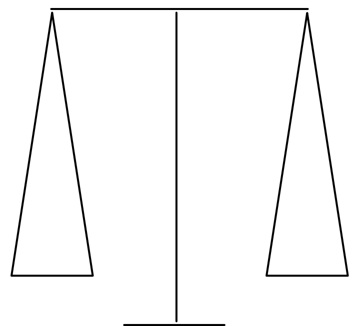
Classical Information Theory:

Game: given a scale



and twelve coins,
one of them is a
counterfeit, so, it
is lighter/heavily,

Find it with min
weighings.

Let's count.
One weighing gives 3 possible answers: L B R

How many possibilities are we distinguishing from?

Label the coins by 1, ..., 12.
Answer looks like 5+, 11-, etc. So, 24 possibilites.

Classical Information Theory:

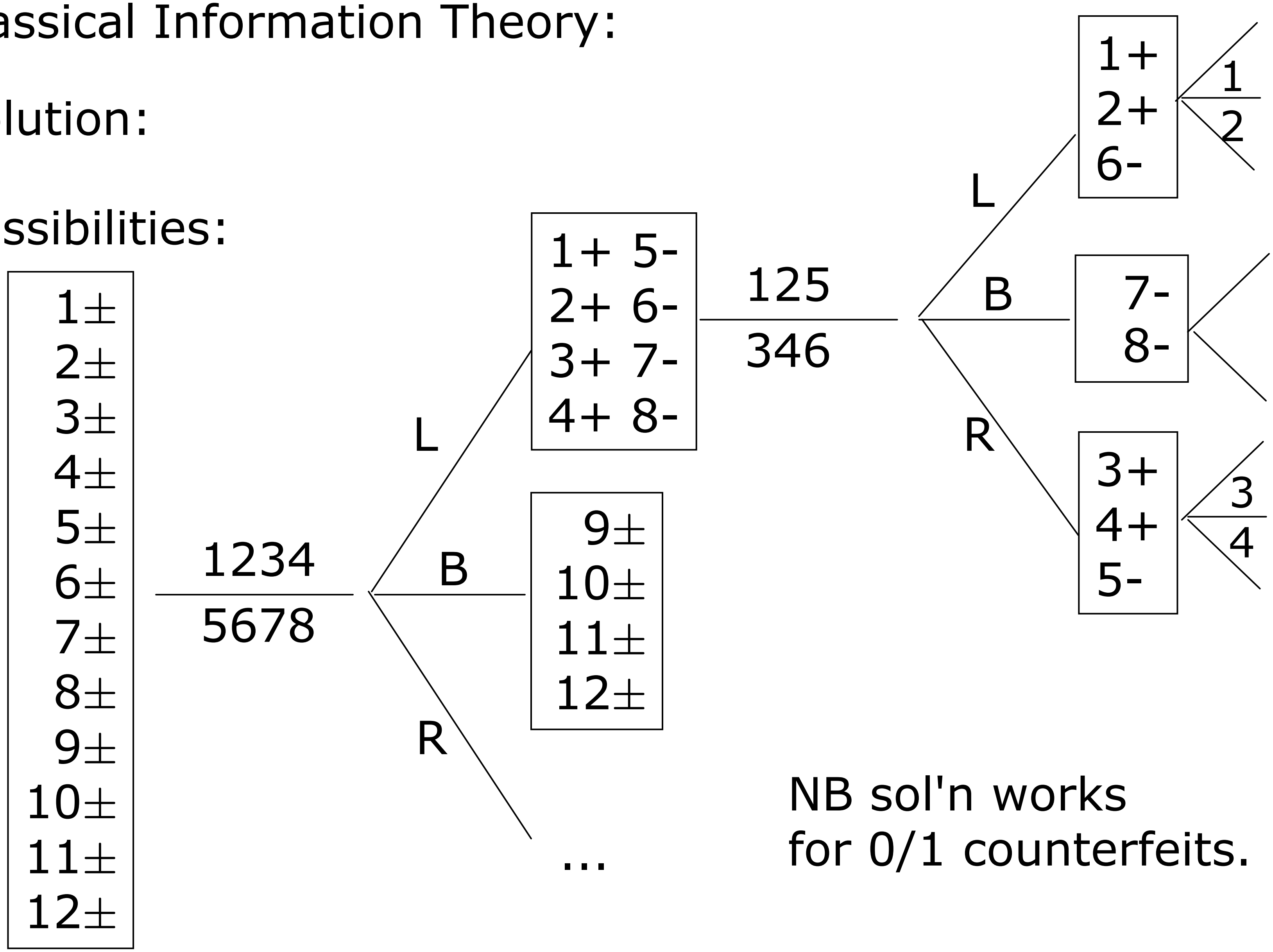
1 weighing: 3 outcomes
2 weighings: 3^2 outcomes
..
n weighings: 3^n outcomes

So, at least 3 weighings.

Classical Information Theory:

Solution:

Possibilities:



NB sol'n works
for 0/1 counterfeits.

X: random variable
Ω: sample space, say, $|\Omega| = m$
p:prob distribution of X
$p: \Omega \to [0,1]$
$x \mapsto \text{prob}(x) = p(x)$

e.g. biased coin toss
$\Omega = \{0,1\}$
$p(0) = 1/3$
$p(1) = 2/3$

iid (independent and identically drawn):
Draw from X, say, n times. How many outcomes?

Qns: m^n

How much does it take to store/represent the outcomes? e.g. in the coin toss, there are 2^n outcomes and we need n bits.

Will see, if we allow a slight risk of mistake, generally takes a lot less.

X: random variable

Ω: sample space, say, $|\Omega|$= m

p:prob distribution of X

$p: \Omega \to [0,1]$

$x \mapsto \text{prob}(x)=p(x)$

Def [Shannon Entropy]:

H(X) or H(p) := $-\sum_{x\in\Omega} p(x) \log p(x)$ [log base 2]

e.g.

For fair coin, H(X) = 1.

For biased coin defined before,

H(X) = -1/3 log(1/3) - 2/3 log(2/3)

= [log3]-2/3 = 0.91830.

X: random variable

Ω: sample space, say, $|\Omega| = m$

p:prob distribution of X

$$p: \Omega \to [0,1]$$
$$x \mapsto \text{prob}(x) = p(x)$$

Maris pointed out that this is false if $a \to p(a)$ is not injective, but in that case, convergence is even faster.

Asymptotic equipartition theorem (AEP)

n iid draws of X, outcome $x^n = x_1\, x_2 \cdots x_n$

By independence, $p(x^n) = p(x_1) \ldots p(x_n) = 2^{\sum_i \log p(x_i)}$

Let $Y = \log p(X)$, $Y_1\, Y_2 \ldots Y_n$ iid draws of Y (via X_i)

$\forall$ a, $\text{Prob}[Y = \log p(a)] = \text{prob}(X=a) = p(a)$

$1/n \sum_i \log p(x_i) = 1/n \sum_{i=1}^n Y_i$

$$\to EY = \sum_a p(a) \log(p(a)) = -H(X) \text{ as } n \to \infty$$

Thus $p(x^n) \to 2^{-nH(X)}$.

Def[typical sequence]:

x^n ε-typical if $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

It means $2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$.

Def[typical set]: $T_{n,\varepsilon} = \{x^n : x^n \ \varepsilon\text{-typical}\}$

Denote $\sum_{x^n \in T_{n,\varepsilon}} p(x^n)$ by $p(T_{n,\varepsilon})$

Consequences of AEP:

$\boxed{\forall\, \varepsilon>0,\ \forall\, \delta>0,\ \exists\ n_0 \text{ s.t. } \forall\ n \geq n_0}$

it means if n large enough, we can make ε, δ as small as we want.

1. $p(T_{n,\varepsilon}) \geq 1-\delta$
2. $(1-\delta)\ 2^{n(H(X)-\varepsilon)} \leq |T_{n,\varepsilon}| \leq 2^{n(H(X)+\varepsilon)}$

Remarks:

- ε: allowed deviation from average to be called typical
  δ: prob of non-typical

Interpretations:

"Typical x^n's are $\approx$ equiprobable (by definition)

taking up most of the prob (item 1: prob nontypical $\leq \delta$)

exponentially few (item 2: $|T_{n,\varepsilon}|/|\Omega^n| \sim 2^{-n(\log|\Omega|-H(X))}$)

Consequences of AEP:

$\forall\, \varepsilon>0, \forall\, \delta>0, \exists\, n_0$ s.t. $\forall\, n \geq n_0$

1. $p(T_{n,\varepsilon}) \geq 1-\delta$
2. $(1-\delta)\, 2^{n(H(X)-\varepsilon)} \leq |T_{n,\varepsilon}| \leq 2^{n(H(X)+\varepsilon)}$

Pf of item 1: just making AEP quantitative

Asymptotic equipartition theorem (AEP)

n iid draws of X, outcome $x^n = x_1\, x_2 \cdots x_n$

By independence, $p(x^n) = p(x_1) \ldots p(x_n) = 2^{\sum_i \log p(x_i)}$

Let $Y = \log p(X)$, $Y_1\, Y_2 \ldots Y_n$ iid draws of Y

$\forall$ a, $\text{Prob}[Y = \log p(a)] = \text{prob}(X=a) = p(a)$

$1/n \sum_i \log p(x_i) = 1/n \sum_{i=1}^n Y_i$

$\to EY = \sum_a p(a) \log(p(a)) = -H(X)$ as $n \to \infty$

$x^n \notin T_{n,\varepsilon} \Leftrightarrow |1/n \sum_i^n Y_i - EY| \geq \varepsilon$

Law of large #: $\Pr(|1/n \sum_i^n Y_i - EY| \geq \varepsilon) \leq (\text{Var } Y/n\, \varepsilon^2)$

$\leq \delta$ if $n \geq n_0 = \text{Var } Y / (\varepsilon^2\, \delta)$

Consequences of AEP:

$\forall\ \varepsilon>0,\ \forall\ \delta>0,\ \exists\ n_0$ s.t. $\forall\ n \geq n_0$

1. $p(T_{n,\varepsilon}) \geq 1-\delta$
2. $(1-\delta)\ 2^{n(H(X)-\varepsilon)} \leq |T_{n,\varepsilon}| \leq 2^{n(H(X)+\varepsilon)}$

Pf of item 2:

$$1 \geq \qquad\qquad\qquad\qquad \geq 1-\delta$$

$$p(T_{n,\varepsilon}) = \sum_{x^n \in T_{n,\varepsilon}} p(x^n)$$

$$|T_{n,\varepsilon}| \max_{x^n \in T_{n,\varepsilon}} p(x^n) \geq \qquad\qquad \geq |T_{n,\varepsilon}| \min_{x^n \in T_{n,\varepsilon}} p(x^n)$$

$$= |T_{n,\varepsilon}|\ 2^{-n\ (H(X)-\varepsilon)} \qquad\qquad = |T_{n,\varepsilon}|\ 2^{-n\ (H(X)+\varepsilon)}$$

Consequences of AEP:

$\forall\, \varepsilon>0, \forall\, \delta>0, \exists\, n_0$ s.t. $\forall\, n \geq n_0$

1. $p(T_{n,\varepsilon}) \geq 1-\delta$
2. $(1-\delta)\, 2^{n(H(X)-\varepsilon)} \leq |T_{n,\varepsilon}| \leq 2^{n(H(X)+\varepsilon)}$

<u>Application [Data compression/Shannon's noiseless coding thm]</u>

Idea: for iid $X_1, \ldots, X_n$, represents only typical outcomes and ignore the rest. Succeeds w.p. $\geq 1-\delta$, and costs only $n(H(X)+\varepsilon)$ bits.

Formally: for iid $X_1, \ldots, X_n$,

$\forall\, R>H(X), \forall\, \delta>0, \exists\, n_0$ s.t. $\forall\, n \geq n_0$

$\exists\, E_n, D_n$ s.t. $\Pr_{x^n} [D_n \circ E_n (x^n) \neq x^n] \leq \delta$

(take $\varepsilon = R-H(X)$, $T_{n,\varepsilon}$ in above.)

Converse: $\forall\, R<H(X)$, no reliable E_n, D_n (pf see N&C)

Note that data compression gives the Shannon entropy H(X) an OPERATIONAL meaning -- how much it takes to represent the data.

It also means how much uncertainty is in the data, or how much we learn by knowing it.

Will cover properties later.

Quantum analogue:

State: $\rho = \sum_v p(v)\ |e_v\rangle\langle e_v|$ (spectral decomposition)

von Neumann entropy: $S(\rho) = H(p) = -\mathrm{tr}\,(\rho \log \rho)$

Idea: $\rho \sim$ a classical rv V with dist^n p *in its eigenbasis* $|e_v\rangle$.

Now, $\rho^{\otimes n}$ is like n iid draws of V.

Let $T_{n,\varepsilon}$ be the typical set of v^n. Their corresponding eigenvectors $|e_{v^n}\rangle$ span typical subspace S with projector:

$$P_{n,\varepsilon} = \sum_{v^n \in T_{n,\varepsilon}} |e_{v^n}\rangle\langle e_{v^n}|$$

(1) $\dim S \leq 2^{n[S(\rho)+\varepsilon]}$ (2) $\mathrm{Tr}(\rho^{\otimes n} P_{n,\varepsilon}) = \sum_{v^n \in T_{n,\varepsilon}} p(v^n) \geq 1-\delta$.

Def: Let X be a classical rv with distribution $q(x)$.

$E=\{q(x),|\psi_x\rangle\}$ is called an *ensemble of quantum states*.

Interpretation: with prob $q(x)$, quantum state is $|\psi_x\rangle$.

Formally, can think of the *"CQ" state*

$$\sum_x q(x)\ |x\rangle\langle x| \otimes |\psi_x\rangle\langle\psi_x|$$

Likewise, can define $E^{\otimes n}$ as ensemble of n states, each drawn iid according to E.

How much space does it take to store these n states if we allow some small error?

