Mathematical methods in communication Updated on Oct 15, 2015

Lecture 1: Method of types and strong typicality

Lecturer: Haim Permuter *Scribe: Avihay Shirazi and Offir Duvdevani*

I. A TYPES: DEFINITION ADN PROPERTIES

The method of types evolved from notions of strong typicality; some of the ideas were used by Wolfowitz [4] to prove channel capacity theorems. The method was fully developed by Csiszar and Korner [1], who derived the main theorems of information theory from this viewpoint.

We will start the lecture by defining a type of a sequence. Let $x^n = (x_1, x_2, ..., x_n)$ be a sequence from alphabet $\mathcal{X} = (a_1, a_2, a_3, ...a_{|\mathcal{X}|})$. Let $N(a|x^n)$ be the number of times that a appears in sequence x^n.

Definition 1 (Type) The type P_{x^n} (or empirical probability distribution) of a sequence x^n is the relative proportion of occurrences of each symbol of $\mathcal{X}$, i.e., $P_{x^n}(a) = \frac{N(a|x^n)}{n}$ for all $a \in \mathcal{X}$.

Example 1 Let $\mathcal{X} = \{0, 1, 2\}$, let $n = 5$ and $x^5 = (1, 1, 2, 2, 0)$. Then $N(0|x^5) = 1$, $N(1|x^5) = 2$ and $N(2|x^5) = 2$. Hence, $P_{x^n} = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right)$.

Definition 2 (all possible types) Let $\mathcal{P}_n$ be the collection of all possible types of sequences of length n.

For example, if $\mathcal{X} = \{0, 1\}$, the set of possible types with denominator n is

$$\mathcal{P}_n = \left\{ (P(0), P(1)) : \left(\frac{0}{n}, \frac{n}{n}\right), \left(\frac{1}{n}, \frac{n-1}{n}\right), ..., \left(\frac{n}{n}, \frac{0}{n}\right) \right\}. \tag{1}$$

Lemma 1 An upper bound for $|\mathcal{P}_n|$:

$$|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|}. \tag{2}$$

Proof: There are $|\mathcal{X}|$ components in the vector that specifies P_{x^n}. The numerator in each component can take on only $n+1$ values. So there are at most $(n+1)^{|\mathcal{X}|}$ choices for the type vector. ■

Definition 3 (Type class) Let $P \in \mathcal{P}_n$, The set of sequences of length n with type P is called type class of P, denoted $T(P)$:

$$T(P) = \{x^n : P_{x^n} = P\} \tag{3}$$

Lets us now define the notation $Q^n(x^n)$ that emphasis that X^n is distributed i.i.d according to $Q(x)$. In other words,

$$Q^n(x^n) \triangleq \prod_{i=1}^{n} Q(x_i). \tag{4}$$

Theorem 1 (Probability of a sequence in the type class) If $X \sim Q$ i.i.d., the probability of x^n depends only on the type of x^n, i.e., P_{x^n}

$$Q^n(x^n) = 2^{-n(H(P_{x^n})+D(P_{x^n}||Q))} \tag{5}$$

Proof: Consider

$$\begin{aligned}
\log Q^n(x^n) &= \sum_{i=1}^{n} \log Q(x_i) && (6)\\
&\overset{(a)}{=} \sum_{a\in\mathcal{X}} N(a|x^n) \log Q(a) && (7)\\
&\overset{(b)}{=} n \sum_{a\in\mathcal{X}} P_{x^n}(a) \log Q(a) && (8)\\
&= n \sum_{a\in\mathcal{X}} P_{x^n}(a) \log \frac{Q(a)}{P_{x^n}(a)} \cdot P_{x^n}(a) && (9)\\
&= n(-H(P) - D(P||Q)), && (10)
\end{aligned}$$

where

(a) follows because each $a \in \mathcal{X}$ contributes exactly $\log Q(a)$ times it's number of occurences in x^n to the sum in (6).

(b) follows from the definition of $P_{x^n}(a)$.

Hence we obtained

$$Q^n(x^n) = 2^{(-nH(P)+D(P||Q))}. \tag{11}$$

■

Corollary 1 if x^n is in the type class of Q, then we get $Q^n(x^n) = 2^{-nH(P_{x^n})}$.

The following theorem tells us how many sequences, asymptotically, exist of type $P \in \mathcal{P}_n$.

Theorem 2 (size of a type class) For any type $P \in \mathcal{P}_n$

$$|T(p)| \doteq 2^{nH(P)} \tag{12}$$

Where $a_n \doteq b_n$ if $\lim\limits_{n\to\infty} \frac{1}{n} \log(\frac{a_n}{b_n}) = 0$.

