Quantum
Lecture 7

- The Holevo bound
- Typical sequences and subspaces
- Compression

The Holevo Bound

Assume a discrete random variable $X \in \mathcal{X}$ with pmf $p(x)$ is embedded on a set of states ρ_x, as the ensemble $\{p(x), \rho_x\}$

A measurement described by $\{M_n\}_{n=1}^{N}$ is performed, resulting in $Y \in \{1, \ldots, N\}$

The Holevo bound states that

$$I(X;Y) \leq S(\rho) - \sum_{x \in \mathcal{X}} p(x) S(\rho_x)$$

over all possible $\{M_n\}$, and with

$$\rho = \sum_{x \in \mathcal{X}} p(x) \rho_x$$

The entity

$$\chi(p(x), \rho_x) = S(\rho) - \sum_{x \in \mathcal{X}} p(x) S(\rho_x)$$

is the Holevo information of the ensemble $\{p(x), \rho_x\}$

Note that the joint entropy of the classical-quantum state

$$\sigma = \sum_{x \in \mathcal{X}} p(x) |e(x)\rangle\langle e(x)| \otimes \rho_x$$

(where $\{e(x)\}$ is a basis) is $H(p) + \sum_{x \in \mathcal{X}} p(x) S(\rho_x)$, hence

$$\chi(p(x), \rho_x) = H(p) + S(\rho) - S(\sigma)$$

$=$ mutual information between the classical and the quantum state

Fano's Inequality

For discrete random variables, consider
- X = variable of interest
- Y = observed variable
- $\hat{X} = f(Y)$ estimate of X based on Y

With $P_e = \Pr(\hat{X} \neq X)$ and $h(x) = -x \log x - (1-x) \log(1-x)$, we have Fano's inequality

$$h(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$$

Hence, in the quantum setting:
For any measurement that tries to conclude X as $\hat{X}$ from ρ,

$$h(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X) - S(\rho) + \sum_{x \in \mathcal{X}} p(x) S(\rho_x)$$

Typical Sequences

For a sequence $x^n = (x_1, \dots, x_n)$ with letters in $\mathcal{X}$ and a pmf $p(x)$ on $\mathcal{X}$, let

$$T(x^n) = -\frac{1}{n}\sum_i \log p(x_i)$$

For fixed n and $\varepsilon > 0$, let

$$\mathcal{T}_\varepsilon^{(n)} = \{x^n : |T(x^n) - H(p)| \le \varepsilon\}$$

be the set of ε-typical sequences (of length n, given p)

By the (weak) LLN, if $X^n \sim \prod_i p(x_i)$ then for any $\varepsilon > 0$ there is an N such that for all $n > N$

$$\Pr(X^n \in \mathcal{T}_\varepsilon^{(n)}) > 1 - \varepsilon$$

We also have

$$|\mathcal{T}_\varepsilon^{(n)}| \le 2^{n(H(p)+\varepsilon)}$$

and there is an N such that for $n \ge N$

$$|\mathcal{T}_\varepsilon^{(n)}| \ge (1-\varepsilon)2^{n(H(p)-\varepsilon)}$$

Compression

We can enumerate all elements of $\mathcal{T}_\varepsilon^{(n)}$ using numbers from $[1 : M_n]$ with $M_n \geq \lceil 2^{n(H(p)+\varepsilon)} \rceil$

Assume $X^n \sim \prod_i p(x_i)$

Compression code: Observe $X^n = x^n$; if $x^n \in \mathcal{T}_\varepsilon^{(n)}$ then produce $i \in [1 : M_n]$ corresponding to x^n; if $x^n \notin \mathcal{T}_\varepsilon^{(n)}$ then declare error

For any $\varepsilon > 0$, there is an N such that for all $n > N$, $\Pr(\text{error}) \leq \varepsilon$ as long as

$$\frac{1}{n} \log M_n \geq H(p) + \varepsilon + \frac{1}{n}$$

On the other hand, from Fano's inequality

$$\Pr(\text{error}) \frac{\log M_n}{n} + \frac{1}{n} \geq H(p) - \frac{1}{n} \log M_n$$

Hence, for large n, choosing $n^{-1} \log M_n$ slightly bigger than $H(p)$ is the best compression we can accomplish

Preservation of Entanglement

For discrete random variables X and Y with join pmf $p(x,y)$, the mutual information $I(X;Y)$ measures the degree of mutual dependence, or (nonlinear) correlation

In quantum systems, two states are dependent on each-other if they are entangled

Consider a mixed state ρ in $\mathcal{H}$ with purification $|\psi\rangle$ in $\mathcal{H} \otimes \mathcal{R}$, i.e. $\rho = \mathrm{Tr}_{\mathcal{R}}|\psi\rangle\langle\psi|$ for some space $\mathcal{R}$

$\mathcal{R}$ can model the unknown environment; if we had access to both $\mathcal{H}$ and $\mathcal{R}$ then we would be considering the pure state $|\psi\rangle\langle\psi|$

The system $\mathcal{H}$ is entangled with the environment $\mathcal{R}$, as characterized by the entangled state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{R}$

Assume $\mathcal{E}$ is applied to ρ in $\mathcal{H}$, resulting in the state σ in $\mathcal{H} \otimes \mathcal{R}$. Then, the entanglement fidelity of $(\rho, \mathcal{E})$ is defined as

$$F(\rho, \mathcal{E}) = \langle\psi|\sigma|\psi\rangle$$

$F(\rho, \mathcal{E})$ does not depend on $\mathcal{R}$, $\quad 0 \le F(\rho, \mathcal{E}) \le 1$