Ans: $2^{n[S(\rho)+\varepsilon]}$ dimensions

where $\rho = \sum_x q(x)\ |\psi_x\rangle\langle\psi_x|$ is the average state of E

Not $2^{n[H(q)+\varepsilon]}$!!

Quantum data compression (Schumacher compression):

Let $E=\{q(x),|\psi_x\rangle\}$ be ensemble with average state ρ.

Then, $\forall\delta>0$, $\exists$ n_0 s.t. $\forall$ $n\geq n_0$, $\exists$ E_n , D_n s.t.

$$\sum_{x^n} q(x^n)\, F(|\psi_{x^n}\rangle\langle\psi_{x^n}|,\, D_n\circ E_n\,(|\psi_{x^n}\rangle\langle\psi_{x^n}|)) \geq 1-\delta$$

& E_n maps to a 2^{nR} dim space with $R>S(\rho)$.

Fidelity

Decoder & encoder

diff=ε

Allowed average error

Thus, von Neumann entropy of the average state represents the space needed for compression of iid source of quantum states.

Quantum data compression:

Let $E=\{q(x),|\psi_x\rangle\}$ be ensemble with average state ρ.

Then, $\forall\delta>0$, $\exists$ n_0 s.t. $\forall$ $n\geq n_0$, $\exists$ E_n , D_n s.t.

$$\sum_{x^n} q(x^n)\, F(|\psi_{x^n}\rangle\langle\psi_{x^n}|, D_n\circ E_n\,(|\psi_{x^n}\rangle\langle\psi_{x^n}|)) \geq 1-\delta$$

& E_n maps to a 2^{nR} dim space with $R>S(\rho)$. ($\varepsilon = R-S(\rho)$)

Proof: Let $\rho = \sum_v p(v)\ |e_v\rangle\langle e_v|$, $P_{n,\varepsilon} = \sum_{v^n\in T_{n,\varepsilon}} |e_{v^n}\rangle\langle e_{v^n}|$

NB T for v
$\neq$ T ' for x

$$E_n(\sigma) = P_{n,\varepsilon}\ \sigma\ P_{n,\varepsilon} + \mathrm{Tr}\ [(1-P_{n,\varepsilon})\sigma(1-P_{n,\varepsilon})]\ |f\rangle\langle f|$$

where $|f\rangle$ is an error (failure) symbol.

i.e. E_n encodes by projecting onto typical space of $\rho^{\otimes n}$

Each input $|\psi_{x^n}\rangle = P_{n,\varepsilon}\ |\psi_{x^n}\rangle + (1-P_{n,\varepsilon})\ |\psi_{x^n}\rangle$ (trivial identity)

Corr output = $P_{n,\varepsilon}\ |\psi_{x^n}\rangle\langle\psi_{x^n}|\ P_{n,\varepsilon}$

$+ \mathrm{Tr}[(1-P_{n,\varepsilon})\ |\psi_{x^n}\rangle\langle\psi_{x^n}|\ (1-P_{n,\varepsilon})]\ |f\rangle\langle f|$

Quantum data compression:

NB ε, δ are those of p13
for typical space of ρ

Let $E=\{q(x),|\psi_x\rangle\}$ be ensemble with average state ρ.

Then, $\forall\delta>0$, $\exists\, n_0$ s.t. $\forall\, n\geq n_0$, $\exists\, E_n$, D_n s.t.

$$\sum_{x^n} q(x^n)\, F(|\psi_{x^n}\rangle\langle\psi_{x^n}|,\, D_n\circ E_n\,(|\psi_{x^n}\rangle\langle\psi_{x^n}|)) \geq 1-\delta$$

& E_n maps to a 2^{nR} dim space with $R>S(\rho)$. ($\varepsilon = R-S(\rho)$)

Corr output $= P_{n,\varepsilon}\, |\psi_{x^n}\rangle\langle\psi_{x^n}|\, P_{n,\varepsilon}$

$\qquad + \mathrm{Tr}[(1-P_{n,\varepsilon})\, |\psi_{x^n}\rangle\langle\psi_{x^n}|\, (1-P_{n,\varepsilon})]\, |f\rangle\langle f|$

$$F(|\psi_{x^n}\rangle\langle\psi_{x^n}|,\, D_n\circ E_n\,(|\psi_{x^n}\rangle\langle\psi_{x^n}|)) = \langle\psi_{x^n}|P_{n,\varepsilon}\, |\psi_{x^n}\rangle$$

$$\sum_{x^n} q(x^n)\, F(\ldots) = \sum_{x^n} q(x^n)\langle\psi_{x^n}|P_{n,\varepsilon}\, |\psi_{x^n}\rangle$$

cyclic prop trace

$$= \sum_{x^n} q(x^n)\, \mathrm{Tr}\, [\, |\psi_{x^n}\rangle\langle\psi_{x^n}|P_{n,\varepsilon}\,]$$

$$= \mathrm{Tr}\, [\, \rho^{\otimes n}\, P_{n,\varepsilon}\,] \geq 1-\delta$$ by prop(2) p13

Converse:

If $R < S(\rho)$, no E_n, D_n will succeed in the compression.

Proof (see N&C).

Back to classical information theory ....

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

Let $q_y = p(X|Y=y)$ be the distribution of X given Y=y.

Def: Conditional entropy $H(X|Y) = \sum_y p(y)\, H(q_y)$.

i.e. it is the (average over y [entropy of X-given-y])

sensible definition

easy to remember consequence (not a definition) $\longrightarrow$ Fact: H(X|Y) = H(XY)-H(Y).

i.e. conditioning removes the uncertainty of the rv conditioned on from the joint uncertainty.

Proof: exercise.

Back to classical information theory ....

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

uncertainty of X before after conditioning on Y

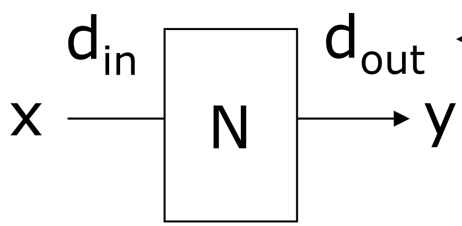
i.e. it equals to the information about X contained in Y
= decrease in uncertainty of X due to conditioning on Y.

Due to "fact": $I(X:Y) = H(X) + H(Y) - H(XY) = I(Y:X)$

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

Back to classical information theory ....

One prominent operational meaning of I(X:Y):



dimension of input and output (often called the alphabet size)

Channel:
$x \to y$ with prob $p(y|x)$

Goal: communicate as many equiprobable messages as possible per use of N, allowing many (n) uses.

The rate R is called achievable (for iid N) if, encode i
$\exists\ \delta_n, \varepsilon_n \to 0$, s.t. $\forall$ n, $\exists\ 2^{n(R-\delta_n)}$ codewords x^n_i
[each labeled by i with length n] $[x_{1i}\ x_{2i}\ \dots\ x_{ni}]$
s.t. $\exists\ D_n$ with Prob $[\ D_n(N^{\otimes n}(x^n_i)) \neq i\ ] \leq \varepsilon_n$

decoder   error   prob of error vanishing with n

Back to classical information theory ....

Channel capacity for N
= supremum over all achievable rates

$= \sup_{p(x)} I(X:Y) = \sup_{p(x)} I(X:N(X))$

Amazing ... # uses n disappear, we sup over one copy of X! Also, how on earth can we prove this?

Poll if we're to see a proof next time.

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

1. $H(X) \leq \log |\Omega|$ [obvious]
2. $H(X|Y) \leq H(X)$ [conditional reduces uncertainty]

   thus ? want prove

   (a) $I(X:Y) \geq 0$

   (b) $H(XY) \leq H(X) + H(Y)$ [subadditivity]
3. Let X_k be a rv for each k, with same Ω (diff dist^n)

   $H(\sum_k p(k) X_k) \geq \sum_k p_k H(X_k)$

   average dist^n obtained by first drawing k, then draw from X_k

   average entropy of X_k

i.e. entropy of the average $\geq$ average entropy

Why? LHS = $H(X)$, RHS = $H(X|K)$. Discarding info on K can only increase uncertainty.

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

4. $H(Z) + H(XYZ) \leq H(XZ) + H(YZ)$

   strong subadditivity (add Z to each term in SA)

5. For p(x) and q(x) with $|| p - q ||_{tr} \leq \varepsilon$,

   $|H(p) - H(q)| \leq \varepsilon \log |\Omega| + H(\varepsilon)$

   i.e. H is asymptotically continuous with Lipschitz constant determined by $\log |\Omega|$.

   [Fannes inequality]

6. $I(A:BC) \leq I(A:B) + H(C)$

   The addition of a system cannot increase MI more than it's size.

   Proof: RHS - LHS

   $= H(A)+H(B)-H(AB)+H(C)-H(A)-H(BC)+H(ABC)$

   $= H(B)-H(AB)+H(C)-H(BC)+H(ABC) = \boxed{H(B)+H(C)-H(BC)} + \boxed{H(ABC)-H(AB)}$ ≥ 0

Now the quantum analogues:

Let A,B be two quantum systems, state ρ (on AB)
$S(\rho)$ defined as before. $S(A) = S(\mathrm{tr}_B\rho)$, $S(B) = S(\mathrm{tr}_A\rho)$.

no quantum analogue

~~Let $q_y = p(X|Y=y)$ be the distribution of X given Y=y.~~

no quantum analogue

~~Def: Conditional entropy $H(X|Y) = \sum_y p(y)\, H(q_y)$.~~

easy to remember consequence (not a definition) $\longrightarrow$ Fact: $H(X|Y) = H(XY)-H(Y)$.

This fact twisted to become a def.

Def: $S(A|B) = S(AB)-S(B)$

Quantum analogue:

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

uncertainty of X before after conditioning on Y

meanings don't hold anymore
nonetheless tweak as quantum def

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

Do these quantities mean anthing anymore?
Next time.

Unused materials.

Recall definitions, meansing, & properties of the following:

$H(X)$ or $H(p) := - \sum_x p_x \log p_x$

$H(X|Y) := \sum_y p_y H(X|Y=y) = H(XY)-H(Y)$

$I(X:Y) := H(X) - H(X|Y) = H(X) + H(Y) - H(XY)$

Add one more, the relative entropy

(aka Kullback-Leibler divergence, information divergence):

$$H(p||q) := \sum_x p_x \log (p_x/q_x)$$

NB. $H(p||q) \neq H(q||p)$ in general

Then [proof as exercise]:

1. $H(X) = \log |\Omega| - H(p||u)$ where u = uniform dist^n

2. $I(X:Y) := H(XY||X\otimes Y)$ or $H(w||p\otimes q)$

where w = distribution of xy, p and q are the marginals,

$\otimes$ connects independent RV.

Second lecture

Recall definitions, meanings, & properties of the following:

$H(X)$ or $H(p) := - \sum_x p_x \log p_x$

$H(X|Y) := \sum_y p_y H(X|Y=y) = H(XY)-H(Y)$

$I(X:Y) := H(X) - H(X|Y) = H(X) + H(Y) - H(XY)$

1. $H(X) \leq \log |\Omega|$
2. $H(X|Y) \leq H(X)$

   thus (a) $I(X:Y) \geq 0$

   (b) $H(XY) \leq H(X)+H(Y)$

2.1 $H(XY)=H(Y)+H(X|Y)$

2.2 (a) $H(X|Y) \geq 0$

(b) $H(XY) \geq H(Y)$

2.3 $H(XY|Z)$

$= H(Y|Z)+H(X|YZ)$

Qn: $H(XY|Z) \stackrel{?}{=} H(Y|Z)+H(X|YZ)$

Ans: Yes. Proof:

$H(XY|Z) = H(XYZ)-H(Z)$

$= H(YZ) + H(X|YZ) - H(Z)$

$= H(Y|Z) + H(X|YZ)$.

Recall definitions, meanings, & properties of the following:

3. Let X_k be a rv for each k, with same Ω (diff dist^n)

   $H(\sum_k p(k) X_k) \geq \sum_k p(k) H(X_k)$

4. $H(Z) + H(XYZ) \leq H(XZ) + H(YZ)$

5. For p(x) and q(x) with $|| p - q ||_{tr} \leq \varepsilon$,

   $|H(p) - H(q)| \leq \varepsilon \log |\Omega| + H(\varepsilon)$

6. Def: we write $X \to Y \to Z$ if $p(x,y,z) = p(x)\, p(y|x)\, p(z|y)$

It is called a Markov Chain. e.g. $Z=f(Y)$.

In general, $p(x,y,z) = p(xy)\, p(z|xy) = p(x)\, p(y|x)\, p(z|xy)$

Thus Markov condition states that Z conditionally depends only on Y but not X.

Facts:

(a) $X \to Y \to Z \Leftrightarrow p(x,z|y) = p(x|y)\, p(z|y)$ [from def]

(b) $X \to Y \to Z \Leftrightarrow Z \to Y \to X$ [follows from (a)]

(c) Data processing inequality:

If $X \to Y \to Z$, then $I(X:Y) \geq I(X:Z)$. [see Cover&Thomas]

(d) If $X \to Y \to Z$, then $I(X:Y|Z) \leq I(X:Y)$. [p32-33]

7. We want to estimate rv X (sample space Ω), via another rv Y, from which we output Z. Let $P_e = \Pr\{X \neq Z\}$.

Thm [Fanos ineq]:

$$H(P_e) + P_e \log(|\Omega|-1) \geq H(X|Y)$$

NB:

- If P_e small, so must $H(X|Y)$. In fact, $P_e=0 \Rightarrow H(X|Y)=0$.
- $P_e \geq [H(X|Y)-1] / |\Omega|$, so, if $H(X|Y)$ is large, so must P_e.

Proof: Define new rv E, E=0 if X=Z, 1 otherwise.

