Indian Institute of Technology Bombay

Department of Electrical Engineering

Handout 8 EE 708 Information Theory and Coding

Tutorial 3 Feb 17, 2022

Question 1) We say the pair of sequences (x^n, y^n) is jointly typical for some $\epsilon > 0$ if,

$$|\frac{1}{n}N\big((a,b)|(x^n,y^n)\big) - p(a,b)| \le \epsilon p(a,b),$$

where $N\big((a,b)|(x^n,y^n)\big)$ counts the number of times the pair $(a,b) \in \{\mathcal{X} \times \mathcal{Y}\}$ occurs in (x^n, y^n). The collection of all such typical (x^n, y^n) is called the jointly typical set $T_\epsilon^n(x,y)$. We will simply write this as T_ϵ^n, when the arguments are understood from the context. Another popular usage for this set is $\epsilon-$jointly typical set, or simply $\epsilon-$typical set.

(a) Show that $(x^n, y^n) \in T_\epsilon^n(x,y)$ implies $x^n \in T_\epsilon^n(x)$ and $y^n \in T_\epsilon^n(y)$.

Solution: Notice that $(x^n, y^n) \in T_\epsilon^n(x,y)$ implies that for each (a,b) with $a \in \mathcal{X}$ and $b \in \mathcal{Y}$, we have

$$\left|\frac{1}{n}N(a,b) - p(a,b)\right| \le \epsilon p(a,b),$$

where $N(a,b)$ counts the number of times (a,b) occurs in (x^n, y^n). In particular

$$\epsilon p(a,b) \le \frac{1}{n}N(a,b) - p(a,b) \le \epsilon p(a,b).$$

Taking the sum of over all $a \in \mathcal{X}$,

$$\epsilon \sum_{a\in\mathcal{X}} p(a,b) \le \frac{1}{n}\sum_{a\in\mathcal{X}} N(a,b) - \sum_{a\in\mathcal{X}} p(a,b) \le \epsilon \sum_{a\in\mathcal{X}} p(a,b).$$

Note that $\sum_a p(a,b) = p(b)$ and $\sum_a N(a,b) = N(b)$, the latter being the number of times b occurs in the sequence y^n. Thus

$$\left|\frac{N(b)}{n} - p(b)\right| \le \epsilon p(b) \Rightarrow y^n \in T_\epsilon^n(y).$$

Along similar lines, an interchange of variables can get the job done for showing $x^n \in T_\epsilon^n(x)$.

(b) Give bounds to the cardinality of $T_\epsilon^n(x,y)$.

Solution: We can proceed as in the case of one random sequence. Imagine we generate a sequence-pair (X^n, Y^n) by IID realizations of (X, Y) according to $p(x,y)$. Notice that for each (x^n, y^n), we have

$$Pr\big((X^n,Y^n) = (x^n,y^n)\big) =: p(x^n,y^n) = \prod_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p(a,b)^{N(a,b)},$$

where $N(a,b)$ is the number of times the ordered pair (a,b) occurs in $\{(x_1,y_1),\dots,(x_n,y_n)\}$, the latter is nothing but (x^n,y^n) written in an ordered manner. Specializing the above probability to the sequence-pairs inside T_ϵ^n, we get

$$\begin{aligned}
\forall (x^n,y^n) \in T_\epsilon^n :\ p(x^n,y^n) &= \prod_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p(a,b)^{N(a,b)} \\
&\le \prod_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p(a,b)^{np(a,b)(1-\epsilon)} \\
&= \prod_{(a,b)\in\mathcal{X}\times\mathcal{Y}} 2^{np(a,b)(1-\epsilon)\log p(a,b)} \\
&= 2^{n(1-\epsilon)\sum_{(a,b)} p(a,b)\log p(a,b)} \\
&= 2^{-n(1-\epsilon)H(X,Y)},
\end{aligned}$$

where the inequality used the definition of T_ϵ^n. Similarly, we can also show

$$\forall (x^n, y^n) \in T_\epsilon^n : \; p(x^n, y^n) \geq 2^{-n(1+\epsilon)H(X,Y)},$$

By law of large numbers, for a large enough n, we can take $1-\epsilon \geq P((X^n, Y^n) \in T_\epsilon^n) \leq 1$. Therefore

$$1-\epsilon \leq \sum_{(x^n,y^n)\in T_\epsilon^n} p(x^n, y^n) \leq 1. \tag{1}$$

Considering the first part of the expression,

$$\begin{aligned} 1-\epsilon &\leq \sum_{(x^n,y^n)\in T_\epsilon^n} p(x^n, y^n) && (2)\\ &\leq \sum_{(x^n,y^n)\in T_\epsilon^n} 2^{-n(1-\epsilon)H(X)} && (3)\\ &= |T_\epsilon^n| 2^{-n(1-\epsilon)H(X,Y)}, && (4) \end{aligned}$$

