Lecture 12
Conditional Types and Strongly Typical Sets

Instructor: Himanshu Tyagi
Scribe: Suprovat Ghoshal

August 31, 2015

1 Overview

In the previous lecture, we introduced the notions of conditional type and the $V$-shell of a conditional type. In this lecture, we extend the properties studied for types to that of conditional types, and show analogous lower and upper bounds on the sizes of the $V$-shells. Finally we shall motivate and introduce the notion of the strongly typical sequences and sets, which will be explored to a greater detail in later lectures.

1.1 Properties of Conditional Types and $V$-shells

Recall that conditional types are analogous notions of conditional PMF’s just as types are analogous to PMF’s. Moreover, a $V$-shell is the set of all sequences with conditional type $V$. Here are some properties that can be easily verified from definition:

\[
\sum_{y \in \mathcal{Y}} N(x, y|\mathbf{x}, y) = N(x|\mathbf{x}) \quad (1)
\]

\[
\sum_{y \in \mathcal{Y}, x \in \mathcal{X}} N(x, y|\mathbf{x}, y) = n \quad (2)
\]

\[
\sum_{y \in \mathcal{Y}} V(y|x) = \sum_{y \in \mathcal{Y}} \frac{N(x, y|x, y)}{N(x|\mathbf{x})} = 1 \quad (3)
\]

Property 1 states that the for a fixed $x \in \mathcal{X}$, counting all the occurrences of $x$ along with any $y \in \mathcal{Y}$ should give back the number of occurrences of $x$. Property
2 is just an extension of the first property i.e., summing over all possible joint occurrences of \( (x, y) \in (\mathcal{X}, \mathcal{Y}) \) gives us the length of sequence. And the third property reflects the fact that \( V(y|x) \) behaves like conditional PMF’s: for a fixed \( x \) the sum of the conditional type evaluated over all \( y \in \mathcal{Y} \) is 1.

Now, as we did for the type classes, we derive lower and upper bounds on the number of sequences in \( V \)-shell of a particular type.

**Lemma 7.** Consider \( x \) of type \( P \). Then,

\[
\frac{2^n H(V|P)}{(n + 1)|\mathcal{X}||\mathcal{Y}|} \leq |T_v(x)| \leq 2^n H(V|P)
\]

**Proof.** We prove the bounds by fixing a \( x \in \mathcal{X} \), and then looking at the subsequences \( y_{i_1}, y_{i_2}, \ldots, y_{i_t} \) that occur along with \( x \). Observe that for a fixed \( x \in \mathcal{X} \), the subsequence \( y_{i_1}, y_{i_2}, \ldots, y_{i_t} \) belongs to the \( V \)-shell of \( x \), which then defines a type in \( \mathcal{Y}^{N(x|x)} \) (see figure for illustration).

\[
\begin{array}{c}
\text{y} = 101100 \\
\text{x} = 100010
\end{array}
\]

\[
\text{\begin{tabular}{c} \dot{0}11 \cdot 0 \quad y|x = 0 \\ \dot{0}00 \cdot 0 \\ \text{\begin{tabular}{c} 1 \cdots 0 \\ 1 \cdots 1 \\
\end{tabular}} \quad x = 1
\end{tabular}}
\]

*Figure 1: Conditional Types*

Therefore, we can use Lemma 6 to lower and upper bound the number of such sequences.

\[
\frac{2^{N(x|x)H(V|x)}}{(N(x|x) + 1)|\mathcal{Y}|} \leq |T_{V|x}(x)| \leq 2^{N(x|x)H(V|x)}
\] (4)

Combining the individual lower bounds from (4) for each \( x \in \mathcal{X} \), we get
\[ |T_v(x)| = \prod_{x \in X} |T_{V(x|x)}(x)| = \geq \prod_{x \in X} \frac{2^{N(x|x)H(V|x)}}{(N(x|x) + 1)^{|X|}} \]

The last inequality follows from the fact that \( N(x|x) \leq n \) for all \( x \in X \). Similarly, we can upper bound the cardinality of the conditional type class:

\[ |T_v(x)| = \prod_{x \in X} |T_{V(x|x)}(x)| \leq \prod_{x \in X} \frac{2^{N(x|x)H(V|x)}}{(N(x|x) + 1)^{|X|}} \]

which completes the proof.

The next lemmas gives us a measure of the amount of probability mass concentrated on the individual type classes and the \( V \)-shell of type classes.

