Joint Asymptotic Equipartition Property (Joint AEP)

In the examples on the previous page, we have gathered some intuition about jointly typical sets. A (jointly) typical set is usually relatively small compared to the total sample space, but for large n, much of the probability lies on its elements.

We also see the effect of mutual information between X and Y. If X and Y are close to independent, the typical set has a rectangular shape: independently selecting a typical X sample and a typical Y sample will certainly yield a jointly typical element. On the other hand, if X and Y are highly dependent on each other, the typical set has a diagonal structure: there is a significant probability that sampling a typical X and a typical Y independently yields a pair that is not jointly typical. These observations lead us to a joint version of the asymptotic equipartition property, describing some properties of jointly typical sets:

Theorem: Joint Asymptotic Equipartition Property (Joint AEP)

Let $X^n Y^n$ be i.i.d. with respect to P_{XY}. Then:

1. $P_{X^n Y^n}(A_\epsilon^{(n)}) \xrightarrow{n \to \infty} 1$.
2. $|A_\epsilon^{(n)}| \leq 2^{n(H(XY)+\epsilon)}$.
3. For random variables $\tilde{X}$ and $\tilde{Y}$, if $\tilde{X}^n \tilde{Y}^n$ is i.i.d. with respect to $P_X \cdot P_Y$, then

$$P[(\tilde{X}^n \tilde{Y}^n) \in A_\epsilon^{(n)}] \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

4. For random variables $\tilde{X}$ and $\tilde{Y}$, if $\tilde{X}^n \tilde{Y}^n$ is i.i.d. with respect to $P_X \cdot P_Y$, then for large enough n,

$$P[(\tilde{X}^n \tilde{Y}^n) \in A_\epsilon^{(n)}] \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}.$$

Proof

We show the statements separately, for an arbitrary $X^n Y^n$ that is i.i.d. with respect to P_{XY}.

1. The first statement follows from the weak law of large numbers. Similarly to the proof of the asymptotic equipartition property, we have that

$$-\frac{1}{n}\log P_{X^n}(X^n) = -\frac{1}{n}\sum_{i=1}^{n}\log P_X(X)$$

$$\xrightarrow{p} -\mathbb{E}[\log P_X(X)]$$

$$= H(X).$$

Writing out the definition of convergence in probability, we have that for all $\epsilon > 0$, there exists an $n_1 \in \mathbb{N}$ such that for all $n > n_1$,

$$P\left[\left|-\frac{1}{n}\log P_{X^n}(X^n) - H(X)\right| \geq \epsilon\right] \leq P\left[\left|-\frac{1}{n}\log P_{X^n}(X^n) - H(X)\right| \geq \frac{\epsilon}{3}\right]$$

$$< \frac{\epsilon}{3}.$$

In the above, we can divide ϵ by 3 because the second inequality holds for *all* $\epsilon > 0$, so in particular also for $\frac{\epsilon}{3}$. We can derive statements about Y^n and $X^n Y^n$ that are similar to this bound above. Doing so, we find $n_2, n_3 \in \mathbb{N}$ for Y^n and $X^n Y^n$, respectively. For all $\epsilon > 0$, we can now choose $n_0 := \max\{n_1, n_2, n_3\}$, and find that for all $n > n_0$,

$$P\left[(X^n, Y^n) \in A_\epsilon^{(n)}\right] = 1 - P\left[(X^n, Y^n) \notin A_\epsilon^{(n)}\right]$$

$$\geq 1 - \left(P\left[\left|-\frac{1}{n}\log P_{X^n}(X^n) - H(X)\right| \geq \epsilon\right] + \right.$$

$$P\left[\left|-\frac{1}{n}\log P_{Y^n}(Y^n) - H(Y)\right| \geq \epsilon\right] +$$

$$\left. P\left[\left|-\frac{1}{n}\log P_{X^n Y^n}(X^n Y^n) - H(XY)\right| \geq \epsilon\right]\right)$$

$$\leq 1 - \left(\frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}\right)$$

$$= 1 - \epsilon.$$

The first inequality is due to the union bound, and the second follows from the inequality in the proof of the previous point, and its analogues. By definition of convergence, the first statement is proven.

2. Observe that by rewriting the last line in the definition of $A_\epsilon^{(n)}$ in the definition of joint typicality, one can derive that for every element $(x^n, y^n) \in A_\epsilon^{(n)}$,

$$P_{X^n Y^n}(x^n y^n) > 2^{-n(H(XY)+\epsilon)}.$$

The second statement follows by rearranging the following inequality:

$$1 = \sum_{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n} P_{X^n Y^n}(x^n, y^n)$$

$$\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n Y^n}(x^n, y^n)$$

$$\geq |A_\epsilon^{(n)}| \cdot 2^{-n(H(XY)+\epsilon)}.$$

3. First, we write out the probability of sampling an element in the typical set, again by rearranging the inequalities in the definition of joint typicality.

$$P[(\tilde{X}^n \tilde{Y}^n) \in A_\epsilon^{(n)}] = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{\tilde{X}^n \tilde{Y}^n}(x^n, y^n)$$

$$= \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n}(x^n) \cdot P_{Y^n}(y^n)$$

$$\leq |A_\epsilon^{(n)}| \cdot 2^{-n(H(X)-\epsilon)} \cdot 2^{-n(H(Y)-\epsilon)}.$$

Combining this with part (2) of this theorem, we get

$$P[(\tilde{X}^n \tilde{Y}^n) \in A_\epsilon^{(n)}] \leq 2^{n(H(XY)+\epsilon)} \cdot 2^{-n(H(X)-\epsilon)} \cdot 2^{-n(H(Y)-\epsilon)}$$

$$= 2^{-n(-H(XY)+H(X)+H(Y)-3\epsilon)}$$

$$= 2^{-n(I(X;Y)-3\epsilon)}.$$

This completes the proof of the third part.

4. By part (1), we have that $P[A_\epsilon^{(n)}] \geq 1 - \epsilon$ for large enough n. Thus,

$$1 - \epsilon \leq P[A_\epsilon^{(n)}]$$

$$= \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n Y^n}(x^n, y^n)$$

$$\leq |A_\epsilon^{(n)}| 2^{-n(H(XY)-\epsilon)}.$$

Rearranging this inequality and using the same type of derivation as in the proof of (3), it follows that

$$P[(\tilde{X}^n \tilde{Y}^n) \in A_\epsilon^{(n)}] = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n}(x^n) \cdot P_{Y^n}(y^n)$$

$$\geq (1 - \epsilon) \cdot 2^{n(H(XY)-\epsilon)} \cdot 2^{-n(H(X)+\epsilon)} \cdot 2^{-n(H(Y)+\epsilon)}$$

$$= (1 - \epsilon) \cdot 2^{-n(I(X;Y)+3\epsilon)}.$$