Theorem [Shannon’s noisy coding theorem]
\[ C(N) = \max_{p(x)} I(X:Y) \]

How to prove this?
1. Direct coding – consider codes that are promising
A clever code doesn’t come by easily.
Instead, consider “random” (M,n) codes with
rate = I(X:Y) and show \( \text{Prob}(E_{P_e} \to 0) > 0 \).
Thus \( \exists \) code with small \( E_{P_e} \) (our 2nd encounter
with “existential proofs”). Extract a subcode with
similar rate but \( P_e \to 0 \).
2. Converse – show that if at higher rates, \( E_{P_e} \to 0 \).
Plan: 2, then heuristic 1, then 1.

Proof of converse:
\[ nR = H(M) = H(M|Y^n) + I(M:Y^n) \leq H(M|Y^n) + I(E_n(M):Y^n) \]
(i) data processing ineq
(ii) by lemma \( \leq n \max_{p(x)} I(X:Y) \)

(i) Data processing inequality \( I(E:F) \geq I(E:G) \)
if \( E \to F \to G \) is a Markov Chain
(i.e. \( I(E:G|F) = 0 \))
Proof:
\[ I(E:F) = H(E) + H(F|E) - H(EF) \]
\[ = I(E:G) + H(G|E) - H(EF) - H(E|G) + H(F|E) \]
\[ = I(E:G) + I(E:F|G) \]
but the LHS is symmetric wrt exchange F and G,
so must the RHS.
\[ \text{So, } I(E:G) + I(E:F|G) = I(E:F) + I(E:G|F) \geq 0 \]
So, \( I(E:G) \geq I(E:F) \).

(iii) Lemma: Let \( Y^n = N^{\otimes n}(X^n) \).
Then, \( I(X^n:Y^n) \leq \sum_{i=1}^n I(X_i;Y_i) \).
Pf:
\[ I(X^n:Y^n) = H(Y^n) - H(Y^n|X^n) \]
\[ = H(Y^n) - \sum_{i=1}^n I(Y_i;Y_1,...Y_{i-1}X^n) \text{ Chain rule} \]
\[ = H(Y^n) - \sum_{i=1}^n h(Y_i|X^n) \]
\[ \leq \sum_{i=1}^n h(Y_i|X) \text{ Subadditivity} \]
\[ \leq \sum_{i=1}^n I(X_i;Y_i) \]

1. Direct coding:
Let \( M = 2^{\Omega(X|Y)^{\otimes n}} \). What’s the \( (M,n) \) code?
Fix any \( p(x) \).
Encoder \( c_i \):
Pick \( n \) codewords \( c_i = x_i \ldots x_n \),
each \( x_i \) chosen iid \( \sim p(x) \).
Fixed & known to Alice & Bob once chosen.
\( c_1 = x_{11}, x_{12}, \ldots, x_{1n} \)
\( c_2 = x_{21}, x_{22}, \ldots, x_{2n} \)
\( \ldots \)
\( c_M = x_{M1}, x_{M2}, \ldots, x_{Mn} \)
Everything refers to this particular code \( c_i \) from now on.
1. Direct coding: \( c_i = x_{1i}, x_{2i}, \ldots, x_{ni} \)
   \( i = 1, 2, \ldots, M \)

   Heuristically why \( P_e \rightarrow 0 \):
   The \( n \) channel outputs \( Y \) is iid with \( p(y) = \sum p(y|x) p(x) \)
   With high prob, output typical \( y_1 \) \( \approx 2^n H(Y) \) of them.

   \( x_{ij} \) chosen iid \( \sim p(x) \)

   For each \( c_i \) sent via \( N \otimes n \), there're \( \approx 2^{nH(Y|X)} \) possible outcomes (call the set \( O_i \) centered around \( c_i \).)

   Since the \( c_i \)'s are random, if \( 2^{nH(Y|X)} M < 2^{nH(Y)} \), these \( O_i \)'s don't overlap much. So, decoder just output "which sphere" contains the output \( y_1 \) \( \ldots y_n \).

Recall:

**Def[typical sequence]:**

- \( x^n \) \( \varepsilon \)-typical if \( \left| -\frac{1}{n} \log(p(x^n)) - H(X) \right| \leq \varepsilon \)

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

**Def[Jointly typical sequence]:**

- \( x^n y^n \) \( \varepsilon \)-jointly-typical if
  \[ \left| -\frac{1}{n} \log(p(x^n y^n)) - H(XY) \right| \leq \varepsilon \]

- where \( p(x^n y^n) = p(x^n) p(y^n) \).