Example 2 Question: How many binary sequences of length n with 50% 0 and 50% 1 exists?
Answer: An exact calculation yields $\binom{n}{\frac{n}{2}}$. An asymptotic calculation Using Theorem 2 yields that $\binom{n}{\frac{n}{2}} \doteq 2^n$.

There are two possible ways to prove Theorem 2, one is a combinatorial proof and the other is a probabilistic. We will provide both proofs in two different subsections.

II. Combinatorial proof of Theorem 2

Lemma 2 (Stirling's formula) :
The combinatorial proof is based on Stirling's Formula:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \tag{13}$$

proof of Theorem 2:

$$|T(P)| = \binom{n}{nP(a_1), nP(a_2), \dots, nP(a_{\mathcal{X}})} = \frac{n!}{(nP(a_1))!(nP(a_2))! \dots (nP(a_{|\mathcal{X}|}))!} \tag{14}$$

Using Stirling's formula with equation (5) we get:

$$n! \ \doteq \ \left(\frac{n}{e}\right)^n \tag{15}$$

a_1	a_2		$a_{\lvert\mathcal{X}\rvert}$
$nP(a_1)$	$nP(a_2)$	$\ldots$	$nP(a_{\lvert\mathcal{X}\rvert})$

Fig. 1. The total length of the sequence is n and the part of the sequence that equals to a_i is $nP(a_i)$

$$\begin{aligned}
|T(P)| &\doteq \frac{n^n}{(nP(a_1))^{nP(a_1)}(nP(a_2))^{nP(a_2)}\ldots(nP(a_{|\mathcal{X}|}))^{nP_{|\mathcal{X}|}}} && (16)\\
&= \frac{n^n}{(n)^{nP(a_1)}(n)^{nP(a_2)}\ldots(n)^{nP_{|\mathcal{X}|}}\prod_{i=1}^{|\mathcal{X}|}P(a_i)^{nP(a_i)}} && (17)\\
&= \frac{1}{\prod_{i=1}^{|\mathcal{X}|}P(a_i)^{nP(a_i)}} && (18)
\end{aligned}$$

Hence:

$$|T(P)| = 2^{\log|T(P)|} \doteq 2^{-n\sum_{i=1}^{|\mathcal{X}|}P(a_i)\log(P(a_i))} = 2^{nH(P)} \qquad (19)$$

■

III. Probabilistic proof of Theorem 2:

From the probabilistic proof we will obtain two bounds that implies Theorem 2. The bounds are:

$$\frac{2^{nH(P)}}{(n+1)^{|\mathcal{X}|}} \le |T(P)| \le 2^{nH(P)}. \qquad (20)$$

Proof: Let's assume the sequence X^n is distributed i.i.d according to $P(x)$. Now consider the following:

$$\begin{aligned}
1 &\ge \Pr(x^n \in T(P)) && (21)\\
&\overset{(a)}{=} \sum_{x^n\in T(P)} \Pr(x^n) && (22)\\
&\overset{(b)}{=} \sum_{x^n\in T(P)} 2^{-nH(P)} && (23)\\
&= |T(P)|2^{-nH(P)}. && (24)
\end{aligned}$$

Equality (a) follows from the fact that the probability of a subset equals to the sum of the probabilities of each element in the subset. For example, if we have a set A, B, C where

the probability of choosing A is P_A, B is P_B and C is P_C, where $P_A + P_B + P_C = 1$, then the probability of choosing from the subset (A, B) is $P_A + P_B$. Equality (b) follows from Theorem 1.
Therefore:

$$|T(P)| \leq 2^{nH(P)} \tag{25}$$

In order to prove the other part we need the following lemma:

Lemma 3 $P^n(T(P)) \geq P^n(T(Q))$

Proof: Let X^n be of a type P. The term $P^n(T(P))$ is the probability of type class $T(P)$ where the sequences of length n are drawn according to $P(x^n) = \prod_{i=1}^n P(x_i)$, and let $Q \in \mathcal{P}_n$.
Consider

$$\begin{aligned}
\frac{P^n(T(P))}{P^n(T(Q))} &\overset{(a)}{=} \frac{|T(P)| \prod_{a\in\mathcal{X}} P(a)^{nP(a)}}{|T(Q)| \prod_{a\in\mathcal{X}} P(a)^{nQ(a)}} &(26)\\
&\overset{(b)}{=} \frac{\binom{n}{nP(a_1),nP(a_2),\ldots,nP(a_{|\mathcal{X}|})} \prod_{a\in\mathcal{X}} P(a)^{nP(a)}}{\binom{n}{nQ(a_1),nQ(a_2),\ldots,nQ(a_{|\mathcal{X}|})} \prod_{a\in\mathcal{X}} P(a)^{nQ(a)}} &(27)\\
&\overset{(c)}{=} \prod_{a\in\mathcal{X}} \frac{(nQ(a))!}{(nP(a))!} P(a)^{n(P(a)-Q(a))} &(28)
\end{aligned}$$

