ECE 520.447
Introduction to Information Theory and Coding

Sanjeev Khudanpur

March 6, 2000

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 $P_{X^nY^n}$ denote the product (i.i.d.) measure on $\mathcal{X}^n \times \mathcal{Y}^n$ obtained from P_{XY}. *i.e.*,

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

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

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

$$\begin{aligned} A_\epsilon^{(n)} \;=\; \Big\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \;:\; & \left| -\frac{1}{n} \log P_{X^nY^n}(x^n, y^n) - H(P_{XY}) \right| < \epsilon, \\ & \left| -\frac{1}{n} \log P_{X^n}(x^n) - H(P_X) \right| < \epsilon, \text{ and } \left| -\frac{1}{n} \log P_{Y^n}(y^n) - H(P_Y) \right| < \epsilon \Big\} \end{aligned}$$

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, $P_{X^nY^n}\left(A_\epsilon^{(n)}\right) > 1 - \epsilon$

2 For every n, $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(P_{XY})+\epsilon)}$.

2a For all sufficiently large n, $(1-\epsilon)2^{n(H(P_{XY})-\epsilon)} \leq \left|A_\epsilon^{(n)}\right| \leq 2^{n(H(P_{XY})+\epsilon)}$.

3 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

$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = P_{X^n}(x^n)P_{Y^n}(y^n) = \prod_{i=1}^n P_X(x_i) \prod_{i=1}^n P_Y(y_i)$ denote the joint pmf of $(\tilde{X}^n, \tilde{Y}^n)$, and note that in general $P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) \neq P_{X^nY^n}(x^n, y^n)$, unless X and Y are independent under P_{XY}.

3a For every n,

$$P_{\tilde{X}^n\tilde{Y}^n}\left(A_\epsilon^{(n)}\right) \leq 2^{-n(I(P_{XY})-3\epsilon)},$$

3b For all sufficiently large n,

$$(1-\epsilon)2^{-n(I(P_{XY})+3\epsilon)} \leq P_{\tilde{X}^n\tilde{Y}^n}\left(A_\epsilon^{(n)}\right) \leq 2^{-n(I(P_{XY})-3\epsilon)}.$$

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

1. Asymptotically, most of the probability mass of $P_{X^nY^n}$ is concentrated on the jointly typical sequences.

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

3. The $\mathcal{X}^n$-component of every pair (x^n, y^n) which is jointly typical for P_{XY} is guaranteed by definition to be typical for P_X, and the $\mathcal{Y}^n$-component to be typical for P_Y. Thus $A_\epsilon^{(n)}$ is contained in the set of P_X-typical $\times$ P_Y-typical pairs. 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 $P_{\tilde{X}^n\tilde{Y}^n}$ 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 $\mathcal{X} \times \mathcal{Y}$, if we generate i.i.d. random variables $\tilde{X}^n$ according to P_X and *independently* generate i.i.d. random variables $\tilde{Y}^n$ according to P_Y, then the $P_{X^n}P_{Y^n}$-probability (not the $P_{X^nY^n}$-probability) that the resulting pair will be jointly typical for P_{XY} is approximately $2^{-nI(X;Y)}$.

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