Multi-User Information Theory November 16th, 2009

Lecture 5

Lecturer: Haim Permuter *Scribe: Offir Duvdevani*

I. METHOD OF TYPES

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.

Let $X_1, X_2, ...X_n$ 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 $x_1, x_2, x_3, ..., x_n$ 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 $a \in \mathcal{X}$ [5].

Definition 2 $\mathcal{P}_n$ is the collection of all possible types of sequences of length n [1].

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 $T(P)$:

$$T(P) = \{x^n : P_{x^n} = P\} \tag{1}$$

Theorem 1

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

Theorem 2 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}

$$Q(x^n) = 2^{-n(H(P_{x^n}) + D(P_{x^n}||Q))} \tag{3}$$

corollary if x^n is in the type class of Q, then we get $Q(x^n) = 2^{-nH(P_{x^n})}$ [5].

Theorem 3 (size of a type class $T(P)$) For any type $P \in \mathcal{P}_n$

$$|T(p)| \doteq 2^{nH(P)} \tag{4}$$

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

There are two possible ways to prove Theorem 3, one is a combinatorical proof and the other is a probabilistic.

- Proof 1 - combinatorical proof:

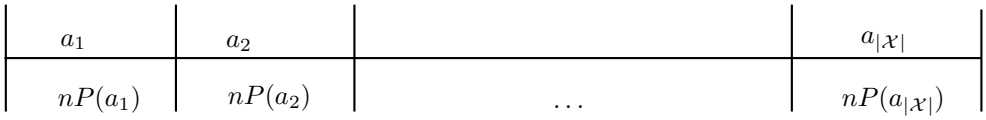

Fig. 1. Length of each a_i

$$|T(P)| = \binom{n}{nP(a_1), nP(a_2), \ldots, nP(a_{\mathcal{X}})} = \frac{n!}{(nP(a_1))!(nP(a_2))! \ldots (nP(a_{|\mathcal{X}|}))!} \tag{5}$$

Lemma 1 (Stirling's formula):

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \tag{6}$$

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

$$n! \doteq \left(\frac{n}{e}\right)^n \tag{7}$$

$$|T(P)| \doteq \frac{n^n}{(nP(a_1))^{nP(a_1)}(nP(a_2))^{nP(a_2)} \ldots (nP(a_{|\mathcal{X}|}))^{nP_{|\mathcal{X}|}}} \tag{8}$$

$$= \frac{n^n}{(n)^{nP(a_1)}(n)^{nP(a_2)} \ldots (n)^{nP_{|\mathcal{X}|}} \prod_{i=1}^{|\mathcal{X}|} P(a_i)^{nP(a_i)}} \tag{9}$$

$$= \frac{1}{\prod_{i=1}^{|\mathcal{X}|} P(a_i)^{nP(a_i)}} \tag{10}$$

Hence:

$$|T(P)| = 2^{\log|T(P)|} \doteq 2^{-n\sum_{i=1}^{|\mathcal{X}|} P(a_i)\log(P(a_i))} = 2^{nH(P)} \tag{11}$$

Example 1 Question: How many binary sequences of length n with 50% 0 and 50% 1 exists? answer: $\binom{n}{\frac{n}{2}} \doteq 2^n$

- Proof 2 - a probabilistic proof:

$$1 \geq \Pr(x^n \in T(P)) \tag{12}$$

$$1 \overset{(a)}{\geq} \sum_{x^n \in T(P)} \Pr(x^n) \tag{13}$$

$$\stackrel{(b)}{=} \sum_{x^n \in T(P)} 2^{-nH(P)} \tag{14}$$

$$= |T(P)|2^{-nH(P)} \tag{15}$$

(a) The sum of probabilities is always less equal to one.

(b) Using theorem 2.

Therefore:

$$|T(P)| \leq 2^{nH(P)} \tag{16}$$

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

Lemma 2 $P(T(P)) \geq P(T(Q))$

Proof:

Let X^n be of a type P, $P^n(T(P))$ is the probability of type class $T(P)$ and let $\hat{P} \in \mathcal{P}_n$.

