Joint Typicality
References:

Cover & Thomas, Chapter 8
From entropy (typicality) to correlations (Joint Typicality)

Consider 2 random variables $X$, $Y$.

Def [Jointly typical sequence]

Given a distribution $p(x,y)$, drawn iid $n$ times

$x^n y^n$ is $\delta$-jointly typical if:

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

b) $\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| \leq \delta$ (\(y^n\) typical)

c) $\left| -\frac{1}{n} \log p(x^n y^n) - H(X,Y) \right| \leq \delta$ (\(xy^n\) typical)

Def [Jointly typical set]

\[ A_{n,\delta} = \left\{ x^n y^n \in \mathcal{S}_x^n \times \mathcal{S}_y^n : x^n y^n \text{ jointly typical} \right\} \]

Will see: under $n$ iid draws of $p(xy)$, the jointly typical set occurs with prob $\to 1$, so we can safely use its properties ...
Recall the chain rule

\[ H(XY) = H(Y) + H(X|Y). \]

Obs: if \( x^n y^n \in \text{An} \), then

\[ p(x^n y^n) = \frac{p(x^n y^n)}{p(y^n)} \leq 2^{\frac{-n(H(x^n y^n) - \delta)}{2 - n(H(y^n) + \delta)}} = 2. \]
For each such $y^n$, the above two conditions say that there are $2^{-nH(X|Y)}$ $\tilde{x}^n$'s such that $\tilde{x}^n y^n \in A_{n,i}$.

Proof: similar to the proof for the AEP.

This observation is crucial for the Joint AEP ...
The 2 parts above are similar to the AEP.

Note that the jointly typical set is defined wrt $p(xy)$ drawn $n$ times iid. Once defined, it is just a set. Now we ask questions about the set concerning OTHER distributions.
The 2 parts above are similar to the AEP.

One quick intuition to see this: there are

\[ 2^{nH(X)} \] typical \( x^n \)'s and \( 2^{nH(Y)} \) typical \( y^n \)'s

If they're chosen indep, there are

\[ 2^{nH(X)} 2^{nH(Y)} \] typical, equiprobable \( x^n y^n \)'s

but there are only \( 2^{nH(XY)} \) elements in the jointly typical set.
The 2 parts above are similar to the AEP.

The third part captures the "jointness" of X and Y
**if** x^n y^n are not from the joint distribution for XY,
but x^n y^n are independent, x^n y^n unlikely in the jointly typical set.

the "unlikelihood" is exp in I(X:Y)

Proof (see Cover & Thomas), based on AEP and union bound etc.
We can summarize joint typicality and JAEP by a matrix:

\[
\begin{array}{ccc}
\text{typical } x^n & \text{typical } y^n \\
\uparrow & \downarrow & \text{Total (2x2 entries)} \\
\text{Typical } y^n & \text{typical } x^n & \text{if } x^n y^n \in \text{Anis} \\
& \text{0 otherwise} \\
\end{array}
\]

1. \( \Pr (\tilde{\text{Anis}}) > 1 - \varepsilon \)
2. \( (1 - \varepsilon) 2^n (\lambda (xy) - 2) \leq 1 \text{Anis} \leq 2^n [\lambda (xy) + 1] \)

Say that tolerating a probability of failure \( \varepsilon \), we can focus on this table.

Say, here are \( \approx 2^n \lambda (xy) \) \( \tilde{\text{Anis}} \)s.
We can summarize joint typicality and JAEP by a matrix:

\[ n H(x) \quad n H(y) \]

Total \( 2 \times 2 \) entries

The entry associated with \( x^n y^n \) is

\[ \begin{cases} 1 & \text{if } x^n y^n \in \text{Anid} \\ 0 & \text{otherwise} \end{cases} \]

Observations:

- If \( x^n y^n \notin \text{Anid} \),
  
  \( p(x^n y^n) = \frac{p(x^n y^n)}{p(y^n)} \leq \frac{-n(H(xy) - d)}{2 - n(H(y) + d)} \leq 2 \)

- If \( x^n y^n \in \text{Anid} \),
  
  \( p(x^n y^n) = \frac{p(x^n y^n)}{p(y^n)} \geq \frac{2 - n(H(x) + d)}{2 - n(H(y) - d)} \geq 2 \)

Obs says each row has approximately \( 2^n H(x y^n) \) entries.
We can summarize joint typicality and JAEP by a matrix:

\[
\begin{array}{c}
\text{Typical } x^n \\
\text{Typical } y^n
\end{array}
\]

The entry associated with \(x^n y^n\) is:

\[
\begin{cases}
1 & \text{if } x^n y^n \in \text{Anis} \\
0 & \text{otherwise}
\end{cases}
\]

We say that tolerating a probability \(\varepsilon\), we can focus on this table.