**Lemma 8.** Consider \( x \in T_P \) and a pmf \( Q \) defined on \( X \). Then,

\[ Q^n(T_P) \leq 2^{-nD(Q\|P)} \]

**Proof.** Fix an \( x \in T_P \). Then,

\[ Q^n(x) = \prod_{i=1}^n Q(x_i) = 2^{\sum_{i=1}^n \log(Q(x_i))} = 2^{\sum_{i=1}^n \log(Q(x_i))} = 2^{n \sum_{i=1}^n P(x|x) \log(Q(x_i))} = 2^{-nH(P) - nD(Q\|P)} \]
Since all the sequences belonging to a fixed type have the same probability with respect to $Q$,

\[
Q^n(\mathcal{T}_P) = |\mathcal{T}_P|Q^n(x) \leq 2^{(nH(P) - nH(P) - nD(Q\parallel P))} = 2^{-nD(Q\parallel P)}
\]

\[\square\]

**Lemma 9.** Consider a $V$-shell of $x$, $\mathcal{T}_V(x)$ and a channel $W: \mathcal{X} \mapsto \mathcal{Y}$. Then,

\[
W^n(\mathcal{T}_V(x)|x) \leq 2^{-nD(V\parallel W)}
\]

**Proof.** As in the proof of Lemma 8, fix a $y \in \mathcal{T}_V(x)$. Then,

\[
W^n(y|x) = \prod_{i \in [n]} W(y_i|x_i) = 2^{\sum_{i \in [n]} \log W(y_i|x_i)} = 2^{\sum_{y \in \mathcal{Y}} N(x,y|x,y) \log W(y|x)} = 2^{\sum_{x \in \mathcal{X}} N(x|x) \sum_{y \in \mathcal{Y}} V(y|x) \log W(y|x)}
\]

We simplify the summation term in the exponent :

\[
\sum_{x \in \mathcal{X}} N(x|x) \sum_{y \in \mathcal{Y}} V(y|x) \log W(y|x) = \sum_{x \in \mathcal{X}} P(x) \sum_{y \in \mathcal{Y}} V(y|x) \log W(y|x) = \sum_{x \in \mathcal{X}} P(x)(-D(V(\cdot|x)\parallel W(\cdot|x)) - H(v(\cdot|x))) = -n(D(V\parallel W) + H(V|P))
\]

Combining the two observations and the upper bound from Lemma 7,

\[
W^n(\mathcal{T}_V(x)|x) = \sum_{y \in \mathcal{T}_V(x)} W^n(y|x) = |\mathcal{T}_V(x)|W^n(y|x) \leq 2^{-nD(V\parallel W)}
\]

\[\square\]
2 Strongly Typical Sequences and Sets

The previous two lemmas illustrate the following fact: Even though the number of sequences in a single type class (or a conditional type class) might be exponential in $n$, it only captures a small amount of the entire probability mass with respect to pmf (or channel). Hence, in order to construct large probability sets with respect to a given pmf, we introduce the notion of strongly typical sequences.

**Definition 1 (P-(Strongly) Typical Sequences).** Let $P$ be a pmf on $\mathcal{X}$. A sequence $x \in \mathcal{X}^n$ is said to be $P$-strongly typical if the following holds for all $x \in \mathcal{X}$

$$\left| \frac{N(x|x)}{n} - P(x) \right| \leq \delta \quad \text{if } P(x) > 0$$

$$N(x|x) = 0 \quad \text{if } P(x) = 0$$

We will drop the prefix "strongly" whenever we refer to strongly typical sequences, and will explicitly use the prefix "weakly" to denote weakly typical sequences instead. The set of all $P$-typical sequences is called the $P$-typical set, and is denoted by $\mathcal{T}[P]_{\delta_n}$. The following lemma states that the $P$-typical set indeed forms a large probability set with respect to $P$ (in the limit).

**Lemma 10.** Let $P$ be a pmf defined on alphabet $\mathcal{X}$. Then,

$$\lim_{n \to \infty} P^n(\mathcal{T}[P]_{\delta_n}) = 1$$

Observe that the $\delta_n$ parameter in Lemma 10 is dependent on $n$. For convergence of probability to the limit, we need the sequence of $\delta$’s to satisfy the following conditions:

$$\lim_{n \to \infty} \delta_n = 0$$

$$\lim_{n \to \infty} \sqrt{n} \delta_n = +\infty$$

In other words, we need $\delta_n$ to approach 0 slower than $1/\sqrt{n}$. Another way to state the above result is the following: there exists $C > 0$ such that for all $n \in \mathbb{N}$

$$P^n(\mathcal{T}[P]_{\delta_n}) \geq 1 - 2^{-nC}$$

The above result can be derived using Chernoff Bound. In the next lecture, we will look at some of the properties of the typical sets, which will eventually lead to the construction of Universal Source codes.