By property 2.3: $H(EX|Y) = H(X|Y) + \cancel{H(E|XY)}$

$$H(EX|Y) = H(E|Y) + H(X|EY)$$

$$\begin{aligned}
\text{So, } H(X|Y) &= H(E|Y) + H(X|EY) \\
&\overset{(2.2)}{\leq} H(E) + \textstyle\sum_y p(y)\, [P_e\, H(X|E=1\ Y=y) + \\
&\qquad\qquad (1-P_e)\, H(X|E=0\ Y=y)] \\
&\leq H(P_e) + P_e \log(|\Omega|-1)
\end{aligned}$$

8. Jensen's inequality:

If f convex function [i.e. $f(py+(1-p)z) \leq pf(y)+(1-p)f(z)$], $p\in[0,1]$
& X rv, then, $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.

9. Let p(x), q(x) be 2 distributions on Ω.

Information divergence, Kullback Leibler divergence, or relative entropy between p and q:

$D(p||q) = \sum_{x\in\Omega} p(x) \log[p(x)/q(x)]$.

Note: $D(p||q) \neq D(q||p)$ in general.

These prove much of the earlier properties.

Simple facts:
(a) $H(p) = \log|\Omega| - D(p||u)$ (u = uniform dist^n on Ω)
(b) $I(X:Y) = D(p(xy)||p(x)p(y))$

Thm: $D(p||q)\geq 0$, with "=" iff p=q. [Cover&Thomas p26]

Recall definitions, meanings, & properties of the following:

For $\rho = \sum_v p(v)\ |e_v\rangle\langle e_v|$ on sys A:

$S(A)_\rho$ or $S(\rho)$:= - tr $\rho \log \rho = H(p)$

For ρ on sys AB:

$S(A|B) := S(AB) - S(B)$ [no analogue to classical interpretation]

$I(A:B) := S(A) - S(A|B) = S(A) + S(B) - S(AB)$

Like classical analogue?

Properties in the quantum setting:

1. $S(A) \le \log(\dim A)$ Y

2. $S(AB) \le S(A) + S(B)$ [subadditivity] Y

$=$ iff AB in product state.

Thus (a) $I(A{:}B) \ge 0$

(b) $S(A|B) \le S(A)$

2.1 $S(AB) = S(B)+S(A|B)$ Y

2.2 (a) $S(A|B) \ge 0$ or ≤ 0 N

(b) $S(AB) \ge$ or $\le S(B)$

2.3 $S(AB|C) = S(B|C)+S(A|BC)$ Y

still holds for classical rv's

e.g. $\rho_{AB} \propto$ projector onto $|00\rangle+|11\rangle$

Like classical analogue?

Properties in the quantum setting:

3. Let ρ_k be a state for each k (on the same system)
   $S(\sum_k p(k) \rho_k) \geq \sum_k p_k S(\rho_k)$ — Y

4. Strong subadditivity
   $S(C) + S(ABC) \leq S(AC) + S(BC)$ — Y

5. For $\rho, \sigma \in B(C^d)$, with $|| \rho - \sigma ||_{tr} \leq \varepsilon$, — Y
   $|S(\rho) - S(\sigma)| \leq \varepsilon \log d + H(\varepsilon)$
   Fannes' Inequality '73

6. For $\rho, \sigma \in B(C^{dA} \otimes C^{dB})$, with $|| \rho - \sigma ||_{tr} \leq \varepsilon$, — NA
   $|S(B|A)_\rho - S(B|A)_\sigma| \leq \varepsilon 4 \log d_B + 2 H(\varepsilon)$
   independent of d_A, Alicki-Fannes '04

Like classical analogue?

9. Let ρ, σ be d-dim quantum states.

Quantum relative entropy between ρ and σ: Y

$$S(\rho||\sigma) = \mathrm{Tr}[\rho \log \rho] - \mathrm{Tr}[\rho \log \sigma]$$ for all

Once again, $S(\rho||\sigma) \neq S(\sigma||\rho)$ in general.

Simple facts:

(a) $S(\rho) = \log d - S(\rho||I/d)$

(b) $I(A:B) = S(\rho_{AB}||\rho_A \otimes \rho_B)$

(c) $S(\rho||\sigma) = S(U\rho U^\dagger \,||\, U\sigma U^\dagger)$

Thm: Klein's inequality

$S(\rho||\sigma) \geq 0$, with "=" iff $\rho = \sigma$.

Thm: $S(\rho||\sigma)$ jointly convex

i.e. $S(\sum_i p_i \rho_i \,||\, \sum_i p_i \sigma_i) \leq \sum_i p_i S(\rho_i||\sigma_i)$

Proofs: see Nielsen & Chuang.

10. Lindblad-Ulhmann monotonicity

For all TCP maps Λ, $S(\rho||\sigma) \geq S(\Lambda(\rho)||\Lambda(\sigma))$.

Proof: [outline only]

(*1*) $\log(\eta\otimes\xi) = (\log\eta)\otimes I + I\otimes(\log\xi)$

Proof: elementary.

(*2*) $S(\mu\otimes\xi \,||\, \eta\otimes\xi) = S(\mu||\eta)$

Proof: use (a), the rest elementary.

Interpretation: attaching or removing an uncorrelated system does not afftect rel entropy.

(*3*) $\exists\, p_i, U_i$ s.t. $\forall\, d\times d$ matrix M, s.t.

$R(M) := \sum_i p_i U_i M U_i^\dagger = (\mathrm{tr}\, M)\, I/d$ for all M.

$I\otimes R(M_{AB}) = (\mathrm{tr}_B M_{AB})\otimes I/d$

Proof: take $p_i=1/d^2$, and U_i = generalized Pauli's.

10. Lindblad-Ulhmann monotonicity

For all TCP maps Λ, $S(\rho||\sigma) \geq S(\Lambda(\rho)||\Lambda(\sigma))$.

(1) $\log(\eta\otimes\xi) = (\log\eta)\otimes I + I\otimes(\log\xi)$

(2) $S(\mu\otimes\xi \,||\, \eta\otimes\xi) = S(\mu||\eta)$

(3) $\exists\, p_i, U_i$ s.t. $I\otimes R(M_{AB}) = (\mathrm{tr}_B M_{AB})\otimes I/d_B$

---

(4) $S(\rho_{AB}||\sigma_{AB}) \geq S(\rho_A||\sigma_A)$ [i.e. for $\Lambda = \mathrm{tr}_B$]

Proof: LHS

$$= \sum_i p_i\, S(I\otimes U_i\, \rho_{AB}\, I\otimes U_i^\dagger \,||\, I\otimes U_i\, \sigma_{AB}\, I\otimes U_i^\dagger)$$

apply simple fact (c) to every term

$$\geq S(\, I\otimes R(\rho_{AB}) \,||\, I\otimes R(\sigma_{AB})\,)$$

joint convexity

$$\overset{(3)}{=} S(\, \rho_A\otimes I/d \,||\, \sigma_A\otimes I/d\,)$$

$$\overset{(2)}{=} \text{RHS}$$

(5) any Λ consists of attaching $|0\rangle\langle 0|$, a unitary, and partial tracing.

11. Monotonicity of QMI under local operations

$$I(A:B)_{\rho_{AB}} \geq I(A:B)_{\Lambda\otimes I(\rho_{AB})}$$

Proof: $I(A:B)_{\rho_{AB}} = S(\rho_{AB}||\rho_A \otimes \rho_B)$

$\geq S(\Lambda \otimes I\,(\rho_{AB})||\,\Lambda(\rho_A) \otimes \rho_B)$

property 10

$= I(A:B)_{\Lambda\otimes I(\rho_{AB})}$

NB same for $I \otimes \Lambda$ and $\Lambda_A \otimes \Lambda_{B.}$

Coherent information:

$$I(A\rangle B) = S(B) - S(AB) = -S(B|A)$$

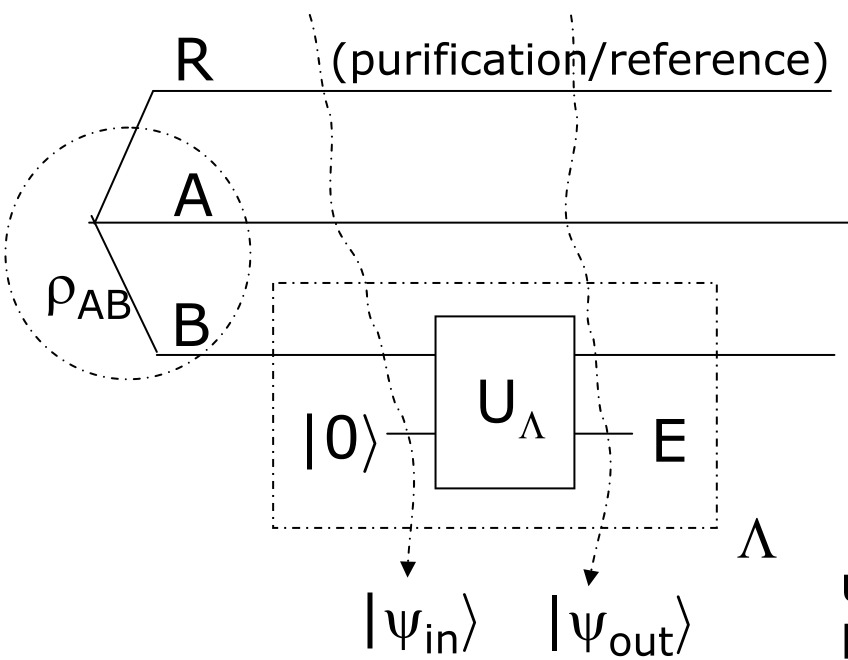
S(B) S(AB)

7. Quantum data processing inequality

$$I(A\rangle B)_{\rho_{AB}} \geq I(A\rangle B)_{\Lambda\otimes I(\rho_{AB})}$$

S(AB)

Proof: [worship in the Church of larger Hilbert space]



$$I(A\rangle B)_{\rho_{AB}} - I(A\rangle B)_{\Lambda\otimes I(\rho_{AB})}$$

$$= [S(BE)-S(R)]_{|\psi_{in}\rangle} - [S(B)-S(RE)]_{|\psi_{out}\rangle}$$

$$= [S(BE)-S(R)]_{|\psi_{out}\rangle} - [S(B)-S(RE)]_{|\psi_{out}\rangle}$$

unitary on BE, inv on R

$$= S(BE)-S(ABE) - S(B)+S(AB) \geq 0$$

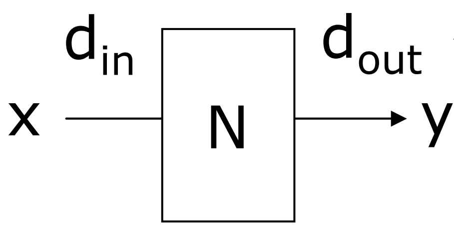
Now study capacities.

1. classical capacity of classical channels
2. classical capacity of quantum channels
3. other capacities of quantum channels

Recall from last time ...

Back to classical information theory ....

One prominent operational meaning of I(X:Y):



dimension of input and output (often called the alphabet size)

Channel:
$x \to y$ with prob $p(y|x)$

Goal: communicate as many equiprobable messages as possible per use of N, allowing many (n) uses.

The rate R is called achievable (for iid N) if,
$\exists\ \delta_n, \varepsilon_n \to 0$, s.t. $\forall$ n, $\exists\ 2^{n(R-\delta_n)}$ codewords $x^n{}_i$ encode i
[each labeled by i with length n] [$x_{1i}\ x_{2i}\ \ldots\ x_{ni}$]
s.t. $\exists\ D_n$ with Prob [$D_n(N^{\otimes n}(x^n{}_i)) \neq i$] $\leq \varepsilon_n$

decoder   error   prob of error vanishing with n

Back to classical information theory ....

Channel capacity for N
= supremum over all achievable rates

$= \sup_{p(x)} I(X:Y) = \sup_{p(x)} I(X:N(X))$

Amazing ... # uses n disappear, we sup over one copy of X! Also, how on earth can we prove this?

---

1. Show that the above is an achievable rate by finding coding schemes that achieves it. This step is called "direct coding."
1'. This is not easy. Instead, analyze a code drawn at random, and show Prob(it works) > 0. This is called an existential proof.
2. Show one cannot beat the above rate -- this is called a "converse."

Recall:

Def[typical sequence]:

x^n ε-typical if $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

It means $2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$.

Def[Jointly typical sequence]:

$x^n y^n$ ε-jointly-typical if

(a) $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

(b) $|-1/n \log(p(y^n)) - H(Y)| \leq \varepsilon$

(c) $|-1/n \log(p(x^n y^n)) - H(XY)| \leq \varepsilon$

where $p(x^n y^n) = \Pi_{i=1}^n p(x_i y_i)$.

[The strong typicality equivalence of (c) implies those of (a,b).]

Def[Jointly-typical set]: $A_{n,\varepsilon} = \{x^n y^n \ \varepsilon\text{-jointly typical}\}$

Joint asymptotic equipartition (Joint AEP) theorem:

Let (X^n,Y^n) be sequences of length n

drawn iid according to $p(x^n\, y^n) = \Pi_{i=1}^n\, p(x_i\, y_i)$.

Then:

1. $\forall\ \delta>0,\ \exists\ n_0$ s.t. $\forall\ n \geq n_0$, $\Pr(X^nY^n \in A_{n,\varepsilon}) > 1-\delta$
2. $(1-\delta)\ 2^{n\,[H(XY)-\varepsilon]} \leq |A_{n,\varepsilon}| \leq 2^{n\,[H(XY)+\varepsilon]}$
3. Let W^n,Z^n be rv's (same sample space as X^n,Y^n) w/ dist^n

   $q(x^n\, y^n) = p(x^n)\, p(y^n)$.

   i.e. q is a dist^n that has the same marginal as p,

   but x^n and y^n are independent.