yielding

$$|T_\epsilon^n| \geq (1-\epsilon) 2^{n(1-\epsilon)H(X,Y)}.$$

Similarly,

$$\begin{aligned} 1 &\geq \sum_{(x^n,y^n)\in T_\epsilon^n} p(x^n, y^n) && (5)\\ &\geq \sum_{(x^n,y^n)\in T_\epsilon^n} 2^{-n(1+\epsilon)H(X)} && (6)\\ &= |T_\epsilon^n| 2^{-n(1+\epsilon)H(X,Y)}, && (7) \end{aligned}$$

yielding

$$|T_\epsilon^n| \leq 2^{n(1+\epsilon)H(X,Y)}.$$

Notice that $\epsilon > 0$ is an arbitrary positive constant, which can be taken arbitrary small. We will write this as ϵ_n to signify the fact that this quantity can be driven arbitrarily close to zero as n gets larger.

Question 2) Consider a pair of sequences (u^n, v^n) generated as IID realizations according to a joint distribution $p(u, v)$. Let w^n be another sequence, generated independent of everything else, as IID realizations according to $p(u)$.

(a) Give a upper bound to $Pr\big[(U^n, V^n) \notin T_\epsilon^n\big]$.

Solution: The event $\big\{(U^n, V^n) \notin T_\epsilon^n\big\}$ will take place if for some pair $(a, b) \in \mathcal{X} \times \mathcal{Y}$, we have its number of occurrences $N(a, b)$ NOT 'close' to $np(a, b)$. For any (a, b), consider the set of random variables $X_i = \mathbb{1}_{\{(U_i,V_i)=(a,b)\}}, 1 \leq i \leq n$. Observe that $X_1, \ldots, X_n$ is a sequence of IID Bernoulli random variables in $\{0, 1\}$, with mean $\mathbb{E}[X] = p(a, b)$. By weak law of large numbers,

$$Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| \geq \epsilon\right) \to 0,$$

for any $\epsilon > 0$. We can now check for every possible (a, b) in $\mathcal{U} \times \mathcal{V}$. Since there are only $|\mathcal{U}||\mathcal{V}| < \infty$ pairs of (a, b), the union bound suggests that we can take $Pr\big[(U^n, V^n) \notin T_\epsilon^n\big] \to 0$. Therefore we can take $Pr\big[(U^n, V^n) \notin T_\epsilon^n\big] \leq \epsilon$.

(b) Give a upper bound to $Pr\big[(W^n, V^n) \in T_\epsilon^n\big]$.

Solution: Notice that W^n is generated independently of V^n. Therefore,

$$Pr((W^n, V^n) = (w^n, v^n)) = Pr(W^n = w^n)Pr(V^n = v^n) = p(w^n)p(v^n),$$

and

$$Pr\big[(U^n, V^n) \in T_\epsilon^n\big] = \sum_{(w^n,v^n)\in T_\epsilon^n} p(w^n)p(v^n).$$

Since $w^n \in T_\epsilon^n(u)$ and $v^n \in T_\epsilon^n(v)$,

$$\sum_{(w^n,v^n)\in T_\epsilon^n} p(w^n)p(v^n) \le \sum_{(w^n,v^n)\in T_\epsilon^n} 2^{-nH(W)(1-\epsilon)}2^{-nH(V)(1-\epsilon)} \tag{8}$$

$$= \big|T_\epsilon^n\big| 2^{-n(1-\epsilon)(H(W)+H(V))}. \tag{9}$$

Notice that W has the same distribution as U, thus $H(W) = H(U)$. Furthermore, by the previous question, we have $|T_\epsilon^n| \le 2^{-n(1+\epsilon)H(U,V)}$. Putting it all together,

$$Pr\big[(U^n, V^n) \in T_\epsilon^n\big] \le 2^{n(1+\epsilon)H(X,Y)}2^{-n(1-\epsilon)(H(U)+H(V))} \tag{10}$$

$$= 2^{-n\Big(H(U)+H(V)-H(U,V)-\epsilon(H(U)+H(V)+H(U,V))\Big)} \tag{11}$$

$$\le 2^{-n[I(U;V)-\epsilon']}, \tag{12}$$

where we have carpeted the terms $\epsilon(H(U) + H(V) + H(U, V))$ by ϵ'. Since the terms used are bounded, $\epsilon' \to 0$ as $\epsilon \to 0$, which is all that we need for our proofs. Since $\epsilon' \to 0$, w.l.o.g. we can assume the exponent to be of the form $-nI(U; V)(1 - \hat{\epsilon})$ for an appropriately small $\hat{\epsilon}$, whenever $I(U; V)$ is non negligible. We will write this way only to keep the notation aligned with what we have been using so far. In the class, I wrote 3ϵ instead of $\hat{\epsilon}$, as there were multiple epsilon terms, please correct it to $\hat{\epsilon}$.