(a) Using the fact that probability of each type $P_{x^n} \in \mathcal{P}_n$ is given by:
$P_{x^n} = \prod_{i=1}^n P(x_i) = \prod_{a\in\mathcal{X}} P(a)^{N(a|x^n)} = \prod_{a\in\mathcal{X}} P(a)^{nP(a)}$.
(b) Using combinatorical math it is known that the number of possibilities to arange a vector $\{x^n : P_{x^n} = P\}$ is: $\binom{n}{nP(a_1),nP(a_2),\ldots,nP(a_{|\mathcal{X}|})}$.
(c) $\frac{\binom{n}{nP(a_1),nP(a_2),\ldots,nP(a_{|\mathcal{X}|})}}{\binom{n}{nQ(a_1),nQ(a_2),\ldots,nQ(a_{|\mathcal{X}|})}} = \prod_{a\in\mathcal{X}} \frac{(nQ(a))!}{(nP(a))!}$
Using the simple bound $\frac{m!}{n!} \geq n^{m-n}$ we obtain:

$$\begin{aligned}
\frac{P^n(T(P)}{P^n(T(Q))} &\geq \prod_{a\in\mathcal{X}} (nP(a))^{nQ(a)-nP(a)} P(a)^{n(P(a)-Q(a))} &(29)\\
&= \prod_{a\in\mathcal{X}} n^{n(Q(a)-P(a))} &(30)\\
&= n^{n(\sum_{a\in\mathcal{X}} Q(a) - \sum_{a\in\mathcal{X}} P(a))} &(31)
\end{aligned}$$

$$= \quad n^{n(1-1)} = 1 \tag{32}$$

■

Using Lemma 3 let us show that $|T(P)| \geq \frac{2^{nH(P)}}{(n+1)^{|\mathcal{X}|}}$:

$$\begin{aligned}
1 &= \sum_{Q \in \mathcal{P}_n} P^n(T(Q)) && (33)\\
&\leq \sum_{Q \in \mathcal{P}_n} \max_Q P^n(T(Q)) && (34)\\
&\overset{(a)}{=} \sum_{Q \in \mathcal{P}_n} P^n(T(P)) && (35)\\
&\overset{(b)}{\leq} (n+1)^{|\mathcal{X}|} P^n(T(P)) && (36)\\
&\overset{(c)}{=} (n+1)^{|\mathcal{X}|} \sum_{x^n \in T(P)} 2^{-nH(P)} && (37)\\
&= (n+1)^{|\mathcal{X}|} |T(P)| 2^{-nH(P)} && (38)
\end{aligned}$$

(a) Using theorem 2 it is clear that: $\max_Q P^n(T(Q)) = P^n(T(P))$.

(b) Using Lemma 1.

(c) Using Theorem 3.

Therefore our final result is:

$$\frac{2^{nH(P)}}{(n+1)^{|\mathcal{X}|}} \leq |T(P)| \leq 2^{nH(P)} \tag{39}$$

which implies that:

$$|T(P)| \doteq 2^{nH(P)} \tag{40}$$

■

IV. Probability of a type and of a set of types (Sanov's Theorem)

Theorem 3 The probability of the type class $T(P)$ where the sequences are drawn i.i.d. $\sim Q$ is

$$Q^n(T(P)) \doteq 2^{-n(D(P||Q)}. \tag{41}$$

Proof:

$$
\begin{aligned}
Q^n(T(P)) &= \sum_{x^n \in T(P)} Q(x^n) && (42)\\
&\overset{(a)}{=} \sum_{x^n \in T(P)} 2^{-n(H(P_{x^n})+D(P_{x^n}||Q))} && (43)\\
&\overset{(b)}{=} \sum_{x^n \in T(P)} 2^{-n(H(P)+D(P||Q))} && (44)\\
&= |T(P)| 2^{-n(H(P)+D(P||Q))} && (45)\\
&\overset{(c)}{\doteq} 2^{-nD(P||Q)}, && (46)
\end{aligned}
$$

where (a) follows from Theorem 1, (b) from the fact that all sequences have the same type $P_{x^n} = P$ and (c) from Theorem 2.

One can also obtain more explicit bounds by using the explicit bounds on $|T(P)|$ given in (39):

$$\frac{2^{-nD(P||Q)}}{(n+1)^{|\mathcal{X}|}} \leq Q^n(T(P)) \leq 2^{-nD(P||Q)} \tag{47}$$

■

Next we state Sanov's theorem, [3] which was generalized by Csiszar [2] using the method of types. It also opened a new field in statistics called *Large Deviation*.

