

Definition: Joint Typicality

Very roughly, the forward direction of Shannon's noisy-channel coding theorem follows from the **asymptotic equipartition property (AEP)**. Increasing n (the size of the input blocks) results in output blocks that are more likely to be typical. But can we be sure that no two typical input sequences result in the same typical output sequence? Such a collision would make it impossible to decode correctly. To show that this does not form a problem, we need stronger mathematical tools.

Before you start on this section, it is a good idea to refresh your understanding of **typical sets** and the **asymptotic equipartition property**.

Definition: Joint typicality

For a joint distribution P_{XY} , the jointly typical set $A_\epsilon^{(n)}$ is defined as

$$A_\epsilon^{(n)} := \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \mid \begin{aligned} &\left| -\frac{1}{n} \log P_{X^n}(x^n) - H(X) \right| < \epsilon, \\ &\left| -\frac{1}{n} \log P_{Y^n}(y^n) - H(Y) \right| < \epsilon, \\ &\left| -\frac{1}{n} \log P_{X^n Y^n}(x^n, y^n) - H(XY) \right| < \epsilon \}. \end{aligned}$$

Here, $P_{X^n Y^n} := \prod_{i=1}^n P_{XY}(x_i y_i)$.

A jointly typical set is a set where both elements of the pair are typical individually under the marginal distributions, and their combination is typical under the joint distribution. To get some intuition for what a typical set looks like, we give two examples: one where X and Y are very correlated, and one where they are independent.

Example

Consider the almost fully correlated random variables X and Y with the following distribution:

| | $X = 0$ | $X = 1$ |
|---------|---------|---------|
| $Y = 0$ | 0.45 | 0.05 |
| $Y = 1$ | 0.05 | 0.45 |

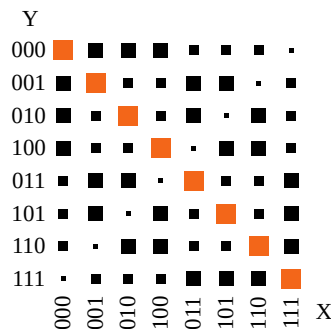
Note that the marginal distributions of X and Y are uniform, and so for all n and all $x^n \in \{0, 1\}^n$,

$$-\frac{1}{n} \log P_{X^n}(x^n) - H(X) = -\frac{1}{n} \log \frac{1}{2^n} - \log 2 = 0,$$

Information Theory | Definition: Joint Typicality

and a similar statement holds for Y . Hence, the first and second inequalities in the definition of joint typicality are always satisfied for any $\epsilon > 0$.

What about the third inequality? The table below visualizes the probabilities of every pair for $n = 3$ and $\epsilon = 0.35$: every square represents a pair of elements from \mathcal{X}^3 and \mathcal{Y}^3 . Bigger squares represent higher probabilities. The typical pairs are marked in orange.



As you might expect, pairs (x^3, y^3) with a higher overlap in bits have a higher probability of occurring. The jointly typical set has a very clear structure: only those pairs with $x^3 = y^3$ are included. Increasing ϵ would result in a bigger typical set. For example, for $\epsilon = 0.8$, all pairs that differ in at most one position (e.g., 000, 001) are part of the jointly typical set. Increasing n would have the same effect: for example, for $n = 10$ and $\epsilon = 0.35$, all pairs of sequences that differ at 0, 1, or 2 positions are part of the jointly typical set.

Similarly to what we saw earlier in the course, in this example, for big n and small ϵ , the jointly typical set contains only those pairs of sequences that differ in roughly 10 percent of the positions. The highest-probability pairs are excluded from the typical set. This might match your expectation, since the probability of sampling a non-matching pair of bits ((0,1) or (1,0)) is 10 percent: pairs with that percentage of non-matching bits are 'typical'. These jointly typical pairs together make up almost all of the probability mass, even though individually they are not the highest-probability pairs under P_{XY} .

Example

Consider the following distribution on X and Y :

| | $X = 0$ | $X = 1$ |
|---------|---------|---------|
| $Y = 0$ | 1/8 | 1/8 |
| $Y = 1$ | 3/8 | 3/8 |

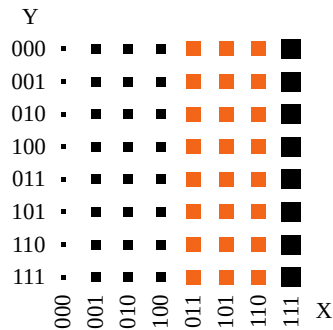
Note that X is uniform, and that X and Y are independent. Just like in the previous example, since X is uniform, the first inequality in the definition of joint typicality is always satisfied for any $\epsilon > 0$. Furthermore, since X and Y are independent, the third inequality is satisfied whenever the second inequality is, because

$$\begin{aligned}
 -\frac{1}{n} \log P_{X^n Y^n}(x^n, y^n) - H(XY) &= -\frac{1}{n} \log P_{X^n}(x^n) - \frac{1}{n} \log P_{Y^n}(y^n) - H(X) - H(Y) \\
 &= -\frac{1}{n} \log P_{Y^n}(y^n) - H(Y).
 \end{aligned}$$

Information Theory | Definition: Joint Typicality

Hence, for this distribution, any pair where the second element is typical for Y is also jointly typical.

More concretely, consider $n = 3$ and $\epsilon = 0.35$. For these values, the sequences 011, 101, and 110 are typical for Y . This results in the following jointly typical set (again, bigger squares represent bigger probabilities, and orange squares indicate elements of the typical set, **typo: the label of the axes label be switched around!**):



Note that $P_Y(111) \approx 0.422$, which is higher than $P_Y(011) \approx 0.141$. However, 111 is not typical for Y while 011 is. 111's probability deviates too far from the 'benchmark' set by $2^{-3H(Y)} \approx 0.185$. Together, the sequences in the typical set have a probability of about 0.422 of occurring. As n increases, this probability will grow, and the total probability of the elements that fall outside of the typical set (such as the all-1 string) will diminish.