It is obvious that the probability of type class $T(P)$ must be greater or equal than the probability of $T(\hat{P})$, hence:

$$P^n(T(P)) \geq P^n(T(\hat{P})), \quad \forall \hat{P} \in \mathcal{P}_n \tag{17}$$

$$\frac{P^n(T(P))}{P^n(T(\hat{P}))} \stackrel{(a)}{=} \frac{|T(P)| \prod_{a\in\mathcal{X}} P(a)^{nP(a)}}{|T(\hat{P})| \prod_{a\in\mathcal{X}} P(a)^{n\hat{P}(a)}} \tag{18}$$

$$\stackrel{(b)}{=} \frac{\binom{n}{nP(a_1),nP(a_2),\ldots,nP(a_{|\mathcal{X}|})} \prod_{a\in\mathcal{X}} P(a)^{nP(a)}}{\binom{n}{n\hat{P}(a_1),n\hat{P}(a_2),\ldots,n\hat{P}(a_{|\mathcal{X}|})} \prod_{a\in\mathcal{X}} P(a)^{n\hat{P}(a)}} \tag{19}$$

$$\stackrel{(c)}{=} \prod_{a\in\mathcal{X}} \frac{(n\hat{P}(a))!}{(nP(a))!} P(a)^{n(P(a)-\hat{P}(a))} \tag{20}$$

(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) $\frac{\binom{n}{nP(a_1),nP(a_2),\ldots,nP(a_{|\mathcal{X}|})}}{\binom{n}{n\hat{P}(a_1),n\hat{P}(a_2),\ldots,n\hat{P}(a_{|\mathcal{X}|})}} = \prod_{a\in\mathcal{X}} \frac{(n\hat{P}(a))!}{(nP(a))!}$

Using the simple bound $\frac{m!}{n!} \geq n^{m-n}$ we obtain:

$$\frac{P^n(T(P)}{P^n(T(\hat{P}))} \geq \prod_{a\in\mathcal{X}} (nP(a))^{n\hat{P}(a)-nP(a)} P(a)^{n(P(a)-\hat{P}(a))} \tag{21}$$

$$= \prod_{a\in\mathcal{X}} n^{n(\hat{P}(a)-P(a))} \tag{22}$$

$$= n^{n(\sum_{a\in\mathcal{X}} \hat{P}(a) - \sum_{a\in\mathcal{X}} P(a))} \tag{23}$$

$$= n^{n(1-1)} = 1 \tag{24}$$

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

$$\begin{aligned} 1 &= \sum_{Q \in \mathcal{P}_n} P^n(T(Q)) && (25)\\ &\leq \sum_{Q \in \mathcal{P}_n} \max_Q P^n(T(Q)) && (26)\\ &\overset{(a)}{=} \sum_{Q \in \mathcal{P}_n} P^n(T(P)) && (27)\\ &\overset{(b)}{\leq} (n+1)^{|\mathcal{X}|} P^n(T(P)) && (28)\\ &\overset{(c)}{=} (n+1)^{|\mathcal{X}|} \sum_{x^n \in T(P)} 2^{-nH(P)} && (29)\\ &= (n+1)^{|\mathcal{X}|} |T(P)| 2^{-nH(P)} && (30) \end{aligned}$$

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

(b) Using theorem 1.

(c) Using theorem 2.