Theorem 4 (Sanov's Theorem) Let $X \sim Q$ i.i.d. and let E be a set of probabilities that is the closure of its interior, then:

$$\lim_{n \to \infty} \log Q^n(E) = -\min_{P \in E} D(P||Q) = -D(P^*||Q), \tag{48}$$

where $Q^n(E)$ is the probability that $x^n \in E$ i.e. $Q^n(E) = \Pr(P \in E)$ and P^* is defined as $P^* = \arg\min_{P \in E} D(P||Q)$.

To get more intuitive understanding we can think of $D(P^*||Q)$ as the minimum distance between E space and Q as shown in the figure:

$$Q^n(E) \doteq 2^{-nD(P^*||Q)} \tag{49}$$

$$P^* = \arg\min_{P \in E} D(P||Q) \tag{50}$$

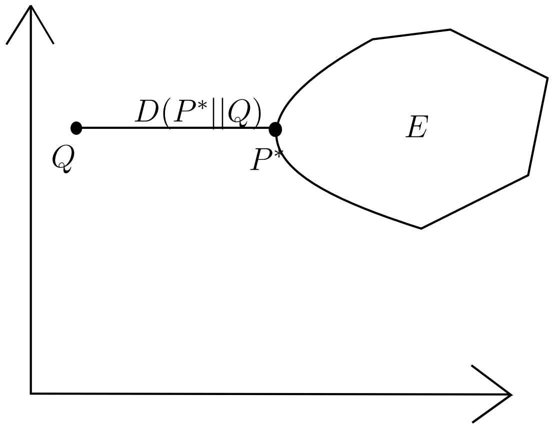

Fig. 2. Let $X \sim Q$ than P^* is the type $P \in E$ that gives the minimum to $D(P||Q)$.

Example 3 Let $Q(x = 1) = Q(x = -1) = \frac{1}{2}$, What is the probability of getting an empirial distrebution that satisfies: $P(x = 1) \geq 0.8$, $P(x = -1) \leq 0.2$?
Answer: P^* is the probability $P(x = 1) = 0.8$, $P(x = -1) = 0.2$ so by using Sanov theorem and Theorem 4 we get our result: $Q(E) \doteq 2^{-nD(P^*||Q)}$

Proof of Theorem 4: First we will find the upper bound

$$\begin{aligned}
Q^n(E) &= \sum_{P \in E \bigcap \mathcal{P}_n} Q^n(T(P)) && (51)\\
&\overset{(a)}{\leq} \sum_{P \in E \bigcap \mathcal{P}_n} 2^{-nD(P||Q)} && (52)\\
&\leq \sum_{P \in E \bigcap \mathcal{P}_n} \max_{p \in E \bigcap \mathcal{P}_n} 2^{-nD(P||Q)} && (53)\\
&= \sum_{P \in E \bigcap \mathcal{P}_n} 2^{-n \min_{P \in E \bigcap \mathcal{P}_n} D(P||Q)} && (54)\\
&\overset{(b)}{\leq} (n+1)^{|\mathcal{X}|} 2^{-n \min_{P \in E \bigcap \mathcal{P}_n} D(P||Q)}, && (55)
\end{aligned}$$

where (a) follows from Theorem 3, and (b) follows from the fact that $|E| \leq |\mathcal{P}_n|$ and the bound on the number of types (Lemma 1).

The minimum of $\min_{P\in E\bigcap\mathcal{P}_n} D(P||Q)$ exists since E is closed, further more because E is the closure of its interior and Divergence is continues, it implies that the $\lim_{n\to\infty}\min_{P\in E\bigcap\mathcal{P}_n} D(P||Q)$ exists and is obtained by some $P^*\in E$.

Now we will find the lower bound:

$$\begin{aligned} Q^n(E) &= \sum_{P\in E\bigcap\mathcal{P}_n} Q^n(T(P)) && (56)\\ &\overset{(a)}{\geq} \min_{P\in E\bigcap\mathcal{P}_n} Q(T(P)) && (57)\\ &\overset{(b)}{\doteq} \min_{P\in E\bigcap\mathcal{P}_n} 2^{-nD(P||Q)} && (58)\end{aligned}$$

where (a) follows from the fact that we take into consideration only one type and (b) According to Theorem 3.

