Recall: H(X) measures the ignorance on the rv X.

Let X,Y be two rv's, with distribution p(xy).
H(XY) = H(p) as before (treat XY as a composite rv).

Fix a particular outcome for Y, say y, with X unknown.

Define $q_y = p(X|Y=y)$ as the distribution of X given Y=y.
$q_y(x) = p(xy) / \sum_x p(xy)$

$H(q_y)$ is the uncertainty of X when Y=y.

Def: Conditional entropy $H(X|Y) = \sum_y p(y) H(q_y)$.
Meaning: average (over unknown Y) uncertainty of X:

easy to remember consequence → Fact: H(X|Y) = H(XY)-H(Y)
aka "Chain rule." Proof: def+algebra

Fact: H(X|Y) = H(XY)-H(Y).
Average (over y) uncertainty of X given Y.

Def [mutual information]: I(X:Y) = H(X) - H(X|Y)

uncertainty of X ↑ before ↑ after conditioning on Y

i.e. it equals to the information about X contained in Y
= decrease in uncertainty of X due to conditioning on Y.

By "fact": I(X:Y) = H(X) + H(Y) - H(XY) = I(Y:X)

I(X:Y) is MUTUAL (information) between X & Y.

Properties of H(X), H(X|Y), I(X:Y):

1. $0 \le H(X) \le \log |\Omega|$ [obvious, but useful]
2. $H(XY) \le H(X) + H(Y)$ [called Subadditivity]
   equivalent to $I(X:Y) \ge 0$
   equivalent to $H(X|Y) \le H(X)$
   [meaning: conditioning reduces uncertainty
   knowing Y cannot hurt]
   "=" iff X, Y independent (MI=0, conditioning useless)

Proof of SA and equality condition: p505 Nielsen & Chuang.

Ideas: (i) define relative entropy $H(p||q) = \sum_x p(x) \log [p(x)/q(x)]$, (ii) show that it is nonnegative [since $-(\ln 2) (\log z) \ge 1-z$, $H(p||q) = -\sum_x p(x) \log[q(x)/p(x)] \ge \sum_x p(x) [1-q(x)/p(x)]/(\ln 2)=0$, with equality hold only iff $q(x)=p(x)\ \forall x$]. (iii) rewritting I(X:Y) as H(p(x,y)||p(x)q(y)).

Properties of H(X), H(X|Y), I(X:Y):

3. Let X_1, X_2, be (different) rv's with the same Ω
   $H(\sum_k p(k) X_k) \ge \sum_k p_k H(X_k)$

   X = rv obtained by (1) draw k, (2) draw from X_k — average entropy of X_k

   [entropy of the average ≥ average entropy]
   [follows from (2): LHS = H(X), RHS = H(X|K)]

Properties of H(X), H(X|Y), I(X:Y):

4. $H(X|Y) \ge 0$ [follows from Def: average over nonnegative entropies]
5. H(XY) = H(Y)+H(X|Y) [Chain Rule, extends to multiple rv's]
6. $H(XY) \ge H(Y)$ [follows from 4&5]
7. $H(Z) + H(XYZ) \le H(XZ) + H(YZ)$
   [called Strong Subadditivity SSA]
   Note that Z special, XY symmetric.
   As if Z added to each term in SA. (Thus the name)

   equiv to $H(Y|ZX) \le H(Y|Z)$ or $H(X|ZY) \le H(X|Z)$
   Conditioning (on a new rv) decreases conditional entropy.
   Here: H(Y|Z) on X or H(Y|Z) on Y.

Quantum analogues:

Recall $S(\rho) = H(\text{spec}(\rho))$

Let A,B be two quantum systems
ρ density matrix representing state on AB

$S(AB) = S(\rho)$, $S(A) = S(\text{tr}_B\rho)$, $S(B) = S(\text{tr}_A\rho)$.

Classical: H(X|Y) = H(XY)-H(Y)

In quantum setting, no obvious meaning to condition on one of the two systems.

Def: S(A|B) = S(AB)-S(B) — Imitate classical expression but not the meaning.