   Then, $\Pr_q(W^n Z^n \in A_{n,\varepsilon}) \leq 2^{-n[I(X:Y)-3\varepsilon]}$

   Also, for large n,

   $(1-\delta)\ 2^{-n[I(X:Y)+3\varepsilon]} \leq \Pr_q(W^n Z^n \in A_{n,\varepsilon})$

<u>Joint asymptotic equipartition (Joint AEP) theorem:</u>

Proof:

[1] Given ε, δ, we can apply AEP on X^n, Y^n, and $(XY)^n$.

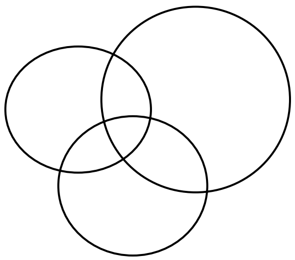
thus, $\exists n_0$ s.t. $\forall\ n \geq n_0$,

the ε-typical sets $T^X_{n,\varepsilon}$, $T^Y_{n,\varepsilon}$, $T^{XY}_{n,\varepsilon}$

all have prob $\geq 1-\delta/3$.

$A_{n,\varepsilon} = T^X_{n,\varepsilon} \cap T^Y_{n,\varepsilon} \cap T^{XY}_{n,\varepsilon}$

$A_{n,\varepsilon}{}^c = T^X_{n,\varepsilon}{}^c \cup T^Y_{n,\varepsilon}{}^c \cup T^{XY}_{n,\varepsilon}{}^c$



By the union bound,

$$\Pr(X^nY^n \in A_{n,\varepsilon}{}^c) \leq \Pr(X^nY^n \in T^X_{n,\varepsilon}{}^c) + \Pr(X^nY^n \in T^X_{n,\varepsilon}{}^c) + \Pr(X^nY^n \in T^{XY}_{n,\varepsilon}{}^c) \leq \delta$$

$\Pr(X^nY^n \in A_{n,\varepsilon}) \geq 1-\delta$.

Joint asymptotic equipartition (Joint AEP) theorem:

Proof:

[2] Using the same proof as in AEP, condition (c) implies

$$\forall\ x^n y^n \in A_{n,\varepsilon}\ ,$$

$$(1-\delta)\ 2^{-n(H(XY)+\varepsilon)} \leq p(x^n y^n) \leq 2^{-n(H(XY)-\varepsilon)}$$

Joint asymptotic equipartition (Joint AEP) theorem:

Proof:

[3] Let W^n, Z^n be rv's (same sample space as X^n, Y^n) w/ dist^n

$$q(x^n y^n) = p(x^n)\, p(y^n).$$

lower bound on $|A_{n,\varepsilon}|$ lower bounds on $p(x^n)$ and $p(y^n)$

$$(1-\delta)\, 2^{n[H(XY)-\varepsilon]} \times 2^{-n[H(X)+\varepsilon]} \times 2^{-n[H(Y)+\varepsilon]} = 2^{-n[I(X:Y)+3\varepsilon]} \leq$$

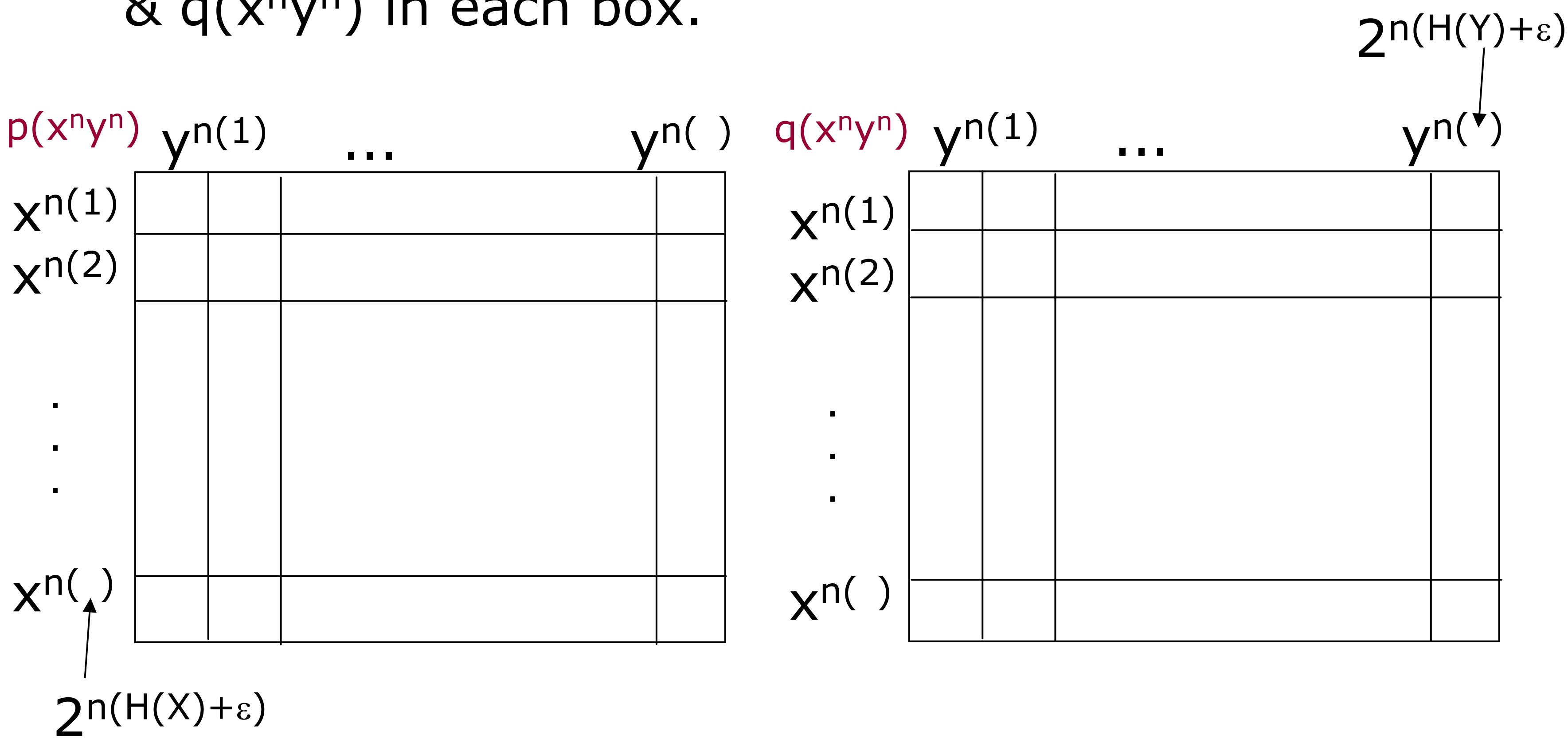
$$\boxed{\begin{array}{l} \Pr_q\left(x^n y^n \in A_{n,\varepsilon}\right) = \\ \sum_{x^n, y^n \in A_{n,\varepsilon}} p(x^n) p(y^n) \end{array}}$$

$$\leq 2^{n[H(XY)+\varepsilon]} \times 2^{-n[H(X)-\varepsilon]} \times 2^{-n[H(Y)-\varepsilon]} = 2^{-n[I(X:Y)+3\varepsilon]}$$

upper bound on $|A_{n,\varepsilon}|$ upper bounds on $p(x^n)$ and $p(y^n)$

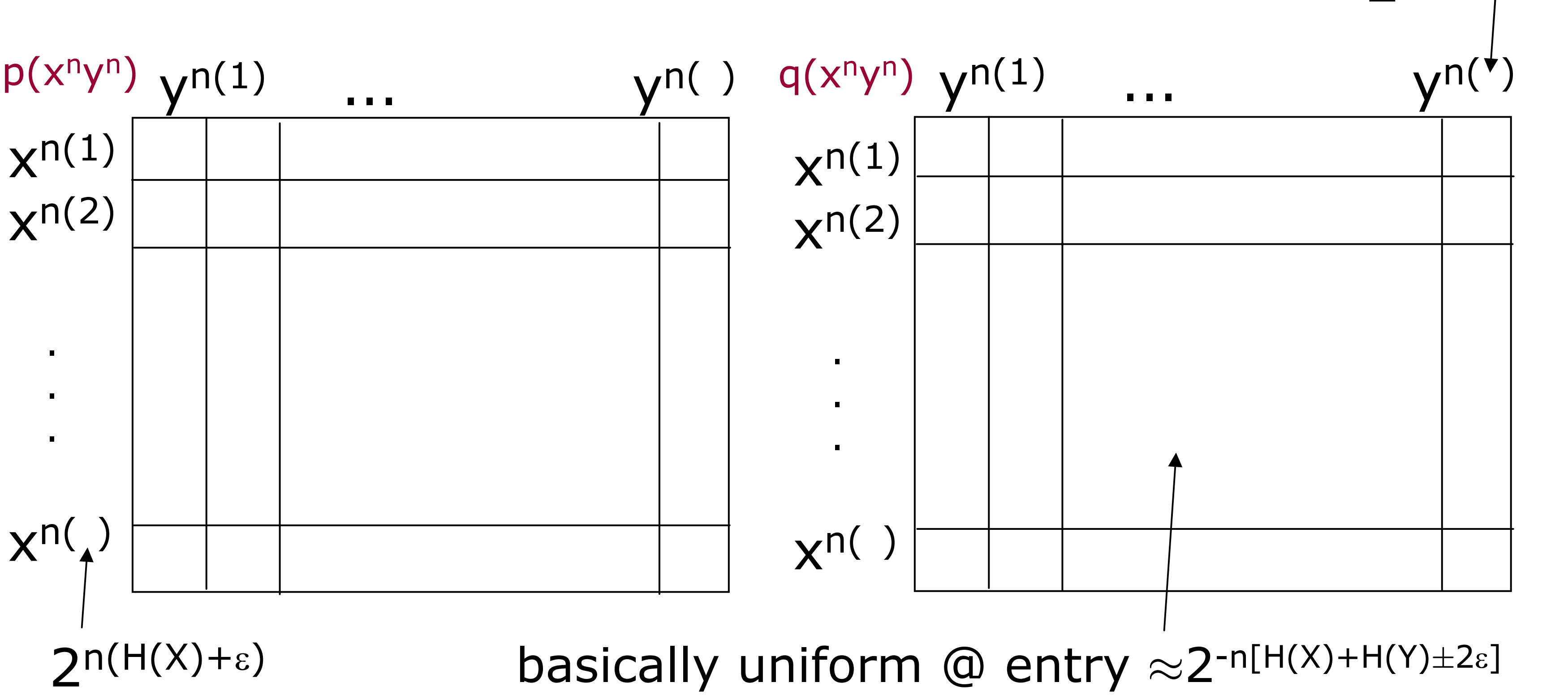
What's going on?

We're comparing 2 distributions, p and q, on $x^n y^n$.
We can list x^n's along a column, y^n's along a row.
For all purpose, only consider x^n's and y^n's typical wrt the common marginal distributions. Put $p(x^n y^n)$ & $q(x^n y^n)$ in each box.



What's going on?

1. Mostly $\approx$ 0's except for $2^{n[H(XY)+\varepsilon]}$ ($\approx$ equiprobable) entries.
2. Fix a y^n (column). $\approx 2^{n[H(X|Y)\pm 2\varepsilon]}$ "nonzero" ($\approx$equiprobable) entries [see next page]. Now, a random x^n (row) will have prob $\approx 2^{n[H(X|Y)\pm 2\varepsilon]} / 2^{n[H(X)+\varepsilon]} = 2^{n[I(X:Y)\pm 3\varepsilon]}$ to be nonzero. Similarly for fix x^n (row). So, LHS $\approx\propto$ 0/1 matrix with $\approx$ equal row & column sums. AEP[3] holds row/column-wise.



3rd lecture

Recall:

Def[typical sequence]:

x^n ε-typical if $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

It means $2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$.

Def[Jointly typical sequence]:

$x^n y^n$ ε-jointly-typical if

(a) $|-1/n \log(p(x^n)) - H(X)| \leq \varepsilon$

(b) $|-1/n \log(p(y^n)) - H(Y)| \leq \varepsilon$

(c) $|-1/n \log(p(x^n y^n)) - H(XY)| \leq \varepsilon$

where $p(x^n\, y^n) = \Pi_{i=1}^n\, p(x_i\, y_i)$.

[The strong typicality equivalence of (c) implies those of (a,b).]

Def[Jointly-typical set]: $A_{n,\varepsilon} = \{x^n y^n\ \varepsilon\text{-jointly typical}\}$

Joint asymptotic equipartition (Joint AEP) theorem:

Let (X^n, Y^n) be sequences of length n
drawn iid according to $p(x^n y^n) = \Pi_{i=1}^{n} p(x_i\ y_i)$.

Then:

1. $\forall\ \delta>0,\ \exists\ n_0$ s.t. $\forall\ n \geq n_0$, $\Pr(X^nY^n \in A_{n,\varepsilon}) > 1-\delta$
2. $(1-\delta)\ 2^{n\,[H(XY)-\varepsilon]} \leq |A_{n,\varepsilon}| \leq 2^{n\,[H(XY)+\varepsilon]}$

3. Let W^n, Z^n be rv's (same sample space as X^n, Y^n) w/ dist^n
$q(x^n\ y^n) = p(x^n)\ p(y^n)$.
i.e. q is a dist^n that has the same marginal as p,
but outcomes x^n, y^n are independent.

Then, $\Pr_q (W^n Z^n \in A_{n,\varepsilon}) \leq 2^{-n[I(X:Y)-3\varepsilon]}$

Also, for large n,

$(1-\delta)\ 2^{-n[I(X:Y)+3\varepsilon]} \leq \Pr_q (W^n Z^n \in A_{n,\varepsilon})$

Joint asymptotic equipartition (Joint AEP) theorem:

Proof:

[1] Given ε, δ, we can apply AEP on X^n, Y^n, and $(XY)^n$.