We can obtain a more explicit bound using (47) in the last step:

$$Q^n(E) \geq \frac{1}{(n+1)^{|\mathcal{X}|}} 2^{-n \min\limits_{P\in E\bigcap\mathcal{P}_n} D(P||Q)}. \tag{59}$$

Combining the lower bound (55) and upper bound (59) we have

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{-n \min\limits_{P\in E\bigcap\mathcal{P}_n} D(P||Q)} \leq Q^n(E) \leq (n+1)^{|\mathcal{X}|} 2^{-n \min\limits_{P\in E\bigcap\mathcal{P}_n} D(P||Q)}. \tag{60}$$

$$\tag{61}$$

which implies

$$Q^n(E) \doteq 2^{-nD(P^*||Q)} \tag{62}$$

■

V. Joint type

Definition 4 (Joint type) The type P_{x^n,y^n} (or empirical probability distribution) of a pair-sequence (x^n, y^n) is the relative proportion of occurrences of each pair-symbol of $\mathcal{X}\times\mathcal{Y}$, i.e., $P_{x^n,y^n}(a,b) = \frac{N(a,b|x^n,y^n)}{n}$ for all $a\in\mathcal{X}$ and $b\in\mathcal{X}$.

Example 4 Let $\mathcal{X} = \{0,1\}$, and $\mathcal{Y} = \{A,B\}$. let $n = 5$ and $x^5 = (1,1,0,1,0)$ and $y^5 = (A,A,B,A,B)$. Then $N(0,A|x^5) = 0$, $N(0,B|x^5) = 2$, $N(1,A|x^5) = 3$ and $N(1,A|x^5) = 0$.

Theorem 5 (Conditional type)

Let us define the *conditional type* $P_{x^n|y^n}$ (or conditional empirical distribution)

$$P_{x^n|y^n}(a|b) \triangleq \frac{N((a,b)|x^n,y^n)}{N(b|y^n)} \tag{63}$$

$$= \frac{P_{X^n,Y^n}(a,b)}{P_{Y^n}(b)}. \tag{64}$$

Let $W(y|x) \in \mathcal{P}^n(x|y)$ be a conational probability, The conditional type $T_W(y^n)$

$$T_W(y^n) = \{x^n \in \mathcal{X}^n : P_{X^n|Y^n}(a|b) = W_{X|Y}(a|b), \forall a, b \in \mathcal{X}, \mathcal{Y}\} \tag{65}$$

$$= \{x^n \in \mathcal{X}^n : P_{X^n,Y^n}(a,b) = W_{X|Y}(a|b)P_{Y^n}(b), \forall a, b \in \mathcal{X}, \mathcal{Y}\} \tag{66}$$

$$H(X|Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x|y) \tag{67}$$

$$P_{X,Y}(a,b) = P_{Y^n}(b)W_{X|Y}(a|b) \tag{68}$$

Than:

$$|T_W(y^n)| \doteq 2^{nH(X|Y)} \tag{69}$$

Proof:

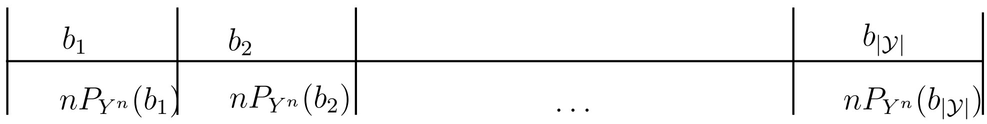

Fig. 3. Length of each b_i.

Now if we have b_1 we get:

Therefore we can use combinatorical proof as we did in the non conditional case:

$$\binom{nP_{y^n}(b_1)}{nP_{x^n,y^n}(a_1,b_1)nP_{x^n,y^n}(a_2,b_1)\ldots nP_{x^n,y^n}(a_{|\mathcal{X}|},b_1)} \doteq 2^{nH(X|y=b_1)P_{y^n}(b_1)} \tag{70}$$

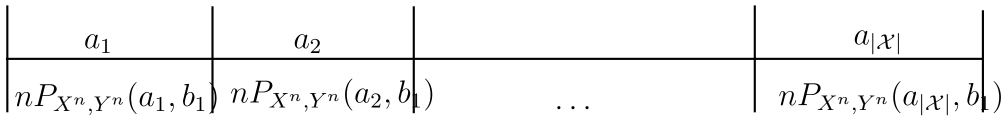

Fig. 4. Length of each a_i given b_1.

$$\binom{nP_{Y^n}(b_1)}{nP_{Y^n}(b_1)P_{x^n|y^n}(a_1|b_1)nP_{Y^n}(b_1)P_{x^n|y^n}(a_2|b_1)\dots nP_{Y^n}(b_1)P_{x^n|y^n}(a_{|\mathcal{X}|}|b_1)} \doteq 2^{nP_{Y^n}(b_1)H(X|y=b_1)} \tag{71}$$

$$|T_W(y^n)| \doteq \prod_{i=1}^{|\mathcal{Y}|} 2^{nH(X|y=b_i)P_{Y^n}(b_i)} = 2^{nH(X|Y)} \tag{72}$$

VI. STRONG TYPICALITY

Definition 5 (ϵ-strongly typical) A sequence $x^n \in \mathcal{X}$ is said to be *ϵ-strongly typical* with respect to a distribution $P(x)$ on $\mathcal{X}$ if

1) For all $a \in \mathcal{X}$ with $P_X(a) > 0$ we have

$$|P_{x^n}(a) - P_X(a)| \leq \frac{\epsilon}{|\mathcal{X}|} \tag{73}$$

2) If $P_X(a) = 0$ then $P_{x^n}(a) = 0$.