- Need also: (a) \( -\frac{1}{n} \log(p(x^n)) - H(X) \leq \varepsilon \)

- (b) \( -\frac{1}{n} \log(p(y^n)) - H(Y) \leq \varepsilon \)

- (The strong typicality has (a) \( \Rightarrow \) (b) but not for entropic typicality.)

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

More observations:

Given \( y^n \in T_{n, \varepsilon}^n \), how many \( x^n \in T_{n, \varepsilon}^n \) s.t. \( x^n y^n \in A_{n, \varepsilon} \)?

Call this set \( S(y^n) \).

(1) \( p(x^n y^n) = p(x^n y^n) / p(y^n) \approx 2^{-n(H(X)+H(Y))} = 2^{-nH(XY)} \)

(2) \( 1 = \sum_{x^n y^n} p(x^n y^n) \approx |S(y^n)| \cdot 2^{-nH(XY)} \)

Hence, \( |S(y^n)| \approx 2^{nH(XY)} \). Fraction of such \( x^n \approx 2^{nH(XY)} \).

Similarly, given \( x^n \in T_{n, \varepsilon}^n \), \( x^n y^n \)'s are jointly typical with it, and the fraction of such \( y^n \approx 2^{nH(Y)} \).

Joint asymptotic equipartition (Joint AEP) theorem:

Let \( (X^n, Y^n) \) be sequences of length \( n \) drawn iid according to \( p(x^n, y^n) = \prod p(x^n) p(y^n) \).

Then:

1. \( \Pr(X^n, Y^n \in A_{n, \varepsilon}) \rightarrow 1 \)

2. \( |A_{n, \varepsilon}| \approx 2^{nH(XY)} \)

3. if we draw \( X^n \) & \( Y^n \) according to \( q(x^n, y^n) = p(x^n) p(y^n) \).

   \( \Pr_q (\text{outcome } \in A_{n, \varepsilon}) \approx 2^{-nI(X;Y)} \)

Proof (with \( \varepsilon, \delta \)) available in the 18 page notes.

Make a table of typical \( x^n \)'s and \( y^n \)'s, and for jointly typical \( x^n y^n \), put a 1, else, put a 0.

| \( x^n(1) \) | \ldots | \( x^n(n) \) | \( y^n(1) \) | \ldots | \( y^n(n) \) |
|---|---|---|---|---|
| \( \approx 2^{nH(X)} \) entries each row has \( \approx 2^{nH(Y)} \) 1's total \( \approx 2^{nH(XY)} \) 1's |
| \( \approx 2^{nH(Y)} \) entries |
| Our random code corresponds to \( M \) randomly chosen rows. |
**Dₙ:** typical set decoding

Given \( y^n \), if there is a unique \( x^n \in S(y^n) \), output \( m' \) s.t. \( c_{m'} = x^n \).
Else, output \( W=M+1 \) (error symbol).

How will this fail for message \( m \)?

Either - no such \( x^n \)
- or \( \exists m' \neq m \) with \( c_{m'} y^n \in A_{\epsilon, n} \)

For the random code \( C_n \), let the average error over all messages be \( \text{EP}_{\epsilon}(C_n) \), same as error if \( m=1 \) (since all messages similar).

\[
\text{EP}_{\epsilon}(C_n) = \text{Pr}_{\epsilon}(W=1|m=1) = \sum_{\text{Err}_m} \text{Pr}_{\epsilon}(\text{Err}_m|m=1) \leq M \sum_{\text{Err}_m} \text{Pr}_{\epsilon}(\text{Err}_m|m=1) \approx 2^{-nI(X:Y)}.
\]

Bounding \( \text{Pr}_{\epsilon}(\text{Err}_2|m=1) = \text{Pr}_{\epsilon}(c_2 y^n \in A_{\epsilon,n}) \):

But \( c_2 \) and \( y^n = N^{\epsilon^n}(x_1) \) independent.

By joint AEP [3], \( \text{Pr}_{\epsilon}(c_2 y^n \in A_{\epsilon,n}) \approx 2^{-n(I(X;Y)} \)

If \( M = 2^{n(I(X;Y) - \delta_n) \text{ and } n\delta_n \text{ growing with } n \text{ but } \delta_n \to 0} \)

Note: \( P_s(m=1) + P_s(m=2) + \ldots + P_s(m=M) = M \text{EP}_{\epsilon}(C_n) \)

Reorder \( m \)'s so that \( P_s(m) \) is increasing.

So, keeping only codewords for \( m=1, \ldots, M/2 \), worse case prob error \( \leq \text{EP}_{\epsilon}(C_n)/2 \).