**Information and Coding Theory** **Autumn 2017**

# Lecture 6: October 16, 2017

Lecturer: Madhur Tulsiani

## 1 The Method of Types

For this lecture, we will take $U$ to be a finite universe $|U| = r$, and use $\overline{\mathbf{x}} = (x_1, x_2, \ldots, x_n)$ to denote a sequence of $n$ elements from $U$.

**Definition 1.1** *The* type $P_{\overline{\mathbf{x}}}$ *of* $\overline{\mathbf{x}}$*, also called the* empirical distribution *of* $\overline{\mathbf{x}}$*, is a distribution* $\hat{P}$ *on* $U$*, defined as*

$$\hat{P}(a) \;:=\; \frac{|\{i : x_i = a\}|}{n} \qquad \forall a \in U\,.$$

*We use* $\mathcal{T}_n$ *to denote the set of all types coming from sequences of length* $n$*. We also use* $\mathcal{C}_P$ *to denote the set of all seuences with the type* $P$*.* $\mathcal{C}_P$ *is called the* type class *of* $P$*.*

$$\mathcal{C}_P \;:=\; \{\overline{\mathbf{x}} \in U^n \;\mid\; P_{\overline{\mathbf{x}}} = P\}\;.$$

**Exercise 1.2** *Check that* $|\mathcal{T}_n| \;=\; \binom{n+r-1}{r-1} \;\leq\; (n+1)^r$.

Next, we bound the size of a given type class in terms of the entropy of that type.

**Proposition 1.3** *For any type* $P \in \mathcal{T}_n$*, we have*

$$\frac{2^{n\cdot H(P)}}{(n+1)^r} \;\leq\; |\mathcal{C}_P| \;\leq\; 2^{n\cdot H(P)}\,.$$

**Proof:** For each $a_i \in U$, let $P(a_i) = k_i/n$. Then $|\mathcal{C}_P| = n!/(k_1!k_2!\ldots k_r!)$. We prove the lower bound by considering

$$\begin{aligned} n^n \;=\; (k_1 + k_2 + \cdots + k_r)^n \;&=\; \sum_{j_1+\cdots+j_r=n} \frac{n!}{j_1!\ldots j_m!}(k_1^{j_1}\ldots k_m^{j_r}) \\ &\leq\; \binom{n+r-1}{r-1}\cdot \max_{j_1+\cdots+j_r=n} \frac{n!}{j_1!\ldots j_r!}\cdot(k_1^{j_1}\ldots k_m^{j_r})\,, \end{aligned}$$

where each tuple $(j_1, \ldots, j_r)$ corresponds to a distinct type. We leave it as an exercise to check that the maximum term in the expression above is when $(j_1, \ldots, j_r) = (k_1, \ldots, k_r)$.

**Exercise 1.4** *Show that*

$$\frac{n!}{j_1! \ldots j_r!} \cdot (k_1^{j_1} \ldots k_r^{j_r}) \;\leq\; \frac{n!}{k_1! \ldots k_r!} \cdot (k_1^{k_1} \ldots k_r^{k_r})$$

*for all* $(j_1, \ldots, j_r)$ *such that* $j_1 + \cdots + j_r = n$. *(Hint: if* $j_s > k_s$ *for some* $s$*, then* $j_t < k_t$ *for some* $t$*.)*

Using the above, we can now prove the lower bound.

$$n^n \;\leq\; \binom{n+r-1}{r-1} \cdot \frac{n!}{k_1! \ldots k_r!} \cdot (k_1^{k_1} \ldots k_r^{k_r}) \;\leq\; (n+1)^r \cdot |\mathcal{C}_P| \cdot (k_1^{k_1} \ldots k_m^{k_m}) \,.$$

We get

$$\begin{aligned} |\mathcal{C}_P| \;&\geq\; \frac{1}{(n+1)^m} \cdot \frac{n^{k_1+k_2+\cdots+k_r}}{k_1^{k_1} \ldots k_r^{k_r}} \\ &=\; \frac{1}{(n+1)^m} \cdot \prod_{i=1}^{r} \left(\frac{n}{k_i}\right)^{k_i} \\ &=\; \frac{1}{(n+1)^m} \cdot \prod_{i=1}^{r} 2^{k_i \cdot \log(n/k_i)} \;=\; \frac{1}{(n+1)^m} \cdot 2^{n \cdot H(P)} \,. \end{aligned}$$

The proof of the upper bound is similar and left as an exercise. ■

Next, we need the observation that the probability of a sequence according to a product distribution only depends on its type.