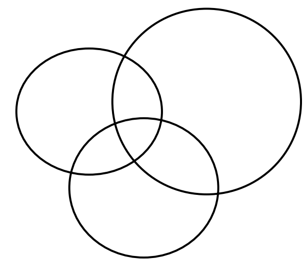
thus, $\exists n_0$ s.t. $\forall\ n \geq n_0$,

the ε-typical sets $T^X_{n,\varepsilon}$, $T^Y_{n,\varepsilon}$, $T^{XY}_{n,\varepsilon}$

all have prob $\geq 1-\delta/3$.

$A_{n,\varepsilon} = T^X_{n,\varepsilon} \cap T^Y_{n,\varepsilon} \cap T^{XY}_{n,\varepsilon}$

$A_{n,\varepsilon}{}^c = T^X_{n,\varepsilon}{}^c \cup T^Y_{n,\varepsilon}{}^c \cup T^{XY}_{n,\varepsilon}{}^c$



By the union bound,

$$\Pr(X^nY^n \in A_{n,\varepsilon}{}^c) \leq \Pr(X^nY^n \in T^X_{n,\varepsilon}{}^c) + \Pr(X^nY^n \in T^X_{n,\varepsilon}{}^c) + \Pr(X^nY^n \in T^{XY}_{n,\varepsilon}{}^c) \leq \delta$$

$$\Pr(X^nY^n \in A_{n,\varepsilon}) \geq 1-\delta.$$

Joint asymptotic equipartition (Joint AEP) theorem:

Proof:

[2] Using the same proof as in AEP, condition (c) implies

$\forall\ x^n y^n \in A_{n,\varepsilon}$,

$(1-\delta)\ 2^{-n(H(XY)+\varepsilon)} \leq p(x^n y^n) \leq 2^{-n(H(XY)-\varepsilon)}$

Joint asymptotic equipartition (Joint AEP) theorem:

Proof:

[3] Let W^n, Z^n be rv's (same sample space as X^n, Y^n) w/ dist^n

$$q(x^n\, y^n) = p(x^n)\, p(y^n).$$

lower bound on $|A_{n,\varepsilon}|$ lower bounds on $p(x^n)$ and $p(y^n)$

$$(1-\delta)\, 2^{n[H(XY)-\varepsilon]} \times 2^{-n[H(X)+\varepsilon]} \times 2^{-n[H(Y)+\varepsilon]} = 2^{-n[I(X:Y)+3\varepsilon]} \leq$$

$$\boxed{\begin{array}{l} Pr_q\,(x^n\, y^n \in A_{n,\varepsilon}) = \\ \sum_{x^n, y^n \in A_{n,\varepsilon}} p(x^n)p(y^n) \end{array}}$$

$$\leq 2^{n[H(XY)+\varepsilon]} \times 2^{-n[H(X)-\varepsilon]} \times 2^{-n[H(Y)-\varepsilon]} = 2^{-n[I(X:Y)-3\varepsilon]}$$

upper bound on $|A_{n,\varepsilon}|$ upper bounds on $p(x^n)$ and $p(y^n)$

More observations:

Given $y^n \in T^Y_{n,\varepsilon}$, how many $x^n \in T^X_{n,\varepsilon}$ is s.t. $x^n\, y^n \in A_{n,\varepsilon}$?

Call this set S_{y^n}.

$p(x^n|y^n) = p(x^n y^n) \,/\, p(y^n) \approx 2^{-n[H(XY)-H(Y)]} = 2^{-n[H(X|Y)]}$

$\uparrow$ since $x^n y^n \in A_{n,\varepsilon}$,

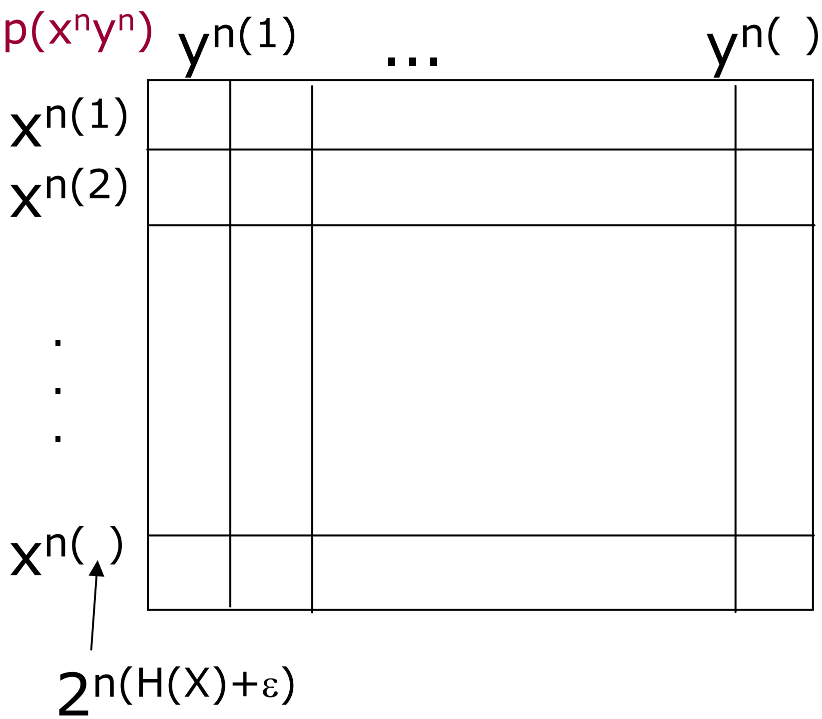
$1 = \sum_{x^n \in S} p(x^n|y^n) \approx |S_{y^n}|\; 2^{-n[H(X|Y)]}$

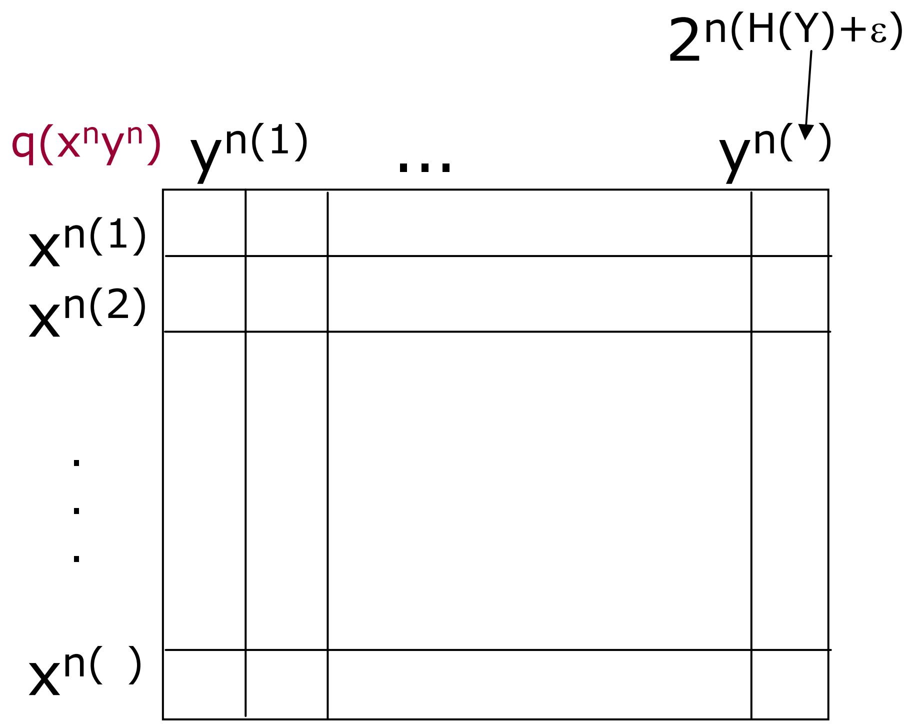
Hence, $|S_{y^n}| \approx 2^{nH(X|Y)}$. Fraction of such $x^n \approx 2^{-nI(X:Y)}$.

Similarly, given $x^n \in T^X_{n,\varepsilon}$, $\approx 2^{nH(Y|X)}$ y^n's are jointly typical with it, and the fraction of such $y^n \approx 2^{-nI(X:Y)}$.

What's going on?

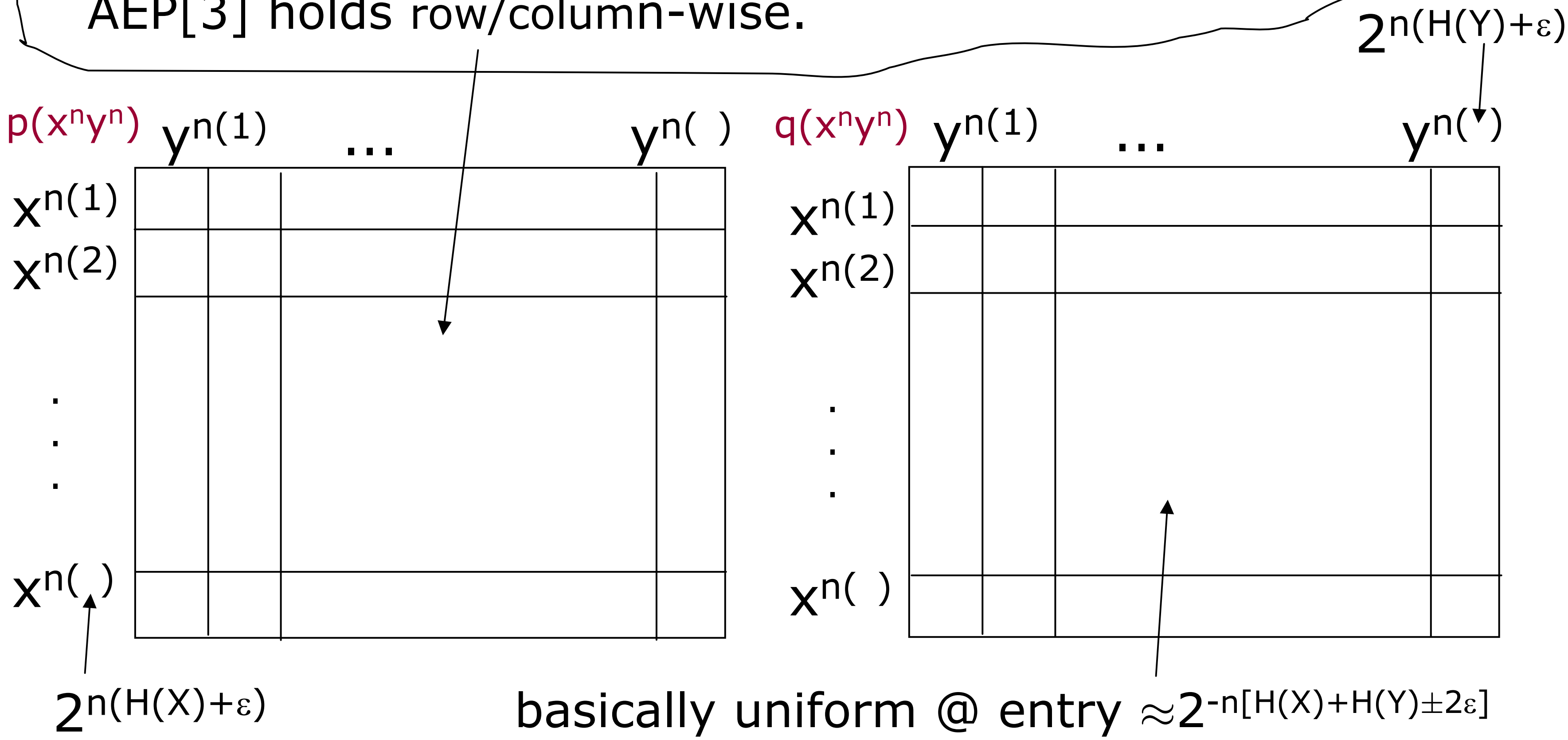
We're comparing 2 distributions, p and q, on $x^n y^n$.
We can list x^n's along a column, y^n's along a row.
Can focus only on x^n's , y^n's typical wrt to the common marginal dist^n's. Put $p(x^n y^n)$,$q(x^n y^n)$ in each box.



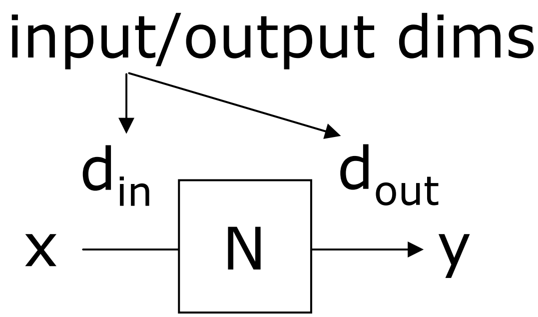


What's going on?

1. Mostly $\approx$ 0's except for $2^{n[H(XY)+\varepsilon]}$ ($\approx$ equiprobable) entries.
2. Fix a y^n (column). $\approx 2^{n[H(X|Y)\pm 2\varepsilon]}$ "nonzero" ($\approx$equiprobable) entries. A random entry (row) $x^n y^n$ is nonzero with prob $\approx 2^{n[H(X|Y)\pm 2\varepsilon]} / 2^{n[H(X)+\varepsilon]} = 2^{n[I(X:Y)\pm 3\varepsilon]}$. Similarly for fix x^n (row). So, LHS $\propto$ 0/1 matrix with $\approx$ equal row & column sums. AEP[3] holds row/column-wise.



Now ready for Shannon's noisy coding theorem.



$N: x \to y$ with
prob $p(y|x)$

The rate R is called achievable if, $\forall$ n,

$\exists\ \eta_n, \zeta_n \to 0$, E_n, D_n encoder & decoder s.t.

$\max_M \Pr(D_n \circ E_n(M) \neq M) \leq \zeta_n$, $M \in \{1, \cdots, k = 2^{n(R-\eta_n)}\}$.

With rules still TBD: Note notation recycling.

$E_n(M) = x_M$ (labeled by M with length n) $= [x_{M1}\ x_{M2}\ \ldots\ x_{Mn}]$

D_n takes y^n to some W.

