Recall:

Def[typical sequence]:

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

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

Def[Jointly typical sequence]:

$x^n y^n$ ε-jointly-typical if

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

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

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

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

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

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. $\forall\, n \ge n_0$, $\Pr(X^nY^n \in A_{n,\varepsilon}) > 1-\delta$
2. $(1-\delta)\, 2^{n\,[H(XY)-\varepsilon]} \le |A_{n,\varepsilon}| \le 2^{n\,[H(XY)+\varepsilon]}$
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, $\Pr_q(W^n\, Z^n \in A_{n,\varepsilon}) \le 2^{-n[I(X:Y)-3\varepsilon]}$
Also, for large n,
$(1-\delta)\, 2^{-n[I(X:Y)+3\varepsilon]} \le \Pr_q(W^n\, Z^n \in A_{n,\varepsilon})$

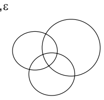
<u>Joint asymptotic equipartition (Joint AEP) theorem:</u>

Proof:

[1] Given ε, δ, we can apply AEP on X^n, Y^n, and $(XY)^n$.
thus, $\exists n_0$ s.t. $\forall\, n \ge n_0$,
the ε-typical sets $T^X_{n,\varepsilon}$, $T^Y_{n,\varepsilon}$, $T^{XY}_{n,\varepsilon}$
all have prob $\ge 1-\delta/3$.

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

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



By the union bound,

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

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

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

Proof:

[2] Using the same proof as in AEP, condition (c) implies
$\forall\, x^ny^n \in A_{n,\varepsilon}$,
$(1-\delta)\, 2^{-n(H(XY)+\varepsilon)} \le p(x^ny^n) \le 2^{-n(H(XY)-\varepsilon)}$

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

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 $p(y^n)$

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

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

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

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}.

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

↑ since $x^ny^n \in A_{n,\varepsilon}$,

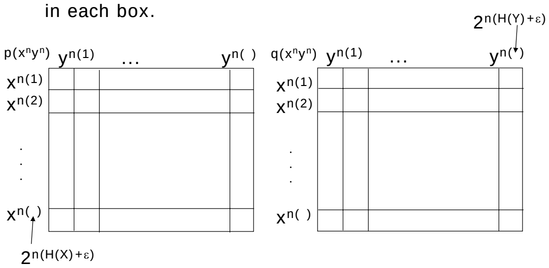
$1 = \sum_{x^n \in S} p(x^n|y^n) \approx |S_{y^n}|\, 2^{-n[H(X|Y)]}$

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 $y^n \approx 2^{-nI(X:Y)}$.

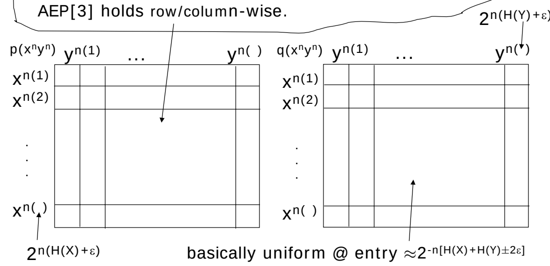
What's going on?

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.

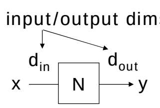


What's going on?

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.

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

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

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.

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

N N N

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 — Err_0
- or $\exists M' \neq M$ with $E_n(M') y^n \in A_{\varepsilon,n}$ — $Err_{M'}$

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

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

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 $\Pr_{\mathcal{C}_n}(Err_0 | M=1)$:

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, $\exists\, \delta_n$, ε_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 $\Pr(x_{M=1}\, y^n \in A_{n,\varepsilon_n}) > 1-\delta_n$.
$\Pr_{\mathcal{C}_n}(Err_0 | M=1) \leq \delta_n$.

BACK 1 SLIDE.

---

for 1 M'

$N^{\otimes n}(x_1)$

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

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, $\exists\, \delta_n$, ε_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

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

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 $\mathcal{C}'_n$.

$P_e{}^{ave}(\mathcal{C}_n) = 1/k \sum_M \lambda_M(\mathcal{C}_n)$

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

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

Thus, $\mathcal{C}'_n$ has worse case error prob $\leq \zeta_n{}^{ave}/2 =: \zeta_n \to 0$.
[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 $\geq \max_{p(x)} I(X:Y)$.