**Proposition 1.5** *Let* $Q$ *be any distribution on* $U$ *and let* $Q^n$ *the product distribution on* $U^n$*. Let* $\overline{\mathbf{x}}, \overline{\mathbf{y}} \in U^n$ *be such that* $P_{\overline{\mathbf{x}}} = P_{\overline{\mathbf{y}}}$*. Then,* $Q^n(\overline{\mathbf{x}}) = Q^n(\overline{\mathbf{y}})$*.*

**Proof:** Let $P = P_{\overline{\mathbf{x}}} = P_{\overline{\mathbf{y}}}$. Then we have:

$$Q^n(\overline{\mathbf{x}}) \;=\; \prod_{a \in U} (Q(a))^{|\{i : x_i = 1\}|} \;=\; \prod_{a \in U} (Q(a))^{n \cdot P(a)} \;=\; Q^n(\overline{\mathbf{y}}) \,.$$

■

Now we give bounds on the probability of a certain type occurring, in terms of the KL divergence between the true distribution and the empirical distribution.

**Theorem 1.6** *For any product distribution* $Q^n$ *and type* $P$ *on* $U^n$*, we have*

$$\frac{2^{-n \cdot D(P \| Q)}}{(n+1)^r} \;\leq\; \mathop{\mathbb{P}}_{\overline{\mathbf{x}} \sim Q^n} [P_{\overline{\mathbf{x}}} = P] \;\leq\; 2^{-n \cdot D(P \| Q)} \,.$$

**Proof:** Let $\overline{\mathbf{x}}$ be of type $P_{\overline{\mathbf{x}}} = P$. For the lower bound, we note that

$$\frac{Q^n(\overline{\mathbf{x}})}{P^n(\overline{\mathbf{x}})} = \frac{\prod_{a \in U} (Q(a))^{nP(a)}}{\prod_{a \in U} (P(a))^{nP(a)}} = \prod_{a \in U} \left( \frac{Q(a)}{P(a)} \right)^{nP(a)} = 2^{n \sum_{a \in U} P(a) \log\left(\frac{Q(a)}{P(a)}\right)} = 2^{-n \cdot D(P \| Q)}$$

We also know from the previous proposition that for any $\overline{\mathbf{x}} \in \mathcal{C}_P$, we have

$$P^n(\overline{\mathbf{x}}) = \prod_{a \in U} (P(a))^{n \cdot P(a)} = 2^{-n \cdot H(P)}.$$

Finally, using Proposition 1.3, we get

$$\begin{aligned}
\underset{\overline{\mathbf{x}} \sim Q^n}{\mathbb{P}} [P_{\overline{\mathbf{x}}} = P] &= \sum_{\overline{\mathbf{x}} \in \mathcal{C}_P} Q^n(\overline{\mathbf{x}}) = \sum_{\overline{\mathbf{x}} \in \mathcal{C}_P} 2^{-n \cdot H(P)} \cdot 2^{-n \cdot D(P \| Q)} \\
&= |\mathcal{C}_P| \cdot 2^{-n \cdot H(P)} \cdot 2^{-n \cdot D(P \| Q)} \\
&\geq \frac{2^{n \cdot H(P)}}{(n+1)^r} \cdot 2^{-n \cdot H(P)} \cdot 2^{-n \cdot D(P \| Q)} \\
&= \frac{2^{-n \cdot D(P \| Q)}}{(n+1)^r}
\end{aligned}$$