Channel capacity for N := sup over all achievable rates

$= \sup_{p(x)} I(X:Y) = \sup_{p(x)} I(X:N(X))$

Proof structure:

1. Direct coding theorem:

a. Show $\forall$ p(X), I(X:Y) is an achievable rate by analyzing the prob of failure of a random code and random message. That it vanishes $\Rightarrow$ $\exists$ at least one code with vanishing average prob of error.

b. Choose a subset of better codewords that gives vanishing worse case prob of error.

2. Converse: At any higher rate, prob of error $\not\to$ 0.

Part 1a. Let $R=I(X:Y)-\eta$ (will find η).

Need E_n, D_n with prob error $\leq \zeta_n$

* Fix any $p(x)$.
* Write down $A_{\varepsilon,n}$ for XY with $pr(Y=y|X=x)$ given by N.
* $\forall$ n (fixed from now on) let $k=2^{n(R-\eta_n)}$. (Will find η_n.)

E_n: Pick k codewords (each x_{Mj} chosen iid $\sim p(x)$).
Call it $\mathcal{C}_n$. Fixed & known to Alice & Bob once choosen.

$$
\begin{aligned}
x_1 &= x_{11}, x_{12}, \ldots, x_{1n}\\
x_2 &= x_{21}, x_{22}, \ldots, x_{2n}\\
&\;\;.\\
&\;\;.\\
x_K &= x_{k1}, x_{k2}, \ldots, x_{kn}
\end{aligned}
$$

Everything refers to this particular code $\mathcal{C}_n$ from now on.

Part 1a. Let R=I(X:Y)-η (will find η).

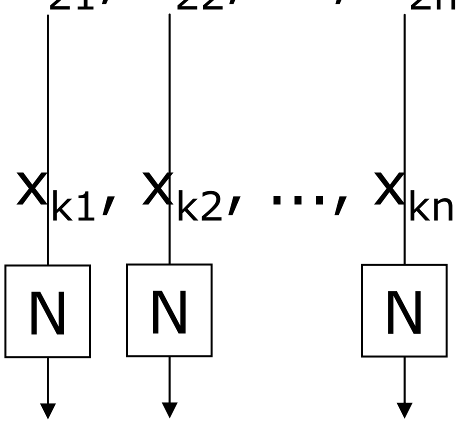
Need E_n, D_n with prob error $\leq \zeta_n$

* Fix any p(x).
* Write down $A_{\varepsilon,n}$ for XY with pr(Y=y|X=x) given by N.
* $\forall$ n (fixed from now on) let $k=2^{n(R-\eta_n)}$. (Will find η_n.)

E_n: Pick k codewords (each x_{Mj} chosen iid $\sim$ p(x)).
Call it $\mathcal{C}_n$. Fixed & known to Alice & Bob once choosen.

$x_1 = x_{11}, x_{12}, \ldots, x_{1n}$
$x_2 = x_{21}, x_{22}, \ldots, x_{2n}$ ← Say, M=2.
.
.
$x_K = x_{k1}, x_{k2}, \ldots, x_{kn}$

$E_n(2)=x_2=x_{21}\ x_{22} \ldots x_{2n}$



n rvs' Y_1 Y_2 Y_n

Let $y^n = y_1\ y_2 \ldots y_n$ be received.
$pr(y^n|x_M) = \Pi_{i=1}^{n}\ pr(y_i|x_{Mi})$

D_n: typical set decoding

Given y^n, let $S_{y^n} = \{x^n \mid x^n y^n \in A_{\varepsilon,n}\}$.

If there is a unique $x^n \in S_y$, output W s.t. $E_n(W)=x^n$.

Else, output W=k+1 (representing an error).

In what ways will this fail?

Either - no such x^n $\quad$ Err_0

- or $\exists M' \neq M$ with $E_n(M')y^n \in A_{\varepsilon,n}$ $\quad$ $Err_{M'}$

Prob of error for a given message M for code $\mathcal{C}_n$:

$$\lambda_M(\mathcal{C}_n) = \Pr(W \neq M \mid M \mathcal{C}_n) = \Pr(Err_0 \cup_{M' \neq M} Err_{M'} \mid M \mathcal{C}_n)$$

Worse case prob of error: $P_e^{max}(\mathcal{C}_n) = \max_M \lambda_M(\mathcal{C}_n)$

Ave (arithmetic) prob of error: $P_e^{ave}(\mathcal{C}_n) = 1/k \sum_M \lambda_M(\mathcal{C}_n)$

Now, upper bound, for this n:

$$\Pr_{\mathcal{C}_n} [\ P_e^{ave}(\mathcal{C}_n)\ ]$$

* just many iid draws to $X \sim p(x)$

wrt a particular $\mathcal{C}_n$ but averaged over M.

$$= \Pr_{\mathcal{C}_n} [\ 1/k \sum_M \lambda_M (\mathcal{C}_n)\ ]$$

each M chosen similarly thus λ_M independent of M

$$= \Pr_{\mathcal{C}_n} \lambda_1 (\mathcal{C}_n)$$

$$= \Pr_{\mathcal{C}_n} (W \neq 1 | M=1) = \Pr_{\mathcal{C}_n}(Err_0 \cup_{M' \neq 1} Err_{M'} | M=1)$$

union bdd

$$\leq \Pr_{\mathcal{C}_n} (Err_0 | M=1) + (k-1) \Pr_{\mathcal{C}_n}(Err_{M' \neq 1} | M=1)$$

Bounding $\Pr_{\mathcal{C}_n}(\mathrm{Err}_0 | M=1)$:

By joint AEP [1], $\forall\ \delta>0,\ \exists\ n_0$ s.t. $\forall\ n \geq n_0$,

$$\Pr(X^n Y^n \in A_{n,\varepsilon}) > 1-\delta$$

Given n, $\exists\ \delta_n,\ \varepsilon_n$ for which $\Pr(X^n Y^n \in A_{n,\varepsilon_n}) > 1-\delta_n$.

[And $\delta_n, \varepsilon_n \to 0$.]

Here:

$x_{M=1} = x_{11} \ldots x_{1n}$ drawn iid $\sim p(x)$, and

$y^n = y_1 \ \ldots\ y_n$ drawn $\sim p(y | x_{1i})$

Thus, $x_{1i} y_i$ iid $\sim p(xy)$ and $\Pr(x_{M=1}\ y^n \in A_{n,\varepsilon_n}) > 1-\delta_n$.

$\Pr_{\mathcal{C}_n}(\mathrm{Err}_0 | M=1) \leq \delta_n$.

BACK 1 SLIDE.

for 1 M' $\qquad\qquad\qquad\qquad$ $N^{\otimes n}(x_1)$

Bounding $\Pr_{\mathcal{C}_n}(\mathrm{Err}_{M'\neq 1}|M=1) = \Pr_{\mathcal{C}_n}(x_{M'}y^n \in A_{n,\varepsilon_n})$:

By joint AEP [3], $\forall\, \delta>0,\ \exists\, n_0$ s.t. $\forall\, n \geq n_0$,

$$W^n, Z^n \sim q(x^n\, y^n) = p(x^n)\, p(y^n).$$

$$(1-\delta)\, 2^{-n[I(X:Y)+3\varepsilon]} \leq \Pr_q(W^n Z^n \in A_{n,\varepsilon}) \leq 2^{-n[I(X:Y)-3\varepsilon]}$$

Given n, $\exists\, \delta_n, \varepsilon_n$ for which

$$(1-\delta_n)\, 2^{-n[I(X:Y)+3\varepsilon_n]} \leq \Pr_q(W^n Z^n \in A_{n,\varepsilon_n}) \leq 2^{-n[I(X:Y)-3\varepsilon_n]}$$

[And $\delta_n, \varepsilon_n \to 0$.]

Here:

$x_{M'} = x_{M'1} \ldots x_{M'n}$ drawn independent of x_1 and

$y^n = y_1 \ldots y_n$ iid $\sim p(y|x_{1i})$, independent of $x_{M'}$.

W^n, Z^n $\qquad$ Thus, $\Pr_{\mathcal{C}_n}(\mathrm{Err}_{M'\neq 1}|M=1) \leq 2^{-n[I(X:Y)-3\varepsilon_n]}$.

Now, upper bound, for this n:

$$\Pr_{\mathcal{C}_n} [\, P_e^{ave}(\mathcal{C}_n) \,]$$

$$\leq \Pr_{\mathcal{C}_n}(Err_0|M=1) + (k-1) \Pr_{\mathcal{C}_n}(Err_{M'\neq 1}|M=1)$$

$$\leq \delta_n + (k-1)\, 2^{-n[I(X:Y)-3\varepsilon_n]}$$

but $k=2^{n(R-\eta_n)}$, $R=I(X:Y)-\eta$

$$\leq \delta_n + 2^{n[I(X:Y)-\eta-\eta_n]}\, 2^{-n[I(X:Y)-3\varepsilon_n]}$$

$$\leq \delta_n + 2^{n[-\eta-\eta_n+3\varepsilon_n]}$$

$$\leq \delta_n + 2^{-n\eta} =: \zeta^{ave}_n \,.$$

choose η = small constant

$\eta_n = 3\varepsilon_n$.

Thus, $\exists\ \mathcal{C}_n\ (E_n, D_n)$ with $P_e^{ave}(\mathcal{C}_n) \leq \zeta^{ave}_n$.

Part 1b.

Worse case prob of error: $P_e^{max}(\mathcal{C}_n) = \max_M \lambda_M(\mathcal{C}_n)$

Ave (arithmetic) prob of error: $P_e^{ave}(\mathcal{C}_n) = 1/k \sum_M \lambda_M(\mathcal{C}_n)$

For the code $\mathcal{C}_n$ obtained in 1a, order M in ascending order of $\lambda_M(\mathcal{C}_n)$. Keep the first half. Call this new code $\mathcal{C}'_n$.

replacing large half of $\lambda_M(\mathcal{C}_n)$ by the median

$$P_e^{ave}(\mathcal{C}_n) = 1/k \sum_M \lambda_M(\mathcal{C}_n)$$
$$\geq 1/k \left[\sum_{M \notin \mathcal{C}'_n} P_e^{max}(\mathcal{C}'_n) + \sum_{M \in \mathcal{C}'_n} \lambda_M(\mathcal{C}_n) \right]$$
$$\geq 1/2\, P_e^{max}(\mathcal{C}'_n).$$

Thus, $\mathcal{C}'_n$ has worse case error prob $\leq \zeta_n^{ave}/2 =: \zeta_n \rightarrow 0$.

[rate for $\mathcal{C}'_n$ = rate for $\mathcal{C}_n$ - 1/n.]

Thus $R=I(X:Y)-\eta$ achievable on $\mathcal{C}'_n$ for any $\eta>0$.

"Sup over R" gives capacity $\geq \max_{p(x)} I(X:Y)$.

Part 2: Converse [If $P_e^{ave} \to 0$, then achievable rate $R \leq C$.]

Lemma: Let $Y^n = N^{\otimes n}(X^n)$, and C be the capacity of N.

Then, $I(X^n:Y^n) \leq nC$.

Pf: $I(X^n:Y^n) = H(Y^n) - H(Y^n|X^n)$

$= H(Y^n) - \sum_{i=1}^n H(Y_i|Y_1 \ldots Y_{i-1}X^n)$ Chain rule

$= H(Y^n) - \sum_{i=1}^n H(Y_i|X_i)$ Y_i only depends on X_i

$\leq \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|X_i)$ Subadditivity

$\leq \sum_{i=1}^n I(X_i:Y_i) = nC.$

Part 2: Converse [If $P_e^{ave} \to 0$, then achievable rate $R \leq C$.]

Lemma: Let $Y^n = N^{\otimes n}(X^n)$, and C be the capacity of N.

Then, $I(X^n:Y^n) \leq nC$.

Thm [Fanos ineq]:

$$H(P_e) + P_e \log(|\Omega|-1) \geq H(X|Y)$$

Proof of converse:

$$nR = H(M) = \underbrace{H(M|Y^n)}_{H(MY^n)-H(Y^n)} + \underbrace{I(M:Y^n)}_{-H(MY^n)+H(Y^n)+H(M)}$$

$$\leq H(M|Y^n) + I(E_n(M):Y^n) \quad \text{data processing ineq}$$

$$\leq \underbrace{1+P_e\, nR}_{\text{Fanos ineq}} + \underbrace{nC}_{\text{Lemma}}$$

Fanos ineq

$M \leftrightarrow X$

$Y^n \leftrightarrow Y$

$2^{nR} \leftrightarrow |\Omega|$

Lemma

Lecture 4 --

Obtaining classical information from quantum states and quantum channels

Concepts and definitions

- Ensemble $\mathcal{E} = \{p_m , \rho_m\}$
- Classical-Quantum state $\tau_{MQ} = \sum_m p_m |m\rangle\langle m| \otimes \rho_m$
- Holevo information for ensemble $\mathcal{E}$

$$\chi(E) := S(\sum_m p_m \rho_m) - \sum_m p_m S(\rho_m) = I(M:Q)_\tau$$

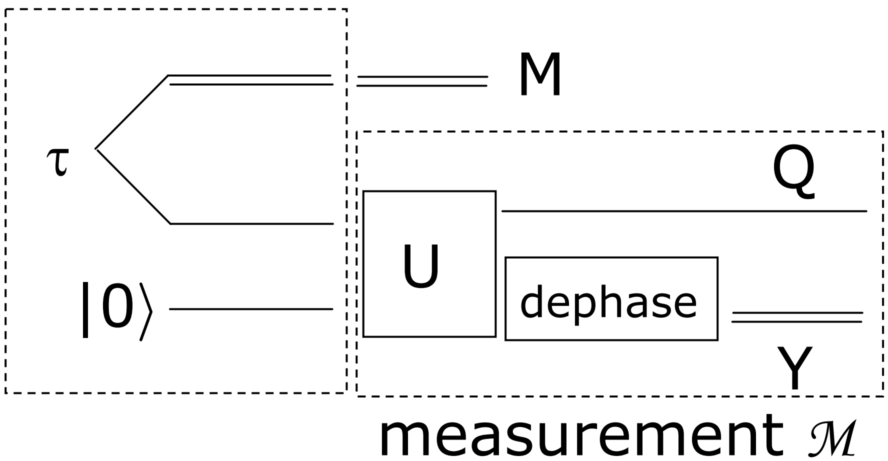
- Generalizes classical mutual information

Add additivity conjecture later.