Obs: if \(x^n y^n \in \text{Anis}\),

\[
\begin{align*}
p(x^n y^n) &= \frac{p(x^n y^n)}{p(y^n)} \\
p(x^n y^n) &= \frac{p(x^n y^n)}{p(y^n)}
\end{align*}
\]

\[
\begin{align*}
\leq \frac{-n(H(x y)-d)}{2^{-n(H(y)+d)}} & \leq 2 \\
\geq \frac{-n(H(x y)+d)}{2^{-n(H(y)+d)}} & \geq 2
\end{align*}
\]

Obs say each column has \(2^{-n(H(x|x))-d}\) '1's

Obs say each row has \(2^{-n(H(y|x))}\) '1's

XY symmetric here ...
We can summarize joint typicality and JAEP by a matrix:

$$P(x^n | y^n) = \frac{1}{n} \left( \frac{1}{2^n} \right)^n$$

1. Say that tolerating a prob of failure $\varepsilon$, we can focus on the table.
2. Say there are $2^{nH(x)}$ "1"'s in each row.
3. Suppose $x^n y^n$ is drawn according to the following distribution:

$$P(x^n y^n) = P(x^n) \cdot P(y^n)$$

Then

$$2^{-n(\delta(x:y) + 3\delta)} \leq \frac{\Pr (x^n y^n \in \mathcal{A}_{n, \delta})}{2^{-n(\delta(x:y) - 3\delta)}} \leq 2$$

4. Suppose a random entry in the table has prob $2^{-n\lambda}$ to be "1".
We can summarize joint typicality and JAEP by a matrix:

\[ \begin{bmatrix} nH(x) & nH(y) \\ \end{bmatrix} \]

Total \(= 2 \times 2\) entries

The entry associated with \(x^n y^n\) is:

\[ \begin{cases} 1 & \text{if } x^n y^n \in A^n \times B^n \\ 0 & \text{otherwise} \end{cases} \]

1. Says that tolerating a prob of failure \(\epsilon\), we can focus on this table.

2. Says there are \(2^{-nH(x)} + 2^{-nH(y)}\) "1"s.

3. Says each column has \(2^{-nH(x)} + 2^{-nH(y)}\) "1"s.

4. Says each row has \(2^{-nH(x)} + 2^{-nH(y)}\) "1"s.

5. Says a random entry in the table has prob \(2^{-nI(x:y)}\) to be "1".
Joint typicality gives the most critical tool to analyse classical comm through classical noisy channels. Here's an example of a joint distribution that will be relevant for this aim.

Example:

\[
\begin{align*}
R_X &= \{0, 1\} \\
R_Y &= \{0, 1\} \\
p(00) &= \frac{1}{2}(1-e) \\
p(01) &= \frac{1}{2}e \\
p(10) &= \frac{1}{2}e \\
p(11) &= \frac{1}{2}(1-e) \\
\end{align*}
\]

NB. \(e = \text{probability for } XY \text{ to disagree, say, } e = 0.1\)

\[
\begin{align*}
H(XY) &= 2 \left( \frac{1}{2} \right) (1-e) \log(\frac{1}{2}(1-e)) + 2 \left( \frac{e}{2} \right) \log(\frac{e}{2}) \\
&= 1 + h(e) = 1.469
\end{align*}
\]

where \(h(a) = -a \log a - (1-a) \log(1-a)\) binary entropy function

From JAEP (1):

\[
\begin{align*}
2^{n \cdot 1.469} \text{ jointly typical } X^n Y^n \text{'s}
\end{align*}
\]

For any \(e\), \(H(X) = H(Y) = 1\)

For \(e = 0.1\),
\[
\begin{align*}
H(X|Y) &= H(XY) - H(Y) \\
&= 0.469
\end{align*}
\]

For \(e = 0.1\),
\[
\begin{align*}
I(X:Y) &= H(X) - H(X|Y) \\
&= 0.531
\end{align*}
\]