The proof of the upper bound is left as an exercise. Note that It may be that $\mathrm{Supp}(Q) \subsetneq \mathrm{Supp}(P)$ i.e., $\exists a \in U : Q(a) = 0, P(a) \neq 0$. Then the $\log(1/Q(a))$ term makes $D(P\|Q)$ undefined, so thinking of $D(P\|Q)$ as $+\infty$, we get $2^{-nD(P\|Q)} = \mathrm{Prob}_{Q^n}(T_P^n) = 0$. ■

## 2 Chernoff bounds

The above counting can be used to prove the Chernoff bound. Let $U = \{0, 1\}$, and let $\overline{\mathbf{x}} = (x_1, \ldots, x_n)$ be a sequence drawn from $U^n$ according to $Q^n$, where

$$Q = \begin{cases} 0 & : \quad \text{with probability } 1/2 \\ 1 & : \quad \text{with probability } 1/2 . \end{cases}$$

We expect there to be around $n/2$ occurrences of 1 in $\overline{\mathbf{X}}$; that is, $\mathbb{E}[\sum_{i=1}^n x_i] = n/2$. It is natural to ask how much the empirical distribution is likely to deviate from $n/2$. If we set

$$P = \begin{cases} 0 & : \quad \text{with probability } 1/2 - \varepsilon \\ 1 & : \quad \text{with probability } 1/2 + \varepsilon , \end{cases}$$

then we have

$$\underset{Q^n}{\mathbb{P}} \left[ X_1 + \cdots + X_n = \frac{n}{2} + \varepsilon n \right] = \underset{\overline{\mathbf{x}} \sim Q^n}{\mathbb{P}} [P_{\overline{\mathbf{x}}} = P] \leq 2^{-n \cdot D(P \| Q)} = 2^{-c \cdot n \cdot \varepsilon^2} ,$$

by Theorem 1.6, for a constant $c$. This is sort of like Chernoff bounds, but we may want to know how likely we are to see *any* sufficiently large deviation, and not just the deviation exactly equal to $\varepsilon n$.

**Theorem 2.1 (Chernoff bound)** *For $\overline{\mathbf{X}} = (X_1, \ldots, X_n) \sim_{Q^n} U^n$ with $Q$ the uniform distribution on $U = \{0, 1\}$, we have*

$$\mathop{\mathbb{P}}_{Q^n}\left[\sum_{i=1}^{n} X_i \geq \frac{n}{2} + \varepsilon n\right] \;\leq\; (n+1) \cdot 2^{-c \cdot n \cdot \varepsilon^2} .$$

**Proof:** Let $U = \{0, 1\}$ and note that that each type class corresponds to a unique value of $x_1 + \cdots + x_n$. From the above bound, we have that for any $\eta > 0$,

$$\mathop{\mathbb{P}}_{Q^n}\left[X_1 + \cdots + X_n = \frac{n}{2} + \eta n\right] \;\leq\; 2^{-c \cdot n \cdot \eta^2} .$$

Going over all types for all $\eta \geq \varepsilon$, and noting that the number of types is at most $n + 1$, we get

$$\mathop{\mathbb{P}}_{Q^n}\left[\sum_{i=1}^{n} X_i \geq \frac{n}{2} + \varepsilon n\right] \;\leq\; (n+1) \cdot 2^{-c \cdot n \cdot \varepsilon^2} ,$$

as claimed. ■

The above idea can be generalized for product distributions over arbitrary (finite) universes to prove a general large deviation result known as Sanov's theorem.

## 3 Sanov's theorem

We generalize the Chernoff bound to understand the probability that $P_{\overline{\mathbf{x}}} \in \Pi$ for an arbitrary set $\Pi$ of distributions over $U$.