Holevo bound (73)

For the classical-quantum state $\tau_{MQ} = \sum_m p_m \, |m\rangle\langle m| \otimes \rho_m$, let a measurement $\mathcal{M}$ be applied to Q, giving a classical outcome in register Y. Then: $I(M{:}Y) \leq I(M{:}Q)_\tau$.

Proof: the measurement attaches Y originally in state $|0\rangle$.

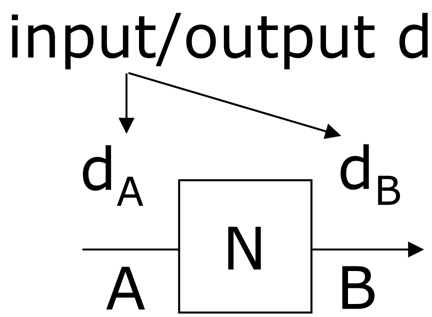


$$I(M{:}Q)_\tau = I(M{:}QY)_{\tau \otimes |0\rangle\langle 0|} \quad \text{p38,39}$$

p41, LO mono

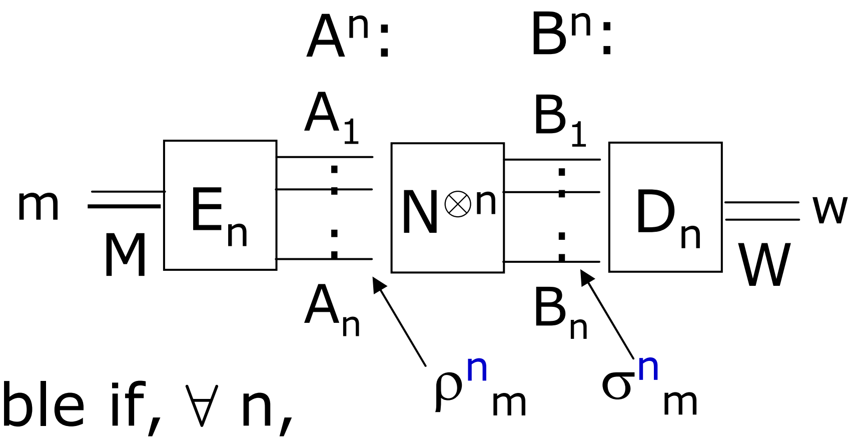
$$\geq I(M{:}QY)_{(I \otimes \mathcal{M})(\tau \otimes |0\rangle\langle 0|)} \geq I(M{:}Y)_{(I \otimes \mathcal{M})(\tau \otimes |0\rangle\langle 0|)}$$

Noisy quantum channel



N:TCP map

Send $m \in M$:



The rate R is called achievable if, $\forall$ n,

$\exists\ \eta_n, \zeta_n \to 0$, E_n, D_n encoder & decoder s.t.

$\max_M \Pr(D_n \circ E_n(m) \neq m) \leq \zeta_n$, $M \in \{1,\cdots,k=2^{n(R-\eta_n)}\}$.

With rules still TBD:

$E_n(m)=\rho^n{}_m$ (labeled by m & lives in $A_1 \otimes \ldots \otimes A_n$)

$N^{\otimes n}(\rho^n{}_m)=\sigma^n{}_m$ (lives in $B_1 \otimes \ldots \otimes B_n$). D_n takes $\sigma^n{}_m$ to some W.

C(N) = classical capacity of N := sup over all achievable rates

$= \lim_{t \to \infty} 1/t \max_\tau I(X:B^t)_\tau$ where

$\tau = \sum_x p_x |x\rangle\langle x| \otimes N^{\otimes t}(\rho^t{}_x)$ (can choose p_x, $\rho^t{}_x$)

HSW Theorem:

$C(N) = \lim_{t \to \infty} 1/t \max_\tau I(X{:}B^t)_\tau$ where

$\tau = \sum_x p_x |x\rangle\langle x| \otimes N^{\otimes t}(\rho^t{}_x)$

Will prove direct coding theorem for t=1. The achievability of the above follows by "double-blocking" -- replacing N with $N^{\otimes t}$.

Part 1a. Let $R = \max_\tau I(X:B)_\tau - \eta$ (will find η).

* Fix any $p(x)$, ρ_x. [Then $\sigma_x = N(\rho_x)$.]

* $\forall$ n (fixed from now on) let $k = 2^{n(R-\eta_n)}$. (Will find η_n.)

Need E_n, D_n with prob error $\leq \zeta_n$.

E_n: Pick k codewords ~~(each x_{Mj} chosen iid $\sim p(x)$)~~.

$x_1 = x_{11}, x_{12}, \ldots, x_{1n}$

$x_2 = x_{21}, x_{22}, \ldots, x_{2n}$

m=2

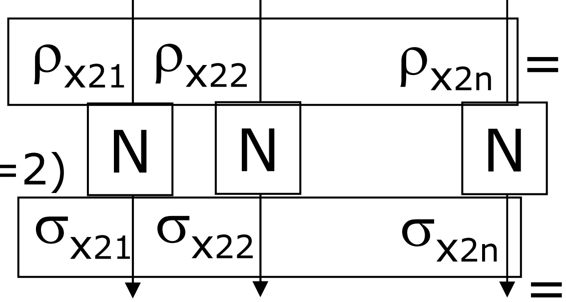
.

.

$x_k = x_{k1}, x_{k2}, \ldots, x_{kn}$

m=k

E_n(m=2)



$= \rho^n_{m=2}$

$= \sigma^n_{m=2}$

Each x^n_m randomly drawn from the strongly typical set $T^s_{n,\varepsilon n}$

An outcome of n iid draws of X, x^n, is "strongly typical" (or freq typical) if each symbol $a \in \Omega$ occurs roughly $np(a)$ times in x^n.

Let q = empirical dist^n as in x^n. Strongly typical if $||p-q||_1 \leq \varepsilon_n$.

Example -- 2 slides down

D_n: distinguishing $\sigma^n_m = \sigma_{x^n_m} = \sigma_{x_{m1}} \otimes \sigma_{x_{m2}} \ldots \sigma_{x_{mn}}$

Recall each x^n_m randomly drawn from the strongly typical set $T^s_{n,\varepsilon}$.

How does σ^n_m look like?

Let $\Omega = \{a_1, a_2, \ldots\}$.

For $i = 1,\ldots,|\Omega|$, σ^n_m has $np(a_i)\pm\varepsilon_n$ copies of σ_{ai} in some order that is known given m.

Example -- 2 slides down

$|\Omega|$ is constant but n is asymptotically large.
Knowing m, for each i, can compress the $np(a_i)\pm\varepsilon_n$ sys [in state $\sigma_{ai}^{n(p(a_i)\pm\varepsilon_n)}$] to $n(p(a_i)\pm\varepsilon_n)S(\sigma_{ai})$ qubits.

on the specific x^n_m

So, the entire σ^n_m can be compressed to a "*conditional* typical subspace" w/ $\leq \sum_i n(p(a_i)\pm\varepsilon_n)S(\sigma_{ai}) \leq n\,[\sum_i p(a_i)S(\sigma_{ai})+\eta_n]$ qubits.

Note, this is due to strong typicality, and it holds $\forall$m.

e.g.

Let $\Omega = \{1,2,3,4\}$, with $p(a) = a/10$.
Draw $X \sim p(a)$ iid $n=20$ times.

Get the following outcome:

3 3 3 4 4   2 1 3 2 2   2 4 3 4 3   4 2 4 4 3

The empirical distribution:
$q(1) = 1/20$
$q(2) = 5/20$
$q(3) = 7/20$
$q(4) = 7/20$

$||p-q||_1 = 0.2$.   So, our sequence is 0.2-strongly typical.

e.g.

Let $\Omega = \{1,2,3,4\}$, with $p(a) = a/10$.
Draw $X \sim p(a)$ iid $n=20$ times.

Get the following outcome:

3 3 3 4 4   2 1 3 2 2   2 4 3 4 3   4 2 4 4 3

Now, we have

$$\sigma_{x^n} = \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_4 \otimes \sigma_4 \otimes \sigma_2 \otimes \sigma_1 \otimes \sigma_3 \otimes \sigma_2 \otimes \sigma_2 \otimes \sigma_2 \otimes \sigma_4 \otimes \sigma_3 \otimes \sigma_4 \otimes \sigma_3 \otimes \sigma_4 \otimes \sigma_2 \otimes \sigma_4 \otimes \sigma_4 \otimes \sigma_3$$

Tensor together the typical subspaces

for σ_3, $n_3 = 7$ on systems 1,2,3,8,13,15,20
for σ_4, $n_4 = 7$ on systems 4,5,12,14,16,18,19
for σ_2, $n_2 = 5$ on systems 6,9,10,11,17
for σ_1, $n_1 = 1$ on system 6

gives the conditional typical subspace for the above outcome.

We make a general statement (disregard how the state arises & omitting m).

Lemma:

Let $\{\sigma_x\}$ and $p(x)$ be fixed.

Let $\sigma_{x^n} = \sigma_{x_1} \otimes \sigma_{x_2} \ldots \sigma_{x_n}$, x_i drawn iid, $x^n = x_1 \cdots x_n$.

Let Π_{x^n} = projection onto the conditional typical subspace.

$\forall$ n, $\exists$ ε_n , $\delta_n \rightarrow 0$ s.t.:

Proof ideas -- just follow the procedure outlined earlier & control the $|\Omega|$ small terms.

1. $\mathrm{Tr}[\sigma_{x^n} \Pi_{x^n}] \geq 1-\delta_n$

2. $\mathrm{Tr}[\Pi_{x^n}] \leq 2^{n[\Sigma_x p(x) H(\sigma_x)+\varepsilon_n]}$

homework

3. $\mathrm{Tr}[\sigma_{x^n} \Pi] \geq 1-\delta_n$ if x^n strongly typical, and Π projector onto typical subspace of $\sigma=\Sigma_x p(x)\sigma_x$

4. $[\Pi\, \sigma^{\otimes n}\, \Pi] \leq 2^{-n[H(\sigma)-\varepsilon_n]}\, \Pi$

from quantum data compression

Back to direct coding for HSW: Want to find D_n that distinguishes $\sigma^n{}_m = \sigma_{x^n{}_m} = \sigma_{x_{m1}} \otimes \sigma_{x_{m2}} \ldots \sigma_{x_{mn}}$

Lemma:

Let $\{\sigma_x\}$ and $p(x)$ be fixed.

Let $\sigma_{x^n} = \sigma_{x_1} \otimes \sigma_{x_2} \ldots \sigma_{x_n}$, x_i drawn iid, $x^n = x_1 \cdots x_n$.

Let Π_{x^n} = projection onto the conditional typical subspace.

$\forall$ n, $\exists\ \varepsilon_n$, $\delta_n \to$ s.t.:

1. $\mathrm{Tr}[\sigma_{x^n} \Pi_{x^n}] \geq 1-\delta_n$

2. $\mathrm{Tr}[\Pi_{x^n}] \leq 2^{n[\Sigma_x p(x) H(\sigma_x) + \varepsilon_n]}$

3. $\mathrm{Tr}[\sigma_{x^n} \Pi] \geq 1-\delta_n$ if x^n strongly typical

These mean that a typical message σ_{x^n} received by Bob occupies $\approx n\Sigma_x\, p(x)H(\sigma_x)$ qubit of space.

& they all live in the typical space of σ, size $2^{nH(\sigma)}$

Thus can have at most $\approx 2^{n[H(\sigma) - \Sigma_x p(x) H(\sigma_x)]}$ distinguishable messages. To achieve it, need to "pack" the messages well :)

Def: Let $S=\{\zeta_m\}$ be a set of quantum states.

The distinguishability error of S is defined as:

$$de(S) = 1 - \max_{\{F_m\}\ \mathrm{POVM}} 1/|S| \sum_m Tr(F_m \zeta_m)$$

Packing lemma: Notations as above. Let p(m) be a distribution and $\zeta = \sum_p p(m)\, \zeta_m$. Suppose $\exists\ \Pi_m, \Pi$ s.t.

(1) $Tr(\zeta_m\, \Pi) \geq 1-\varepsilon$

(2) $Tr(\zeta_m\, \Pi_m) \geq 1-\varepsilon$

(3) $Tr(\Pi_m) \leq d_1$

(4) $\Pi\, \zeta\, \Pi \leq \Pi/d_0$.

Let $X_1, \ldots, X_n$ be iid $\sim$ p(m).

$S' = \{\zeta_{xi}\}$. ($|S'|=k$.)

$k = \lfloor (d_0/d_1)\gamma \rfloor$ for $0<\gamma<1$.

Then,

E $de(S') \leq 2[\varepsilon+\sqrt{(8\varepsilon)}]+4\gamma$.

(E=expection)

Given: $\zeta = \sum_m p(m)\, \zeta_m$.

(1) $\mathrm{Tr}(\zeta_m \Pi) \geq 1-\varepsilon$, (2) $\mathrm{Tr}(\zeta_m \Pi_m) \geq 1-\varepsilon$, (3) $\mathrm{Tr}(\Pi_m) \leq d_1$, (4) $\Pi\zeta\Pi \leq \Pi/d_0$.

$X_1, \ldots, X_k$ iid $\sim p(m)$, $S' = \{\zeta_{xi}\}$, $k = \lfloor (d_0/d_1)\gamma \rfloor$, $0<\gamma<1$.

Claim: $E\, de(S') \leq 2[\varepsilon + \sqrt{(8\varepsilon)}] + 4\gamma$.