Definition 6 (ϵ-joint strongly typical) A pair of sequences $(x^n, y^n) \in \mathcal{X} \times \mathcal{Y}$ is said to be *ϵ-joint strongly typical* with respect to a distribution $P(x,y)$ on $\mathcal{X} \times \mathcal{Y}$ if

1) For all $(a,b) \in \mathcal{X} \times \mathcal{Y}$ with $P_{X,Y}(a,b) > 0$ we have

$$|P_{x^n,y^n}(a,b) - P_{X,Y}(a,b)| \leq \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|} \tag{74}$$

2) If $P_{X,Y}(a,b) > 0$ then $P_{x^n,y^n}(a,b) = 0$.

Definition 7 (Strongly typical set) The set of sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ that are ϵ-joint strongly typical is called *strongly typical set* and is denoted as $T_\epsilon^{(n)}(X,Y)$ or $T_\epsilon^{(n)}(P_{X,Y})$. I.e.,

$$T_\epsilon^{(n)}(X,Y) \triangleq \{(x^n,y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : |P_{x^n,y^n} - P(x,y)| \leq \epsilon\} \tag{75}$$

In a shorter notation we write it as

$$T_\epsilon^{(n)}(X,Y) \triangleq \{x^n, y^n : |P_{x^n,y^n} - P(x,y)| \leq \epsilon\}. \tag{76}$$

Definition 8 (Strongly conditional typical set $T_\epsilon^{(n)}(X|y^n)$)

$$\begin{aligned} T_\epsilon^{(n)}(Y|x^n) &\triangleq \{y^n : (x^n, y^n) \in T_\epsilon^{(n)}(X,Y)\} \\ &= \{y^n : |P_{x^n,y^n} - P(x,y)| \le \epsilon\}. \end{aligned} \tag{77}$$

The following lemma follows directly from the Sanov's Theorem.

Lemma 4 Suppose w.l.o.g. that $Q(x) > 0$ for all $x \in \mathcal{X}$ (otherwise shrink the alphabet to its effective size) and that X_i are i.i.d. $\sim Q(x)$. Then there exists $\epsilon'(\epsilon)$ such that $\epsilon'(\epsilon) \to 0$ as $\epsilon \to 0$ such that for all sufficiently large n

$$2^{-n[D(P||Q)+\epsilon']} \le \Pr(X^n \in T_\epsilon^{(n)}(P)) \le 2^{-n[D(P||Q)-\epsilon']}. \tag{78}$$

Lemma 5 Consider a joint distribution $P_{X,Y}$ with marginal P_X and P_Y. Generate X^n i.i.d. $\sim P_X$ and $\sim P_Y$. Then

$$2^{-n[I(X;Y)+\epsilon']} \le \Pr((X^n, Y^n) \in T_\epsilon^{(n)}(X,Y)) \le 2^{-n[I(X;Y)-\epsilon']}. \tag{79}$$

VII. Alternative proofs of properties of strongly typical set based only on probabilistic methods

Let us define the strong type slightly differen but equivalently as follows:

$$T_\epsilon^{(n)}(P_X) = T_\epsilon^{(n)}(X) = \left\{x^n \in \mathcal{X}^n : \ |P_{x^n}(a) - P_X(a)| \le \epsilon P_X(a)\right\} \tag{80}$$

and

$$T_\epsilon^{(n)}(P_{XY}) = T_\epsilon^{(n)}(X,Y) = \left\{x^n, y^n : \ |P_{x^n y^n}(a,b) - P_{XY}(a,b)| \le \epsilon P_{XY}(a,b)\right\}. \tag{81}$$

Properties:

1) If $X^n \sim P_X \ \text{ i.i.d. }$, then

$$\lim_{n\to\infty} \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} = 1 \tag{82}$$

or equivalently

$$1 - \delta_n(\epsilon) \le \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} \le 1, \tag{83}$$

where $\forall \epsilon > 0, \ \lim_{n\to\infty} \delta_n(\epsilon) = 0.$

Proof: Because of the L.L.N. ■

2) For $X^n \sim P_X,$ i.i.d. , then for any sequence $x^n \in T_\epsilon^{(n)}(X)$ we have

$$2^{-nH(X)(1+\epsilon)} \stackrel{(i)}{\leq} p(x^n) = \prod_{i=1}^{n} P_X(x_i) \stackrel{(ii)}{\leq} 2^{-nH(X)(1-\epsilon)}. \tag{84}$$