**Theorem 3.1 (Sanov)** *Let $\Pi$ be a set of distributions on $U$, and $|U| = r$. Then*

$$\mathop{\mathbb{P}}_{Q^n}\left[P_{\overline{\mathbf{x}}} \in \Pi\right] \;\leq\; (n+1)^r \cdot 2^{n \cdot \delta} ,$$

*where $\delta = \inf_{P \in \Pi} D(P \| Q)$. Moreover, if $\Pi$ is the closure of an open set and*

$$P^* \;:=\; \mathop{\arg\min}_{P \in \Pi} D(P \| Q) ,$$

*then*

$$\frac{1}{n} \cdot \log\left(\mathop{\mathbb{P}}_{\overline{\mathbf{x}} \sim Q^n}\left[P_{\overline{\mathbf{x}}} \in \Pi\right]\right) \;\;\to\;\; -D(P^* \| Q) .$$

**Proof:** For any $P \in \mathcal{T}_n$, we have by Theorem 1.6 that

$$\mathop{\mathbb{P}}_{Q^n}[\overline{\mathbf{x}} \in \mathcal{C}_P] \;\leq\; 2^{-nD(P\|Q)}\,.$$

Let $\mathcal{T}_\delta = \{P \in \mathcal{T}_n \;\mid\; D(P\|Q) \geq \delta\}$. Then, we have

$$\mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}[D(P_{\overline{\mathbf{x}}}\|Q) \geq \delta] \;=\; \sum_{P\in\mathcal{T}_\delta} 2^{-nD(P\|Q)} \;\leq\; (n+1)^r \cdot 2^{-n\delta}.$$

We now use this to prove Sanov's theorem. Take $\delta = \inf_{P\in\Pi} D(P\|Q)$, so for all $P \in \Pi$ we have $D(P\|Q) \geq \delta$. Then we get

$$\mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}[P_{\overline{\mathbf{x}}} \in \Pi] \;=\; \mathop{\mathbb{P}}_{Q^n}[P_{\overline{\mathbf{x}}} \in \Pi \cap \mathcal{T}_n] \;\leq\; \mathop{\mathbb{P}}_{Q^n}[D(P_{\overline{\mathbf{x}}}\|Q) \geq \delta] \;\leq\; (n+1)^r \cdot 2^{-n\delta}$$

as desired. Now let's prove the other direction. Since $\Pi$ is the closure of an open set and $P^* \in \Pi$, there is an $n_0$ such that we can find a sequence $\{P^{(n)}\}_{n\geq n_0}$ satisfying $P^{(n)} \to P^*$ and $P^{(n)} \in \mathcal{P}_n \cap \Pi$ for each $n$. Then we have

$$\begin{aligned}\mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}[P_{\overline{\mathbf{x}}} \in \Pi] \;=\; \mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}[P_{\overline{\mathbf{x}}} \in \Pi] \;&=\; \mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}[P_{\overline{\mathbf{x}}} \in \Pi \cap \mathcal{T}_n] \\ &\geq\; \mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}\left[P_{\overline{\mathbf{x}}} = P^{(n)}\right] \\ &\geq\; \frac{1}{(n+1)^r} \cdot 2^{-nD(P^{(n)}\|Q)}\end{aligned}$$

Thus we get

$$-D(P^{(n)}\|Q) - \frac{r\log(n+1)}{n} \;\leq\; \frac{1}{n}\log\left(\mathop{\mathbb{P}}_{\overline{\mathbf{x}}\sim Q^n}[P_{\overline{\mathbf{x}}} \in \Pi]\right) \;\leq\; -\,D(P^*\|Q) + \frac{r\log(n+1)}{n}$$

which gives

$$\frac{1}{n}\mathop{\mathbb{P}}_{Q^n}[P_{\overline{\mathbf{x}}} \in \Pi] \to -D(P^*\|Q)\,,$$

as claimed. ■

Note that the upper bound on the probability in Sanov's theorem holds for any $\Pi$. However, for the lower bound we need some conditions on $\Pi$. This is necessary since if (for example) $\Pi$ is a set of distributions such that all probabilities in all the distributions are irrational, then $\mathbb{P}_{Q^n}[P_{\overline{\mathbf{x}}} \in \Pi] = 0$. In particular, we cannot get any lower bound on this probability for such a $\Pi$.