Some Remarks on Joint Typicality

Let $\mathcal{X}$ and $\mathcal{Y}$ denote discrete and finite sets and let $P_{XY}$ be a probability mass function (pmf) on $\mathcal{X} \times \mathcal{Y}$. Let $x^n = (x_1, \ldots, x_n)$ denote an $n$-length sequence of symbols from $\mathcal{X}$, i.e., $x^n \in \mathcal{X}^n$. Similarly, let $y^n$ denote elements of $\mathcal{Y}^n$.

Let $(X^n,Y^n) = ((X_1,Y_1),\ldots,(X_n,Y_n))$ denote an $n$-length sequence of random variables drawn according to the i.i.d. (product) measure on $\mathcal{X}^n \times \mathcal{Y}^n$ obtained from $P_{XY}$, i.e.,

$$\text{Prob}(X^n = x^n, \ Y^n = y^n) = P_{X^nY^n}(x^n,y^n) = \prod_{i=1}^{n} P_{XY}(x_i,y_i).$$

Next, let

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{XY}(x,y) \quad \text{and} \quad P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x,y)$$

denote the marginal pmfs induced by $P_{XY}$ on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $P_{X^n}$ denote the i.i.d. (product) measure on $\mathcal{X}^n$ obtained from $P_X$, and similarly $P_{Y^n}$ on $\mathcal{Y}^n$ from $P_Y$. It is clear that $P_{X^n}$ and $P_{Y^n}$ are also the marginal pmfs induced on $\mathcal{X}^n$ and $\mathcal{Y}^n$ respectively by $P_{X^nY^n}$.

Finally, let $\tilde{X}^n = (\tilde{X}_1,\ldots,\tilde{X}_n)$ be an $n$-length sequence of i.i.d. random variable drawn according to the common pmf $P_X$. Let $\tilde{Y}^n = (\tilde{Y}_1,\ldots,\tilde{Y}_n)$ be another $n$-length sequence of i.i.d. random variables drawn according to $P_Y$ and independently of $\tilde{X}^n$. Let

$$\text{Prob}(\tilde{X}^n = x^n, \ \tilde{Y}^n = y^n) = P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n) = P_{X^n}(x^n)P_{Y^n}(y^n) = \prod_{i=1}^{n} P_X(x_i) \prod_{i=1}^{n} P_Y(y_i)$$

denote the joint pmf of $(\tilde{X}^n, \tilde{Y}^n)$. Note that in general

$$\text{Prob}(\tilde{X}^n = x^n, \ \tilde{Y}^n = y^n) = P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n) \neq P_{X^nY^n}(x^n,y^n) = \text{Prob}(X^n = x^n, \ Y^n = y^n),$$

unless, trivially, $X$ and $Y$ are independent under $P_{XY}$.

**Definition:** The set of *jointly $\epsilon$-typical sequences* for the pmf $P_{XY}$ is given by

$$A_{\epsilon}^{(n)} = \left\{(x^n,y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| -\frac{1}{n} \log P_{X^nY^n}(x^n,y^n) - H(P_{XY}) \right| < \epsilon, \left| -\frac{1}{n} \log P_{X^n}(x^n) - H(P_X) \right| < \epsilon, \text{ and } \left| -\frac{1}{n} \log P_{Y^n}(y^n) - H(P_Y) \right| < \epsilon \right\}$$
Properties of the jointly typical set: The following hold for any fixed \( \epsilon > 0 \). (For proofs, see Chapter 8, Section 8.6, of the Cover and Thomas book.)

1. For all sufficiently large \( n \), \( P_{X^n Y^n}(A^{(n)}_\epsilon) > 1 - \epsilon \)

2a. For every \( n \), \( |A^{(n)}_\epsilon| \leq 2^{n(H(P_{X,Y})+\epsilon)} \).

2b. For all sufficiently large \( n \), \((1-\epsilon)2^{n(H(P_{X,Y})-\epsilon)} \leq |A^{(n)}_\epsilon| \leq 2^{n(H(P_{X,Y})+\epsilon)} \).

3a. For every \( n \),

\[
P_{X^n Y^n}(A^{(n)}_\epsilon) \leq 2^{-n(I(P_{X,Y})-3\epsilon)} ,
\]

3b. For all sufficiently large \( n \),

\[
(1-\epsilon)2^{-n(I(P_{X,Y})+3\epsilon)} \leq P_{X^n Y^n}(A^{(n)}_\epsilon) \leq 2^{-n(I(P_{X,Y})-3\epsilon)} .
\]

Remarks: The properties \( A^{(n)}_\epsilon \) of may be qualitatively interpreted as follows.

1. Asymptotically, most of the probability mass of \( P_{X^n Y^n} \) is concentrated on the jointly \( \epsilon \)-typical sequences.

2. There are approximately \( 2^{nH(X,Y)} \) sequences in the jointly \( \epsilon \)-typical set, each with roughly the same \( P_{X^n Y^n} \)-probability of \( 2^{-nH(X,Y)} \).

3. The \( X^n \)-component of every pair \( (x^n, y^n) \) which is jointly \( \epsilon \)-typical for \( P_{XY} \) is guaranteed by definition to be \( \epsilon \)-typical for \( P_X \), and the \( Y^n \)-component to be \( \epsilon \)-typical for \( P_Y \). Thus \( A^{(n)}_\epsilon \) is contained in the cross-product \( P_X \)-typical \( \times \) \( P_Y \)-typical of sets. However, not every pair \( (x^n, y^n) \) obtained by choosing a \( P_X \)-typical \( x^n \) and a \( P_Y \)-typical \( y^n \) is jointly typical.

To see what fraction of them are indeed jointly typical, recall that the \( P_X \)-typical set has roughly \( 2^{nH(X)} \) \( n \)-length sequences and the \( P_Y \)-typical set has roughly \( 2^{nH(Y)} \) \( n \)-length sequences. Thus there are \( 2^{n(H(X)+H(Y))} \) ways of choosing a pair \( (x^n, y^n) \), both of whose components are individually typical. However, about \( 2^{nH(X,Y)} \) are jointly typical and therefore, the fraction of pairs which are jointly typical is roughly \( \frac{2^{nH(X,Y)}}{2^{n(H(X)+H(Y))}} = 2^{-n(H(X)+H(Y)-H(X,Y))} = 2^{-nI(X;Y)} \). When this is considered together with the fact that

(a) each pair of individually typical \( (x^n, y^n) \) has roughly equal probability under the the product (i.i.d.) pmf \( P_{X^n Y^n} \) and

(b) the product (i.i.d.) pmf puts most of its mass on the \( 2^{n(H(X)+H(Y))} \) pairs of individually typical sequences,

it follows that given a joint pmf \( P_{XY} \) on \( X \times Y \), if we generate i.i.d. random variables \( X^n \) according to \( P_X \) and independently generate i.i.d. random variables \( Y^n \) according to \( P_Y \), then the \( P_{X^n Y^n} \)-probability (not the \( P_{X^n Y^n} \)-probability) that the resulting pair will just happen to be jointly \( \epsilon \)-typical for \( P_{XY} \) is approximately \( 2^{-nI(X;Y)} \).

The last observation is crucial in understanding the proof of the channel coding theorem.