Quantum analogues:

Classical: $I(X:Y) = H(X) - H(X|Y)$

Def [quantum mutual information]:
$S(A:B) = S(A) - S(A|B) = S(A) + S(B) - S(AB)$.

Imitate classical expression, due to $S(A|B)$, meaning of $S(A:B)$ not immediately clear. (Investigate later.)

Example 1:

Suppose we have a pure state $|\psi\rangle$ on AB.

There is always a Schmidt decomposition:

$$|\psi\rangle = \sum_x \sqrt{p(x)}\, |e_x\rangle_A\, |f_x\rangle_B$$

where $\{|e_x\rangle\}$ is orthonormal in C^A, $\{|f_x\rangle\}$ o.n. in C^B.

Note that $\rho_A = \sum_x p(x)\, |e_x\rangle\langle e_x|$, $\rho_B = \sum_x p(x)\, |f_x\rangle\langle f_x|$

$S(AB) = 0$.

$S(A) = S(B) = H(p)$.

$S(A:B) = 2\, H(p)$, $S(A|B) = -H(p)$

Example 2: Consider a density matrix $\rho = \sum_x \lambda_x\, |e_x\rangle\langle e_x|$

Suppose we measure in some basis (WLOG the computation basis) and the outcome is y.

Let $|e_x\rangle = \sum_y V_{xy}\, |y\rangle$.
Make a matrix V where V_{xy} = entry for the x-col & y-row so V transforms the comp basis to the eigenbasis of ρ.

Note that $p(y) = \sum_x \lambda_x\, |V_{xy}|^2$ and $p(y|x) = |V_{xy}|^2$

The distributed given by p(y) (as a vector labeled by y) is obtained from the distribution λ_x (as a vector labeled by x) by applying the matrix D (where $D_{xy} = |V_{xy}|^2$).

In general, we say that D is a stochastic map taking X to Y. if D has nonnegative entries with columns sum to 1.

Here, the rows of D also sum to 1, and we call it doubly stochastic. It is known to be entropy nondecreasing.

Therefore $S(\rho) = H(X) \leq H(Y)$ (meas outcomes are more random than the prep)

Properties of $S(A)$, $S(A|B)$, $S(A:B)$: Like classical?

0. $S(\rho) = S(U\rho U^\dagger)$, $S(A:B)_\rho = S(A:B)_{U\otimes V\, \rho\, U^\dagger \otimes V^\dagger}$
1. $0 \leq S(A) \leq \log(\dim A)$ Y
2. $S(AB) \leq S(A) + S(B)$ [subadditivity] Y
   - equiv to $S(A:B) \geq 0$
   - equiv to $S(A|B) \leq S(A)$
   - "=" iff $\rho_{AB} = \rho_A \otimes \rho_B$ product state

   Proof: NC511. Def $S(\rho||\sigma) := \mathrm{tr}(\rho \log \rho) - \mathrm{tr}(\rho \log \sigma)$. Show $S(\rho||\sigma) \geq 0$ and $S(A:B)_\rho = S(\rho_{AB}||\rho_A \otimes \rho_B)$
3. Let $\tau_1, \tau_2, \ldots$ be states on the same system and $\{p_k\}$ a distribution. Then,
   $S(\sum_k p_k \tau_k) \geq \sum_k p_k S(\tau_k)$ Y
   [entropy of the average $\geq$ average entropy]
   Why: consider $\rho = \sum_k p_k \tau_k \otimes |k\rangle\langle k|$.
   $S(AB) = H(p) + \sum_k p_k S(\tau_k)$
   $S(A) = S(\sum_k p_k \tau_k)$, $S(B) = H(p)$. Follows from 2.

Properties in the quantum setting: Like classical analogue?

4. $S(A|B) \geq 0$ or $S(A|B) \leq 0$ N
5. $S(AB) = S(B)+S(A|B)$ [by def] Y
6. $S(AB) \geq S(B)$ or $S(AB) \leq S(B)$

7. Strong subadditivity (for any tripartitie state on ABC)
   $S(C) + S(ABC) \leq S(AC) + S(BC)$ Y
   - equiv to $S(A|BC) \leq S(A|B)$ conditioning reduces conditional entropy
   - equiv to $S(A:B|C) \geq 0$ nonnegativity of conditional QMI
   - equiv to $S(A:B) \leq S(A:BC)$ local discarding cannot ↑ QMI
   - equiv to $S(A:B)_{(I\otimes E)(\rho)} \leq S(A:B)_\rho$ $\forall$TCP E local processing cannot ↑ QMI

Proof of SSA is very involved, see N&C 11.4.1.

