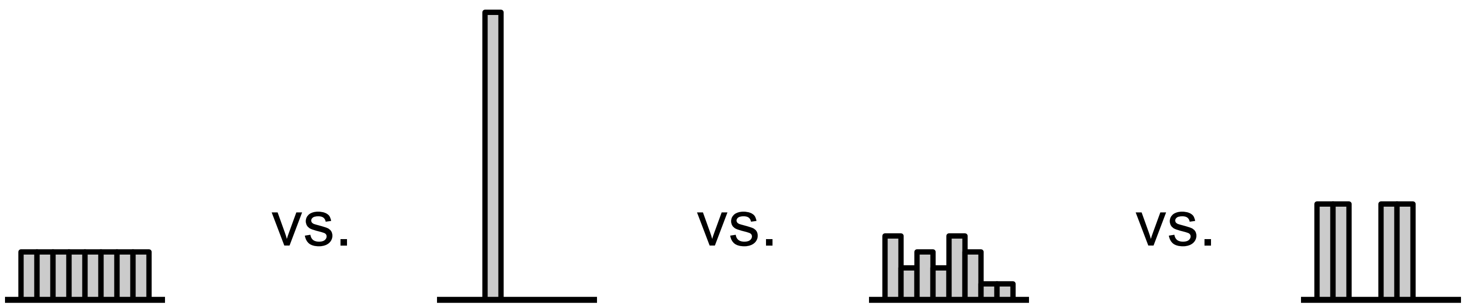
Entropy and compression

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 $S(\rho) = -\operatorname{Tr}\rho \log\rho$ is a good quantum analogue 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, $g(j_1,\ldots,j_n) = -\dfrac{1}{n}\log\left(p_{j_1}\cdots p_{j_n}\right)$

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

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

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

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

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

- By definition, each such sequence has probability $\geq 2^{-n(H(\boldsymbol{p}) + \varepsilon)}$
- 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:

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

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

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 …

ε-good:

No procedure can distinguish between these two states

(a) compressing and then uncompressing the data

(b) the original data left as is

with probability more than $\frac{1}{2} + \frac{1}{4}\varepsilon$

Quantum compression (2)

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

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

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

The compression method:

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

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 $\begin{cases} |\psi_1\rangle & \text{prob. } q_1 \\ \vdots & \vdots \quad\quad \vdots \\ |\psi_d\rangle & \text{prob. } q_d \end{cases}$

Quantum compression (4)

Calculation of the expected fidelity:

$$\sum_I p_I \langle \phi_I | \Pi_{\rm typ} | \phi_I \rangle = \sum_I p_I \mathrm{Tr}\left(\Pi_{\rm typ} | \phi_I \rangle \langle \phi_I | \right) = \mathrm{Tr}\left(\sum_I p_I \Pi_{\rm typ} | \phi_I \rangle \langle \phi_I | \right)$$

$$= \mathrm{Tr}\left(\Pi_{\rm typ} \rho^{\otimes n} \right) \quad \geq \quad 1 - \varepsilon$$

Abbreviations

$$I = i_1 i_2 \ldots i_n$$

$$p_I = p_{i_1 i_2 \ldots i_n}$$

$$|\phi_I\rangle = |\phi_{i_1} \phi_{i_2} \ldots \phi_{i_n}\rangle$$

What does this mean?

If the generated state is $|\phi_I\rangle$ and the compression process first applies the measurement $\Pi_{\rm typ}, \Pi_{\rm typ}^{\perp}$ then the success probability is $\langle \phi_I | \Pi_{\rm typ} | \phi_I \rangle$ (call outcome $\Pi_{\rm typ}$ "success")

Averaging over the possible choices of the index I, the success probability for the compression part is $\geq 1 - \varepsilon$

Quantum compression (5)

How good an approximation of the true data is the compressed state when the compression part succeeds?

The ***true data*** is of the form $(I, |\phi_I\rangle)$ where the I is generated with probability p_I

The ***approximate data*** is of the form $\left(I, \frac{1}{\gamma_I}\Pi_{\text{typ}}|\phi_I\rangle\right)$ where I is generated with probability p_I

$\gamma_I = \sqrt{\langle\phi_I|\Pi_{\text{typ}}|\phi_I\rangle}$ normalization factor

Above two states ***at least*** as hard to distinguish as these two:

$$|\Phi\rangle = \sum_I \sqrt{p_I}|I\rangle \otimes |\phi_I\rangle \qquad |\Phi'\rangle = \frac{1}{\gamma}\sum_I \sqrt{p_I}|I\rangle \otimes \Pi_{\text{typ}}|\phi_I\rangle$$

Fidelity: $\langle\Phi|\Phi'\rangle = \frac{1}{\gamma}\sum_I p_I\langle\phi_I|\Pi_{\text{typ}}|\phi_I\rangle \geq \frac{1}{\gamma}(1-\varepsilon) \geq \sqrt{1-\varepsilon}$

Trace distance: $\| \, |\Phi\rangle - |\Phi'\rangle \, \|_{\text{tr}} \leq 2\sqrt{\varepsilon}$