We can easily verify that

$$F(\rho, \mathcal{E}) = (F(|\psi\rangle\langle\psi|, \sigma))^2$$

where $F(|\psi\rangle\langle\psi|, \sigma)$ is the regular (static) fidelity between the pure state $|\psi\rangle\langle\psi|$ and σ $\quad$ (remember $F(\rho, \sigma) = \mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}$)

$F(\rho, \mathcal{E})$ measures how well entanglement is preserved by $\mathcal{E}$

Let $\{E_i\}$ be the operation elements of $\mathcal{E}$, then we also have

$$F(\rho, \mathcal{E}) = \sum_i |\mathrm{Tr}(\rho E_i)|^2$$

Typical Subspaces

Any density operator ρ associated with a system $\mathcal{H}$ has an eigen-decomposition $\rho = \sum_i \lambda_i |x_i\rangle\langle x_i|$

Since $\sum_i \lambda_i = 1$, we can interpret this representation for ρ as an information source; $|x_i\rangle$ is emitted with probability $p(x_i) = \lambda_i$

Let $\rho^n = \rho \otimes \cdots \otimes \rho$, $|x^n\rangle = |x_{i_1} \cdots x_{i_n}\rangle = |x_{i_1}\rangle \otimes \cdots \otimes |x_{i_n}\rangle$ and $\mathcal{H}^n = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$ (n times)

The states ρ^n and $|x^n\rangle$ correspond to "using the information source" $(\rho, \mathcal{H})$ a number of n independent times

With $T(|x^n\rangle) = -n^{-1} \sum_{m=1}^{n} \log p(x_{i_m})$ let

$$\mathcal{T}_\varepsilon^{(n)} = \{|x^n\rangle : |T(|x^n\rangle) - S(\rho)| \leq \varepsilon\}$$

and define the typical subspace

$$\mathcal{S}_\varepsilon^{(n)} = \text{span}\, \mathcal{T}_\varepsilon^{(n)} = \text{span}\{|x^n\rangle : |x^n\rangle \in \mathcal{T}_\varepsilon^{(n)}\}$$

Let $P_\varepsilon^{(n)}$ denote the projection operator from $\mathcal{H}^n$ to $\mathcal{S}_\varepsilon^{(n)}$

For any $\varepsilon > 0$ there is an N such that for $n > N$

$$\mathrm{Tr}(P_\varepsilon^{(n)} \rho^n) \geq 1 - \varepsilon$$

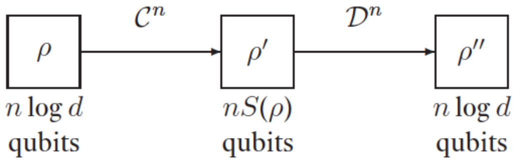
Furthermore, for any n and ε

$$\mathrm{Tr} P_\varepsilon^{(n)} \leq 2^{n(S(\rho)+\varepsilon))}$$

and for any $\varepsilon > 0$ there is an N such that for $n > N$

$$\mathrm{Tr} P_\varepsilon^{(n)} \geq (1 - \varepsilon) 2^{n(S(\rho)-\varepsilon))}$$

Compression



$\mathcal{C}^n$ maps states in $\mathcal{H}^n$ to states in a space $\mathcal{G}_n$ of dimension D_n

$\mathcal{D}^n$ maps states in $\mathcal{G}_n$ back to states in $\mathcal{H}^n$

Assume $|\psi\rangle$ is a purification of ρ^n in $\mathcal{H}^n \otimes \mathcal{R}$, and let $\mathcal{E}^n = \mathcal{D}^n \circ \mathcal{C}^n$

Let σ^n be the resulting state in $\mathcal{H}^n \otimes \mathcal{R}$

The corresponding entanglement fidelity is

$$F(\rho^n, \mathcal{E}^n) = \langle\psi|\sigma^n|\psi\rangle$$

A compression scheme

Select $\mathcal{G}_n \supset \mathcal{S}_\varepsilon^{(n)} \Rightarrow \mathrm{Tr} P_\varepsilon^{(n)} \le D_n$

Set $\mathcal{C}^n = P_\varepsilon^{(n)}$ and $\mathcal{D}^n = I$ (identity)

Then for any $\varepsilon > 0$ there is an N such that for $n > N$

$$F(\rho^n, \mathcal{E}^n) \ge |\mathrm{Tr}(\rho^n P_\varepsilon^{(n)})|^2 \ge |1-\varepsilon|^2 \ge 1 - 2\varepsilon$$

It also holds that

$$\mathrm{Tr} P_\varepsilon^{(n)} \le 2^{n(S(\rho)+\varepsilon))}$$

Thus $F(\rho^n, \mathcal{E}^n) > 1 - 2\varepsilon$ as long as

$$\frac{1}{n}\log D_n > S(\rho) + \varepsilon$$

Converse: It can be shown that, if

$$\lim_{n\to\infty} \frac{1}{n} \log D_n < S(\rho)$$

then $F(\rho^n, \mathcal{S}^n) \to 0$ for any projector $\mathcal{S}^n$

If $\mathcal{H}$ is d-dimensional, $\mathcal{H}^n$ is d^n-dimensional; i.e. it takes $n \log d$ qubits to describe a state in $\mathcal{H}^n$

Then the best compression we can have is from $\log d$ qubits to $S(\rho)$ ($\leq \log d$) qubits, per use of the source $(\rho, \mathcal{H})$,

with preserved entanglement $F(\rho^n, \mathcal{S}^n) \to 1$

Since $1 - \sqrt{F(\rho,\sigma)} \leq V(\rho,\sigma) \leq \sqrt{1 - (F(\rho,\sigma))^2}$ we could also use $V(\rho,\sigma) = 2^{-1}\mathrm{Tr}|\rho - \sigma|$ as fidelity metric,

$$F(\rho,\sigma) \to 1 \iff V(\rho,\sigma) \to 0$$