Therefore our final result is:

$$\frac{2^{nH(P)}}{(n+1)^{|\mathcal{X}|}} \leq |T(P)| \leq 2^{nH(P)} \tag{31}$$

which imply that:

$$|T(P)| \doteq 2^{nH(P)} \tag{32}$$

Lets summarize our results so far:

- $|\mathcal{P}_n| \leq (n+1)^{\mathcal{X}}$
- $|T(P)| \doteq 2^{nH(P)}$
- $Q(x^n) = 2^{-n(H(P_{x^n})+D(P_{x^n}||Q)}$

Theorem 4

$$Q(T(P)) \doteq 2^{-n(D(P_{x^n}||Q)} \tag{33}$$

Proof:

$$\begin{aligned} Q(T(P)) &= \sum_{x^n \in T(P)} Q(x^n) && (34)\\ &\overset{(a)}{=} \sum_{x^n \in T(P)} 2^{-n(H(P_{x^n})+D(P_{x^n}||Q))} && (35)\\ &= |T(P)| 2^{-n(H(P_{x^n})+D(P_{x^n}||Q))} && (36)\\ &\overset{(b)}{\doteq} 2^{-nD(P_{x^n}||Q)} && (37) \end{aligned}$$

(a) Using theorem 2.

(b) using theorem 3.

Theorem 5 (Sanov theorem - Large deviation)
Let $X \sim Q$ i.i.d. and let E be a closed set of probabilities, then:

$$\lim_{n\to\infty} \log Q^n(E) = -\min_{P\in E} D(P_{x^n}||Q) = -D(P^*||Q) \tag{38}$$

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

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:

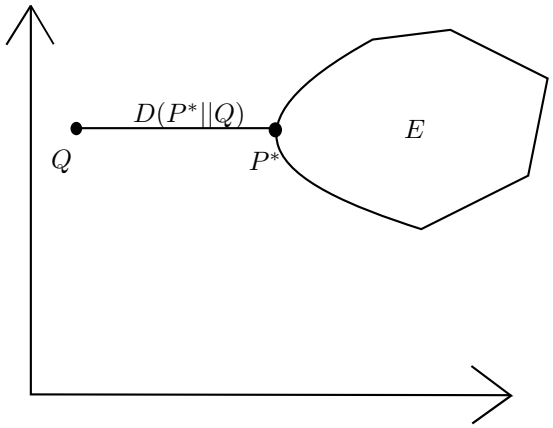

Fig. 2. Let $X \sim Q$ than P^* is the type $P \in E$ that gives the minimum to $D(P||Q)$.

$$Q^n(E) \doteq 2^{-nD(P^*||Q)} \tag{39}$$

$$P^* = \arg\min_{P\in E} D(P||Q) \tag{40}$$

Historical note: Sanovs theorem [3] was generalized by Csiszar [2] using the method of types.

Example 2 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 5:

First we will find the upper bound:

$$\begin{aligned} Q^n(E) &= \sum_{P\in E\bigcap \mathcal{P}_n} Q(T(P)) && (41)\\ &\overset{(a)}{\leq} \sum_{P\in E\bigcap \mathcal{P}_n} 2^{-nD(P||Q)} && (42)\\ &\leq \sum_{P\in E\bigcap \mathcal{P}_n} \max_{p\in E\bigcap \mathcal{P}_n} 2^{-nD(P||Q)} && (43)\\ &= \sum_{P\in E\bigcap \mathcal{P}_n} 2^{-n\min_{P\in E\bigcap \mathcal{P}_n} D(P||Q)} && (44)\\ &\overset{(b)}{\leq} (n+1)^{|\mathcal{X}|}2^{-n\min_{P\in E\bigcap \mathcal{P}_n} D(P||Q)} && (45) \end{aligned}$$

(a) According to theorem 5.

(b) Using the fact that $|E| \leq |\mathcal{P}_n|$ and theorem 1.

Now we will find the lower bound:

$$\begin{aligned} Q^n(E) &= \sum_{P\in E\bigcap \mathcal{P}_n} Q(T(P)) && (46)\\ &\overset{(a)}{\geq} Q(T(P^*)) && (47)\\ &\overset{(b)}{\geq} \frac{1}{(n+1)^{|\mathcal{X}|}}2^{-nD(P^*||Q)} && (48) \end{aligned}$$

(a) Taking only one type class is less equal of the sum of all type classes.