Proof: Note that $p(x^n) = \prod_{a\in\mathcal{X}} P_X(a)^{N(a|x^n)}$, hence, the left hand side of the inequality (i) follows from

$$\frac{1}{n}\log p(x^n) = \frac{1}{n}\sum_{a\in\mathcal{X}} N(a|x^n)\log P_X(a) \tag{85}$$

$$= \sum_{a\in\mathcal{X}} P_{x^n}(a)\log P_X(a) \tag{86}$$

$$\stackrel{(a)}{\geq} \sum_{a\in\mathcal{X}} (1+\epsilon)P_X(a)\log P_X(a) \tag{87}$$

$$= -H(X)(1+\epsilon), \tag{88}$$

where step (a) follows from (80).

The right hand side of the inequality (ii) follows from

$$\frac{1}{n}\log p(x^n) = \frac{1}{n}\sum_{a\in\mathcal{X}} N(a|x^n)\log P_X(a) \tag{89}$$

$$= \sum_{a\in\mathcal{X}} P_{x^n}(a)\log P_X(a) \tag{90}$$

$$\leq \sum_{a\in\mathcal{X}} (1-\epsilon)P_X(a)\log P_X(a) \tag{91}$$

$$= -H(X)(1-\epsilon). \tag{92}$$

■

3) The size of the strongly typical set can be bounded as

$$(1-\delta_{\epsilon,n})2^{nH(X)(1-\epsilon)} \stackrel{(i)}{\leq} |T_\epsilon^{(n)}(X)| \stackrel{(ii)}{\leq} 2^{nH(X)(1+\epsilon)}. \tag{93}$$

Proof: Recall that under the assumption that $X^n \sim P_X,$ i.i.d. , we have $(1-\delta) \leq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} \leq 1$. Now we prove the left hand side inequality (i):

$$\Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} = \sum_{x^n\in T_\epsilon^{(n)}(X)} p(x^n) \tag{94}$$

$$\leq \sum_{x^n \in T_\epsilon^{(n)}(X)} 2^{-n(H(X)(1-\epsilon)} \tag{95}$$

$$= |T_\epsilon^{(n)}(X)| 2^{-nH(X)(1-\epsilon)}, \tag{96}$$

and because $(1-\delta) \leq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\}$, we get that $(1-\delta)2^{nH(X)(1-\epsilon)} \leq |T_\epsilon^{(n)}(X)|$. The right hand side inequality (ii) follows from

$$1 \geq \Pr\left\{X^n \in T_\epsilon^{(n)}(X)\right\} \tag{97}$$

$$= \sum_{x^n \in T_\epsilon^{(n)}(X)} p(x^n) \tag{98}$$

$$\geq \sum_{x^n \in T_\epsilon^{(n)}(X)} 2^{-nH(X)(1+\epsilon)} \tag{99}$$

$$= |T_\epsilon^{(n)}(X)| 2^{-nH(X)(1+\epsilon)}, \tag{100}$$

therefore, $|T_\epsilon^{(n)}(X)| \leq 2^{nH(X)(1+\epsilon)}$. ■

Conditionally strong typical set

For a given $x^n \in \mathcal{X}^n$, let us define

$$T_\epsilon^{(n)}(Y|x^n) = \left\{y^n : \ (x^n, y^n) \in T_\epsilon^{(n)}(X,Y)\right\}. \tag{101}$$

Notice that if $(x^n, y^n) \in T_\epsilon^{(n)}(X,Y)$, then surely $x^n \in T_\epsilon^{(n)}(X)$.

Properties:

1) If $x^n \in T_{\epsilon_x}^{(n)}(X),\ p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ (DMC) , then for all $0 < \epsilon_x < \epsilon$

$$1 - \delta_{\epsilon,\epsilon_x,n} \leq \Pr\left\{y^n \in T_\epsilon^{(n)}(Y|x^n)\right\} \leq 1 \tag{102}$$

where $\delta_{\epsilon,\epsilon_x,n} \to 0$ as $n \to \infty$ for all $0 < \epsilon_x \leq \epsilon$.