---

7. Strong subadditivity
   $S(C) + S(ABC) \leq S(AC) + S(BC)$
   - (i) equiv to $S(A|BC) \leq S(A|B)$
     (so, $S(A:C|B) := S(A|B) - S(A|BC) \geq 0$)
   - (ii) equiv to $S(A:B) \leq S(A:BC)$
   - (iv) equiv to $S(A:B)_{(I\otimes E)(\rho)} \leq S(A:B)_\rho$

Proof of equivalences:

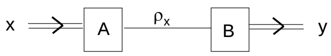
(i) $S(A|BC) = S(ABC)-S(BC)$, $S(A|B) = S(AB)-S(B)$

(ii) $S(A:B) = S(A)+S(B)-S(AB)$, $S(A:BC) = S(A)+S(BC)-S(ABC)$

(iii) recall any TCP map E can be realized by an isometry $B \rightarrow B'E$ where E is a suitable environment initially in a pure state, followed by discarding the environment.
The von Neumann entropy is invariant under a unitary change of basis. Thus $S(A:B) = S(A:B'E)$. By (ii), $S(A:B'E) \geq S(A:B')$.
Conversely, discarding is a TCP map.

How much info can we learn about a quantum state by measuring it?

Given an ensemble $\mathscr{E} = \{p_x, \rho_x\}$, consider a game:



Alice draws x wp p(x), prepares ρ_x and sends to Bob.

Bob performs meas $\mathscr{M}$ on ρ_x with operators $\{M_y\}$ ($\sum_y M_y = I$).

Probability to obtain outcome y if state is ρ_x :
$p(y|x) = \mathrm{tr}(M_y\, \rho_x)$

Joint distribution $p(xy) = p(y|x)\, p(x)$

Classical mutual info I(X:Y) quantities the information on which state X given by the measurement outcome Y

Def: $I_{acc}(\mathscr{E}) = \max_{\mathscr{M}} I(X:Y)$ [accessible info of $\mathscr{E}$]

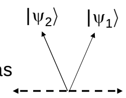
I_{acc} is a natural quantity to define but difficult to compute.

Examples (proof of optimality of meas left as Ex/HW):

e.g.1 $\rho_1 = |\psi_1\rangle\langle\psi_1|$ for $|\psi_1\rangle = \cos\theta\ |0\rangle + \sin\theta\ |1\rangle$
$\rho_2 = |\psi_2\rangle\langle\psi_2|$ for $|\psi_2\rangle = \cos\theta\ |0\rangle - \sin\theta\ |1\rangle$

drawn with $p_1 = p_2 = ½$

opt meas



$M_2 = |-\rangle\langle-|$ $M_1 = |+\rangle\langle+|$

$|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$ $|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$

Work for any ensemble with 2 pure states

x y
$p(1|1) = ½\ (\cos\theta + \sin\theta)^2 = \alpha$
$p(2|1) = ½\ (\cos\theta - \sin\theta)^2 = 1-\alpha$

$p(1|2) = ½\ (\cos\theta - \sin\theta)^2 = 1-\alpha$
$p(2|2) = ½\ (\cos\theta + \sin\theta)^2 = \alpha$

$p(1) = p(2) = ½$

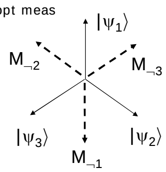
$H(Y|X) = H(\alpha)$, $H(Y) = 1$
$I_{acc} = I(X:Y)$
$= H(Y) - H(Y|X) = 1 - H(\alpha)$
$\leq H(X) = 1.$