Question 3) Consider $K \times n$ matrix B with IID entries generated according to the distribution $p(x)$. Let $X^n(k), 1 \le k \le K$ denote the rows of the matrix thus generated. Consider a joint distribution $p(x, y)$ for which we compute the marginal $p(y) = \sum_x p(x, y)$, and generate an independent vector Y^n by IID drawings using $p(y)$. Let $\mathcal{E}$ be the event that at least one row of the matrix B is found jointly typical according to $p(x, y)$ with the sequence Y^n generated above. Find a bound on $Pr(\mathcal{E})$.

Solution: This is just an application of the previous question. In particular each row of the matrix is generated independent of Y^n. Assuming that $I(X; Y) > 0$,

$$Pr\big[(X^n(k), Y^n) \in T_\epsilon^n\big] \le 2^{nI(X;Y)(1-\hat{\epsilon})}. \tag{13}$$

Using union bound

$$Pr\Big[\bigcup_{1\le k\le K} \{(X^n(k), Y^n) \in T_\epsilon^n\}\Big] \le K.2^{nI(X;Y)(1-\hat{\epsilon})}. \tag{14}$$

Question 4) Suppose a fixed $M \times n$ codebook $\mathcal{C}$ is given to you. While using it for transmitting uniformly chosen messages $W \in \{1, \ldots, M\}$, it was found that the average message error probability $Pr(\hat{W} \ne W) \le \epsilon$. Show that there exists a codebook $\hat{\mathcal{C}}$ with at least $\frac{M}{2}$ codewords such that $Pr(\hat{W} \ne m | W = m) \le 2\epsilon, 1 \le m \le \frac{M}{2}$ for each codeword in $\hat{\mathcal{C}}$.

Solution: Imagine the average marks obtained for an exam in a class. Not more than half the people in that class can have double the average marks or more. Clearly if student each in the best half has exactly double the average, then every student in the bottom half should have zero to meet the average. More generally, for any set of non-negative numbers, we can assume that the maximum value of the bottom half is less than or equal to twice the average. Thus, if we arrange the codewords according to the ascending probability of error that it encounters, and then throw the bottom half (with more error), each of remaining will have at most 2ϵ error probability, where ϵ was the original average.

Question 5) Suppose a given $M \times r$ codebook $\mathcal{C}$ achieves an average message error probability of ϵ for a uniform choice of messages $W \in \{1, \dots, M\}$. The message error probability is also known as block error probability. Clearly the error probability depends on the decoding technique also. Let Y^n denote the received symbols in each block, and we apply the optimal ML decoding to recover the transmitted messages with the minimum error probability, i.e. $Pr(\hat{W} \neq W) \leq \epsilon$.

Show that the rate $R = \frac{1}{n}\log_2 M$ obeys the inequality

$$R(1+\epsilon) \leq \frac{I(W;Y^n)+1}{n}.$$

Solution: Will be done in the coming class.

Question 6) Show the convexity of $I(X;Y)$ in $p(y|x)$ for a fixed $p(x)$.

Solution: We will show that

$$\lambda I(X;Y_1) + (1-\lambda) I(X;Y_2) \geq I(X;Y_\lambda), \tag{15}$$

where (X, Y_j) has the joint distribution $p(x)p_j(y|x)$ for $j = 1, 2, \lambda$, with $0 \leq \lambda \leq 1$ and $p_\lambda(y|x) = \big(\lambda p(y_1|x) + (1-\lambda)p(y_2|x)\big)$. With this, notice that

$$I(X;Y_j) = D\Bigg(\Big[p(x)p_1(y|x)\Big]\Big|\Big|\Big[p(x)p_j(y)\Big]\Bigg) \text{ for } j \in \{1, 2, \lambda\}.$$

We also know (from the class) that $D(r||s)$ is convex in the pair (r, s), i.e.

$$\lambda D(r_1||s_1) + (1-\lambda) D(r_2||s_2) \geq D(r_\lambda||s_\lambda). \tag{16}$$

Let us take $r_1(x,y) = p(x)p_1(y|x)$; $r_2(x,y) = p(x)p_2(y|x)$; $s_1(x) = p(x)p_1(y)$ and $s_2(x) = p(x)p_2(y)$, where $p_j(y) = \sum_x p(x)p_j(y|x)$ for $j = 1, 2, \lambda$. With this, we get

$$\begin{aligned} r_\lambda(x,y) = \lambda r_1(x,y) + (1-\lambda) r_2(x,y) = p(x)(\lambda p_1(y|x) + (1-\lambda)p_2(y|x)) \quad &= p(x)p_\lambda(y|x) \\ s_\lambda(x,y) = \lambda s_1(x,y) + (1-\lambda) s_2(x,y) = p(x)(\lambda s_1(y) + (1-\lambda)s_2(y)) \quad &= p(x)p_\lambda(y). \end{aligned}$$

The convexity can be observed by substituting the above definition in (16).