Proof: Follows directly from the L.L.N. ■

2) If $y^n \in T_\epsilon^{(n)}(Y|x^n), \quad p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ (DMC), then

$$2^{-nH(Y|X)(1+\epsilon)} \overset{(i)}{\leq} p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i) \overset{(ii)}{\leq} 2^{-nH(Y|X)(1-\epsilon)}. \tag{103}$$

Proof: Notice that

$$p(y^n|x^n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) = \prod_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} P_{Y|X}(b|a)^{N(a,b|x^n,y^n)}. \tag{104}$$

Now, the left hand side of the inequality (i) follows from

$$\begin{aligned}
\frac{1}{n}p(y^n|x^n) &= \frac{1}{n}\sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} N(a,b|x^n,y^n)\log P_{Y|X}(b|a) && (105)\\
&= \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} P_{x^n,y^n}(a,b)\log P_{Y|X}(b|a) && (106)\\
&\geq \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} (1+\epsilon)P_{XY}(a,b)\log P_{Y|X}(b|a) && (107)\\
&= -H(Y|X)(1+\epsilon). && (108)
\end{aligned}$$

The right hand side of the inequality (ii) follows from

$$\begin{aligned}
\frac{1}{n}p(y^n|x^n) &= \frac{1}{n}\sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} N(a,b|x^n,y^n)\log P_{Y|X}(b|a) && (109)\\
&= \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} P_{x^n,y^n}(a,b)\log P_{Y|X}(b|a) && (110)\\
&\leq \sum_{\substack{a\in\mathcal{X}\\ b\in\mathcal{Y}}} (1-\epsilon)P_{XY}(a,b)\log P_{Y|X}(b|a) && (111)\\
&= -H(Y|X)(1-\epsilon). && (112)
\end{aligned}$$

■

3) Given $x^n \in T_{\epsilon_x}^{(n)}(X)$, then

$$(1-\delta_{\epsilon,\epsilon_x,n})2^{nH(Y|X)(1+\epsilon)} \overset{(i)}{\leq} |T_{\epsilon}^{(n)}(Y|x^n)| \overset{(ii)}{\leq} 2^{nH(Y|X)(1-\epsilon)}. \tag{113}$$

The proof is done in a similar way to the proof of (93).

Lemma 6 Consider a joint PMF P_{XY}. Let $x^n \in T_{\epsilon_x}^{(n)}(X)$ and y^n drawn i.i.d. according to P_Y and independent of x^n, then

$$2^{-n\left(I(X;Y)+\delta_\epsilon\right)} \overset{(i)}{\leq} \Pr\left\{Y^n \in T_{\epsilon}^{(n)}(Y|x^n)\right\} \overset{(ii)}{\leq} 2^{-n\left(I(X;Y)-\delta_\epsilon\right)} \tag{114}$$

for $\delta_\epsilon \to 0$ as $\epsilon \to 0$.

Proof: The left hand side inequality (i) follows from

$$\Pr\left\{Y^n \in T_\epsilon^{(n)}(Y|x^n)\right\} = \sum_{y^n \in T_\epsilon^{(n)}(Y|x^n)} p(y^n) \tag{115}$$

$$\geq \sum_{y^n \in T_\epsilon^{(n)}(Y|x^n)} 2^{-nH(Y)(1+\epsilon)} \tag{116}$$

$$= |T_\epsilon^{(n)}(Y|x^n)\big\}|2^{-nH(Y)(1+\epsilon)} \tag{117}$$

$$\geq (1-\delta_{\epsilon,n})2^{n(H(Y|X)(1-\epsilon)}2^{-nH(Y)(1+\epsilon)} \tag{118}$$

$$= (1-\delta_{\epsilon,n})2^{-n(I(X;Y)+\delta_{\epsilon,n})}. \tag{119}$$

The right hand side inequality (ii) follows from

$$\Pr\left\{Y^n \in T_\epsilon^{(n)}(Y|x^n)\right\} = \sum_{y^n \in T_\epsilon^{(n)}(Y|x^n)} p(y^n) \tag{120}$$

$$\leq \sum_{y^n \in T_\epsilon^{(n)}(Y|x^n)} 2^{-nH(Y)(1-\epsilon)} \tag{121}$$

$$= |T_\epsilon^{(n)}(Y|x^n)|2^{-nH(X)(1-\epsilon)} \tag{122}$$

$$\leq 2^{n(H(Y|X)(1+\epsilon)}2^{-nH(Y)(1-\epsilon)} \tag{123}$$

$$= 2^{-n\left(I(X;Y)-\delta_\epsilon\right)}. \tag{124}$$

■

REFERENCES

[1] I. Csiszar and J. Korner. Information Theory: Coding Theorems for Discrete Memoryless Systems. Academic Press, New York, 1981.

[2] I Csiszar. Sanov property, generalized I-projection and a conditional limit theorem. Ann. Prob., 12:768793, 1984.

[3] I. N. Sanov. On the probability of large deviations of random variables. Mat. Sbornik, 42:1144, 1957. English translation in Sel. Transl. Math. Stat. Prob., Vol. 1, pp. 213-244, 1961.

[4] J. Wolfowitz. Coding Theorems of Information Theory. Springer-Verlag, Berlin, and Prentice-Hall, Englewood Cliffs, NJ, 1978.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New-York: Wiley, 2006.