---

e.g.2 $\rho_1 = |\psi_1\rangle\langle\psi_1|$ for $|\psi_1\rangle = |0\rangle$
$\rho_2 = |\psi_2\rangle\langle\psi_2|$ for $|\psi_2\rangle = \cos\pi/3\ |0\rangle + \sin\pi/3\ |1\rangle$
$\rho_3 = |\psi_3\rangle\langle\psi_3|$ for $|\psi_3\rangle = \cos\pi/3\ |0\rangle - \sin\pi/3\ |1\rangle$

drawn with $p_1 = p_2 = p_3 = 1/3$

opt meas



$M_{\neg k} = 2(I - |\psi_k\rangle\langle\psi_k|)/3$

$H(X|Y) = 1$
(each measurement outcomes rules out 1-out-of-3 states)

$H(X) = \log 3$

$I_{acc} = I(X:Y)$
$= \underbrace{H(X) - H(X|Y)}_{\text{do it however way is easier}} = (\log 3) - 1 \approx 0.5850$

---

e.g.3 $\rho_x = |x\rangle\langle x|$ for $x = 0,1,\cdots,n-1$ and
$\rho_{x+n} = U|x\rangle\langle x|U^\dagger$ U = fourier transform

When n=2, these are the 4 BB84 states

Each state drawn with uniform probability 1/2n.

(x is encoded in the computational or conjugate basis wp ½ each)

Optimal measurement turns out to be $M_y = ½\ \rho_y$
i.e. randomly measure in one of the two possible bases

Let TY denote Bob's entire data set, where T is the coin toss specifying his measurement basis, and Y is the outcome of that measurement.

With prob ½, Bob's random basis equals the actual one, giving Y=X, so, I(X:Y|t correct) = log n. With prob ½, he measures in the "conjugate basis" so Y is random and independent of his quantum state (elaborate). So, I(X:Y|t wrong) = 0. So, I(X:Y) = ½ log n.

---

e.g.3 $\rho_x = |x\rangle\langle x|$ for $x = 0,1,\cdots,n-1$ and
$\rho_{x+n} = U|x\rangle\langle x|U^\dagger$ U = fourier transform

When n=2, these are the 4 BB84 states

Each state drawn with uniform probability 1/2n.

(x is encoded in the computational or conjugate basis wp ½ each)

Optimal measurement turns out to be $M_y = ½\ \rho_y$
i.e. randomly measure in one of the two possible bases

We happen to find an upper bound of ½ log n for I_{acc} .

Note: if 1 more bit (which basis) is given to Bob, he can always make the correct measurement, and $I_{acc} = 1 + \log n$.

So, I_{acc} can increase by 1 + ½ log n bits when the system size increases by 1 bit. Since the increment >> the extra bit, it does not "carry" the increment, but rather "unlocks" it from the other log n qubits. (More on locking later.)

<u>A lower bound to accessible information</u>

Jozsa, Robb, Wootters 94

For a density matrix ρ in d dimensions with eigenvalues $\{\lambda_k\}$, define the "subentropy":

$$Q(\rho) = -\sum_{k=1}^{d} \left[\Pi_{l\neq k}\ \lambda_k\ /(\ \lambda_k - \lambda_l)\right]\ \lambda_k \log \lambda_k$$

For the ensemble $\mathscr{E} = \{p_x, \rho_x\}$,
$I_{acc}(\mathscr{E}) \geq Q(\sum_x p_x \rho_x) - \sum_x p_x Q(\rho_x)$ achieved by meas in random basis

If ρ_x are pure and $I/d = \sum_x p_x \rho_x$ (an ensemble of pure state that averages to the maximally mixed state),

$I_{acc} \geq \log(d) - (\log e)(1/2 + 1/3 + \ldots + 1/d)$ (in bits)

For d = 2, $I_{acc} \geq 0.2787$, for $d \to \infty$, $I_{acc} \geq \approx 0.60995$.

An upper bound to accessible information

For the ensemble $\mathcal{E} = \{p_x, \rho_x\}$, define

Holevo information $\chi(\mathcal{E}) = S(\sum_x p_x \rho_x) - \sum_x p_x S(\rho_x)$

Theorem: $I_{acc}(\mathcal{E}) \leq \chi(\mathcal{E})$

Proof:

The ensemble can be represented by the "CQ" state

$\tau_{XQ} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$

We interpret Alice as using the info x in system X to prepare the state ρ_x in system Q which is then transmitted to Bob.