Proof: Let $\Lambda_m = \Pi\, \Pi_m\, \Pi$, $Z = \sum_{i=1}^k \Lambda_{xi}$

Take $F_i = Z^{-1/2} \Lambda_{xi} Z^{-1/2}$ for the POVM elements (PGM).

$\sum_{i=1}^k F_i = Z^{-1/2} \sum_{i=1}^k \Lambda_{xi} Z^{-1/2} = Z^{-1/2}\, Z\, Z^{-1/2} = I_{supp(Z)} \leq I.$

Can add $F_{err} = I - I_{supp(Z)}$ to complete the POVM.

$de(S') \leq 1 - 1/k \sum_i \mathrm{Tr}(\zeta_{xi} F_i) = 1/k \sum_i \mathrm{Tr}[\zeta_{xi} (I-F_i)]$

Aside: useful operator ineq for this

$I - (X+Y)^{-1/2} X (X+Y)^{-1/2} \leq 2(I-X) + 4Y.$

Write $Z = \Lambda_{xi} + \sum_{j \neq i} \Lambda_{xj}$.

$I - Z^{-1/2} \Lambda_{xi} Z^{-1/2} \leq 2(I-\Lambda_{xi}) + 4 \sum_{j \neq i} \Lambda_{xj}$

Given: $\zeta = \sum_m p(m)\, \zeta_m$.

(1) $Tr(\zeta_m \Pi) \geq 1-\varepsilon$, (2) $Tr(\zeta_m \Pi_m) \geq 1-\varepsilon$, (3) $Tr(\Pi_m) \leq d_1$, (4) $\Pi\zeta\Pi \leq \Pi/d_0$.

$X_1, \ldots, X_k$ iid $\sim p(m)$, $S' = \{\zeta_{xi}\}$, $k = \lfloor (d_0/d_1)\gamma \rfloor$, $0<\gamma<1$.

Claim: $E\ de(S') \leq 2[\varepsilon+\sqrt{(8\varepsilon)}]+4\gamma$.

Proof: Let $\Lambda_m = \Pi\ \Pi_m\ \Pi$, $Z = \sum_{i=1}^{k} \Lambda_{xi}$

Take $F_i = Z^{-1/2}\, \Lambda_{xi}\, Z^{-1/2}$ for the POVM elements.

$\sum_{i=1}^{k} F_i = Z^{-1/2} \sum_{i=1}^{k} \Lambda_{xi}\, Z^{-1/2} = Z^{-1/2}\, Z\, Z^{-1/2} = I_{supp(Z)} \leq I$.

Can add $F_{err} = I - I_{supp(Z)}$ to complete the POVM.

$de(S') \leq 1 - 1/k \sum_i Tr(\zeta_{xi} F_i) = 1/k \sum_i Tr[\zeta_{xi}(I-F_i)]$

$\leq 1/k \sum_i Tr[\zeta_{xi}\, (2(I-\Lambda_{xi})+4 \sum_{j \neq i} \Lambda_{xj})]$

$E\ de(S') \leq E\ Tr[\zeta_{x1}\, (2(I-\Lambda_{x1})+4 \sum_{j\geq 2} \Lambda_{xj})]$ symmetry due to E

$\leq 2\ [1-E\ Tr(\zeta_{x1}\Lambda_{x1})]+4 \sum_{j\geq 2} E\ Tr[\zeta_{x1}\Lambda_{xj}]$

Given: $\zeta = \sum_m p(m)\, \zeta_m$.

(1) $Tr(\zeta_m \Pi) \geq 1-\varepsilon$, (2) $Tr(\zeta_m \Pi_m) \geq 1-\varepsilon$, (3) $Tr(\Pi_m) \leq d_1$, (4) $\Pi\zeta\Pi \leq \Pi/d_0$.

$X_1, \ldots, X_k$ iid $\sim p(m)$, $S' = \{\zeta_{xi}\}$, $k = \lfloor (d_0/d_1)\gamma \rfloor$, $0<\gamma<1$.

Claim: $E\, de(S') \leq 2[\varepsilon + \sqrt{(8\varepsilon)}] + 4\gamma$.

$\Lambda_m = \Pi\, \Pi_m \Pi$

Proof:

$$E\, de(S') \leq \underbrace{2\,[1 - E\, Tr(\zeta_{x1} \Lambda_{x1})]}_{\text{1st term}} + 4 \sum_{j\geq 2} \underbrace{E\, Tr[\zeta_{x1} \Lambda_{xj}]}_{\text{2nd term}}$$

For the 1st term:

$$\boxed{\text{Gentle measurement lemma [Winter]: Let } \sigma \geq 0,\ tr(\sigma) \leq 1,\ 0 \leq Y^\dagger Y \leq I. \text{ If } Tr(\sigma\, Y^\dagger Y) \geq 1-\varepsilon, \text{ then } \|\, Y\sigma Y^\dagger - \sigma \,\|_1 \leq \sqrt{(8\varepsilon)}.}$$

$$\boxed{\text{By (1) } Tr(\zeta_m\, \Pi) \geq 1-\varepsilon, \text{ thus } \|\, \Pi\, \zeta_m\, \Pi - \zeta_m \,\|_1 \leq \sqrt{(8\varepsilon)}.}$$

Thus, $\forall\ 0 \leq P \leq I$, $|\, Tr[\, P\, (\Pi\zeta_m\Pi - \zeta_m)]\, | \leq \sqrt{(8\varepsilon)}$.

Taking $P=\Pi_m$, $|\, - Tr[\, \Pi_m\, \Pi\zeta_m\Pi] + Tr[\Pi_m\zeta_m]\, | \leq \sqrt{(8\varepsilon)}$.

$- Tr[\Lambda_m\zeta_m] \leq - Tr[\Pi_m\zeta_m] + \sqrt{(8\varepsilon)} \leq -1+\varepsilon+\sqrt{(8\varepsilon)}$ from (2)

Given: $\zeta = \sum_m p(m)\, \zeta_m$.

(1) $\mathrm{Tr}(\zeta_m \Pi) \geq 1-\varepsilon$, (2) $\mathrm{Tr}(\zeta_m \Pi_m) \geq 1-\varepsilon$, (3) $\mathrm{Tr}(\Pi_m) \leq d_1$, (4) $\Pi\zeta\Pi \leq \Pi/d_0$.

$X_1, \ldots, X_k$ iid $\sim p(m)$, $S' = \{\zeta_{x_i}\}$, $k = \lfloor (d_0/d_1)\gamma \rfloor$, $0<\gamma<1$.

Claim: $E\ \mathrm{de}(S') \leq 2[\varepsilon+\sqrt{(8\varepsilon)}]+4\gamma$.

$\Lambda_m = \Pi\, \Pi_m \Pi$

Proof:

$$E\ \mathrm{de}(S') \leq 2\,[1-\underbrace{E\,\mathrm{Tr}(\zeta_{x_1}\Lambda_{x_1})}_{\text{1st term}}] + 4 \sum_{j\geq 2} \underbrace{E\,\mathrm{Tr}[\zeta_{x_1}\Lambda_{x_j}]}_{\text{2nd term}}$$

For the 2nd term:

$$\begin{aligned}
E\,\mathrm{Tr}[\zeta_{x_1}\Lambda_{x_j}] &= E\,\mathrm{Tr}[\zeta_{x_1}\Pi\, \Pi_{x_j}\, \Pi] \\
&= \mathrm{Tr}[\,(E\zeta_{x_1})\,\Pi\,(E\,\Pi_{x_j})\,\Pi] \quad j\neq 1 \Rightarrow \text{independence} \\
&= \mathrm{Tr}[\quad \zeta \quad \Pi\,(E\,\Pi_{x_j})\,\Pi] \\
&= \mathrm{Tr}[\,\Pi\zeta\Pi\ \ (E\,\Pi_{x_j})] \\
&\leq \mathrm{Tr}[\,\Pi/d_0\ \ (E\,\Pi_{x_j})] \qquad \text{by (4)} \\
&= E\,\mathrm{Tr}[\,\Pi\,\Pi_{x_j}\,\Pi]/d_0 \leq d_1/d_0. \qquad \text{by (3)}
\end{aligned}$$

Given: $\zeta = \sum_m p(m)\, \zeta_m$.

(1) $\mathrm{Tr}(\zeta_m \Pi) \geq 1-\varepsilon$, (2) $\mathrm{Tr}(\zeta_m \Pi_m) \geq 1-\varepsilon$, (3) $\mathrm{Tr}(\Pi_m) \leq d_1$, (4) $\Pi\zeta\Pi \leq \Pi/d_0$.

$X_1, \ldots, X_k$ iid $\sim p(m)$, $S' = \{\zeta_{x_i}\}$, $k = \lfloor (d_0/d_1)\gamma \rfloor$, $0<\gamma<1$.

Claim: $E\, de(S') \leq 2[\varepsilon+\sqrt{(8\varepsilon)}]+4\gamma$.

$\Lambda_m = \Pi\, \Pi_m \Pi$

Proof:

$$E\, de(S') \leq 2\, [1 - \underbrace{E\, \mathrm{Tr}(\zeta_{x_1} \Lambda_{x_1})}_{\text{1st term}}] + 4 \sum_{j\geq 2} \underbrace{E\, \mathrm{Tr}[\zeta_{x_1} \Lambda_{x_j}]}_{\text{2nd term}}$$

For the 1st term: $-\, \mathrm{Tr}[\Lambda_m \zeta_m] \leq -1+\varepsilon+\sqrt{(8\varepsilon)}$

For the 2nd term: $E\, \mathrm{Tr}[\zeta_{x_1} \Lambda_{x_j}] \leq d_1/d_0$.

$$E\, de(S') \leq 2\, [\varepsilon+\sqrt{(8\varepsilon)}] + 4\, (k-1)\, d_1/d_0$$
$$\leq 2\, [\varepsilon+\sqrt{(8\varepsilon)}] + 4\, \gamma$$

Back to direct coding for HSW: Want to find D_n that distinguishes $\sigma^n_m = \sigma_{x^n_m} = \sigma_{x_{m1}} \otimes \sigma_{x_{m2}} \dots \sigma_{x_{mn}}$

Take $S = \{\zeta_m = \sigma^n_m\}$ in the packing lemma.

We saw earlier:

$\forall$ n, $\exists$ ε_n , $\delta_n \to 0$ s.t.:

condition # in packing lemma

add m back

1. $\mathrm{Tr}[\sigma_{x^n} \Pi_{x^n}] \geq 1-\delta_n$ $\qquad \to (2)$, $\delta_n \to \varepsilon$

2. $\mathrm{Tr}[\Pi_{x^n}] \leq 2^{n[\Sigma_x p(x) H(\sigma_x) + \varepsilon_n]}$ $\qquad \to (3)$, $2^{n[\Sigma_x p(x) H(\sigma_x) + \varepsilon_n]} \to d_1$

3. $\mathrm{Tr}[\sigma_{x^n} \Pi] \geq 1-\delta_n$ if x^n strongly typical $\qquad \to (1)$, $\delta_n \to \varepsilon$

4. $[\Pi\, \sigma^{\otimes n}\, \Pi] \leq 2^{-n[H(\sigma)-\varepsilon_n]}\, \Pi$ $\qquad \to (4)$, $2^{n[H(\sigma)-\varepsilon_n]} \to d_0$

---

(1) $\mathrm{Tr}(\zeta_m \Pi) \geq 1-\varepsilon$, (2) $\mathrm{Tr}(\zeta_m \Pi_m) \geq 1-\varepsilon$, (3) $\mathrm{Tr}(\Pi_m) \leq d_1$, (4) $\Pi \zeta \Pi \leq \Pi / d_0$.

Back to direct coding for HSW: Want to find D_n that distinguishes $\sigma^n{}_m = \sigma_{x^n{}_m} = \sigma_{x_{m1}} \otimes \sigma_{x_{m2}} \ldots \sigma_{x_{mn}}$

Take $S = \{\zeta_m = \sigma^n{}_m\}$ in the packing lemma.

By the packing lemma, take

$k = \gamma\, d_0/d_1 = \gamma\, 2^{n[I(X:B)_\tau - 2\varepsilon_n]}$

randomly drawn states, and average distinguishability error $\leq 2\,[\delta_n + \sqrt{(8\delta_n)}] + 4\,\gamma$.

Take $\gamma = 2^{-n\eta}$. Then, $\exists$ code with average error $\to 0$

Rate deficit $= \eta + 2\varepsilon_n \to 0$.

Remove worst half of the codewords to make worse case error $\to 0$.

condition # in packing lemma

$\to (2)$

$\delta_n \to \varepsilon$

$\to (3)$

$2^{n[\Sigma_x p(x) H(\sigma_x) + \varepsilon_n]} \to d_1$

$\to (1)$

$\delta_n \to \varepsilon$

$\to (4)$

$2^{n[H(\sigma) - \varepsilon_n]} \to d_0$

$\tau = \Sigma_x\, p(x) |x\rangle\langle x| \otimes \sigma_x$

$I(X:B)_\tau = H(\sigma) - \Sigma_x p(x) H(\sigma_x)$

Part 2: Converse [If $P_e^{ave} \to 0$, then achievable rate $R \leq C$.]

~~Lemma: Let $Y^n = N^{\otimes n}(X^n)$, and C be the capacity of N.~~
~~Then, $I(X^n:Y^n) \leq nC$.~~ not for quantum channels due to additivity issue

Thm [Fanos ineq]:

$$H(P_e) + P_e \log(|\Omega|-1) \geq H(X|Y)$$

H(MY)-H(Y)

Proof of converse:

-H(MY)+H(Y)+H(M)

$$nR = H(M) = H(M|Y) + I(M:Y)$$

$$\leq H(M|Y) + \boxed{I(E_n(M):Y) \quad \text{data processing ineq}}$$

$$\leq 1 + P_e\, nR + nC$$

no need

Fanos ineq
$M \leftrightarrow X$
$2^{nR} \leftrightarrow |\Omega|$

by Holevo's bound
& definition of C(N).