(b) According to theorem 5.

Using the lower bound from (51) and the upper bound from (47) we get:

$$\frac{1}{(n+1)^{|\mathcal{X}|}}2^{-nD(P^*||Q)} \leq Q^n(E) \leq (n+1)^{|\mathcal{X}|}2^{-n\min_{P\in E\bigcap \mathcal{P}_n} D(P||Q)}. \quad (49)$$

$$(50)$$

which proves that:

$$Q^n(E) \doteq 2^{-nD(P^*||Q)} \quad (51)$$

Example 3 Let X, Y be i.i.d. $X, Y \sim P_X P_Y$.

We look at a specific sequence (X^n, Y^n) with type $P_{X,Y}$, what is the probability that a sequence (x^n, y^n) wich was generated from iid $P_X P_Y$ has a joint type $P_{X,Y}$?

Answer: $Q(T(P)) \doteq 2^{-nD(P||Q)} = 2^{-nD(P_{X,Y}||P_X P_Y)} = 2^{-nI(X;Y)}$

Theorem 6 (Conditional type)

Let $W(y|x)$ be a conditional PMF.

and let:

$$P_{x^n|y^n}(a|b) = \frac{N((a,b)|x^n,y^n)}{N(b|y^n)} \quad (52)$$

$$= \frac{P_{X^n,Y^n}(a,b)}{P_{Y^n}(b)} \quad (53)$$

$$T_W(y^n) = \{x^n \in \mathcal{X}^n : P_{X^n|Y^n}(a|b) = W_{X|Y}(a|b), \forall a,b \in \mathcal{X},\mathcal{Y}\} \quad (54)$$

$$= \{x^n \in \mathcal{X}^n : P_{X^n,Y^n}(a,b) = W_{X|Y}(a|b)P_{Y^n}(b), \forall a,b \in \mathcal{X},\mathcal{Y}\} \quad (55)$$

$$H(X|Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x|y) \quad (56)$$

$$P_{X,Y}(a,b) = P_{Y^n}(b)W_{X|Y}(a|b) \quad (57)$$

Than:

$$|T_W(y^n)| \doteq 2^{nH(X|Y)} \quad (58)$$

Proof:

b_1	b_2		$b_{\|\mathcal{Y}\|}$
$nP_{Y^n}(b_1)$	$nP_{Y^n}(b_2)$	...	$nP_{Y^n}(b_{\|\mathcal{Y}\|})$

Fig. 3. Length of each b_i.

Now if we have b_1 we get:

a_1	a_2		$a_{\|\mathcal{X}\|}$
$nP_{X^n,Y^n}(a_1,b_1)$	$nP_{X^n,Y^n}(a_2,b_1)$	...	$nP_{X^n,Y^n}(a_{\|\mathcal{X}\|},b_1)$

Fig. 4. Length of each a_i given b_1.

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

$$\binom{nP_{y^n}(b_1)}{nP_{x^n,y^n}(a_1,b_1)nP_{x^n,y^n}(a_2,b_1)\ldots nP_{x^n,y^n}(a_{|\mathcal{X}|},b_1)} \doteq 2^{nH(X|y=b_1)P_{y^n}(b_1)} \tag{59}$$

$$\binom{nP_{Y^n}(b_1)}{nP_{Y^n}(b_1)P_{x^n|y^n}(a_1|b_1)nP_{Y^n}(b_1)P_{x^n|y^n}(a_2|b_1)\ldots nP_{Y^n}(b_1)P_{x^n|y^n}(a_{|\mathcal{X}|}|b_1)} \doteq 2^{nP_{Y^n}(b_1)H(X|y=b_1)} \tag{60}$$

$$|T_W(y^n)| \doteq \prod_{i=1}^{|\mathcal{Y}|} 2^{nH(X|y=b_i)P_{Y^n}(b_i)} = 2^{nH(X|Y)} \tag{61}$$

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.