Bob makes a measurement with POVM $\{M_y\}$ and outcome y stores in Y, and discards the system Q. The joint system is

$\tau'_{XY} = \sum_x p_x |x\rangle\langle x| \otimes \sum_y tr(M_y \rho_x) |y\rangle\langle y|$

Proof (ctd): $\tau_{XQY} = \sum_x p_x|x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0|$

$\tau'_{XQY} = \sum_x p_x|x\rangle\langle x| \otimes \sum_y M_y^{1/2}\rho_x M_y^{1/2} \otimes |y\rangle\langle y|$

For any measurement by Bob:

$$S(X:Y)_{\tau'} \leq S(X:Q)_{\tau}$$

by monotonicity of QMI (since the state change is a TCP map on Bob's side).

For the optimal measurement, $S(X:Y)_{\tau'} = I_{acc}(\mathcal{E})$,

whereas $S(X:Q)_\tau = S(X) + S(Q) - S(XQ)$

$= H(p) + S(\sum_x p_x \rho_x) - [H(p) + \sum_x p_x S(\rho_x)]$

$= \chi(\mathcal{E})$

Therefore $I_{acc}(\mathcal{E}) \leq \chi(\mathcal{E})$.

Note a useful fact -- the Holevo information is the QMI of the "XQ" system defining the ensemble.

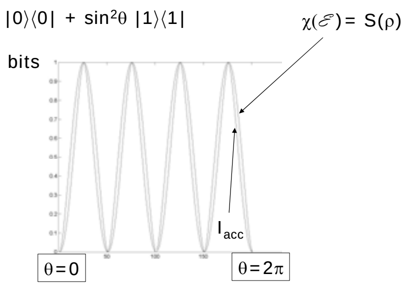
---

e.g.1 $\rho_1 = |\psi_1\rangle\langle\psi_1|$ for $|\psi_1\rangle = \cos\theta\, |0\rangle + \sin\theta\, |1\rangle$

$\rho_2 = |\psi_2\rangle\langle\psi_2|$ for $|\psi_2\rangle = \cos\theta\, |0\rangle - \sin\theta\, |1\rangle$

drawn with $p_1 = p_2 = ½$

$\rho = \cos^2\theta\, |0\rangle\langle 0| + \sin^2\theta\, |1\rangle\langle 1|$



---

e.g.2 $\rho_1 = |\psi_1\rangle\langle\psi_1|$ for $|\psi_1\rangle = |0\rangle$

$\rho_2 = |\psi_2\rangle\langle\psi_2|$ for $|\psi_2\rangle = \cos \pi/3\, |0\rangle + \sin \pi/3\, |1\rangle$

$\rho_3 = |\psi_3\rangle\langle\psi_3|$ for $|\psi_3\rangle = \cos \pi/3\, |0\rangle - \sin \pi/3\, |1\rangle$

drawn with $p_1 = p_2 = p_3 = 1/3$

$\rho = I/2$, $\chi = S(\rho) = 1$, $I_{acc} \approx 0.5850$, $Q(\rho) = 0.2787$

A beautiful result is that, given 2 iid draws of this ensemble, the best joint measurement on the 4-dim system gives more than 2*0.5850 bits of information, so I_{acc} is not additive! Will learn later that I_{acc} on many copies is nearly $n\chi$.

e.g.3 $\rho_x = |x\rangle\langle x|$ for x = 0,1,⋯,n-1 and

$\rho_{x+n} = U|x\rangle\langle x|U^\dagger$ U = fourier transform

Each state drawn with uniform probability 1/2n.

$\rho = I/n$, $\chi = S(\rho) = \log n$, $I_{acc} = ½ \log n$, $Q(\rho) \approx 0.6$ for large n

---

Note that in general, there are many many 1-qubit states, and to specify one such state takes many bits.

Preparing the quantum state (and not knowing the classical label) less than S(I/2) = 1 bit of info can be extracted.

It is highly irrreversible.

Holevo's bound also says that we cannot use 1 qbit cannot transmit more one 1 bit of data.