Information and Coding Theory **Autumn 2014**

Lecture 8: October 23, 2014

Lecturer: Madhur Tulsiani Scribe: Alexander Dunlap

As in the notes from the previous lecture, $\mathbf{x} = (x_1, \ldots, x_n)$ will denote a sequence of n elements, each drawn from a finite universe U with $|U| = m$. For a sequence $\mathbf{x}$, we use $P_{\mathbf{x}}$ to denote its type (empirical distribution). We will use $\mathcal{P}_n$ to denote the set of all types for sequences of length n. Recall from the previous lecture that $|\mathcal{P}_n| \leq (n+1)^m$.

1 Sanov's theorem (continued)

Theorem 1.1 (Sanov) *Let Π be a set of distributions on U, and $m = |U|$. Let*

$$P^* = \operatorname{argmin}_{P \in \Pi} D(P \| Q).$$

Then

$$\underset{Q^n}{\mathbb{P}}\left[P_{\mathbf{x}} \in \Pi\right] \leq (n+1)^m 2^{-D(P^* \| Q)}.$$

If Π is the closure of an open set, then

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

We will need the following bound proved in the last lecture:

$$\underset{Q^n}{\mathbb{P}}\left[D(P_{\mathbf{x}} || Q) \geq \delta\right] \leq (n+1)^m \cdot 2^{-n\delta}.$$

Let's review the proof. We have

$$\underset{Q^n}{\mathbb{P}}\left[\mathbf{x} \in \mathcal{T}_P\right] \leq 2^{-nD(P \| Q)}.$$

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

$$\begin{aligned} \underset{Q^n}{\mathbb{P}}\left[D(P_{\mathbf{x}} || Q) \geq \delta\right] &= \underset{Q^n}{\mathbb{P}}\left[\bigcup_{P \in \mathcal{C}_\delta} (\mathbf{x} \in \mathcal{T}_P)\right] \\ &\leq |\mathcal{C}_\delta| \cdot 2^{-n\delta} \\ &\leq (n+1)^m \cdot 2^{-n\delta} \end{aligned}$$

We now use this to prove Sanov's theorem.

Proof: Take $\delta = D(P^*\|Q)$, so for all $P \in \Pi$ we have $D(P\|Q) \geq \delta$. Then we get

$$\begin{aligned}\mathop{\mathbb{P}}_{Q^n}[P_{\mathbf{x}} \in \Pi] &= \mathop{\mathbb{P}}_{Q^n}[P_{\mathbf{x}} \in \Pi \cap \mathcal{P}_n] \\ &\leq \mathop{\mathbb{P}}_{Q^n}[D(P_{\mathbf{x}}\|Q) \geq \delta] \\ &\leq (n+1)^m 2^{-n\delta} \\ &= (n+1)^m 2^{-nD(P^*\|Q)}\end{aligned}$$

as desired. Now let's prove the other direction. Since Π 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

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

Thus we get

$$-D(P^{(n)}\|Q) - \frac{m\log(n+1)}{n} \leq \frac{1}{n}\log \mathop{\mathbb{P}}_{Q^n}[P_{\mathbf{x}} \in \Pi] \leq -D(P^*\|Q) + \frac{m\log(n+1)}{n}$$

and

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

■

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

We now show how to compute P^* for a special family of distributions Π. Such a family is sometimes called a *linear family*.

An example: finding P^* for a linear family Π

Let $f : U \to \mathbb{R}$. Let's try to compute $\mathbb{P}_{Q^n}\left[\frac{1}{n}\sum_{i=1}^n f(x_i) \geq \alpha\right]$. Note that

$$\frac{1}{n}\sum_{i=1}^n f(x_i) = \sum_{a \in U} P_{\mathbf{x}}(a) f(a).$$

Let

$$\Pi = \left\{P : \sum_{a \in U} P(a) f(a) \geq \alpha\right\}.$$

Then the probability we want is $\mathbb{P}_{Q^n}[P_{\mathbf{x}} \in \Pi]$. We have that

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

And

$$P^* = \text{argmin}_{P \in \Pi} D(P \| Q)$$

(Assume that $\sum Q(a) f(a) < \alpha$.) Then we want to minimize $D(P\|Q)$ so that $\sum P(a) f(a) = \alpha$ (which must be true for P^*) and $\sum P(a) = 1$. The Lagrangian is $D(P\|Q) + \lambda_1(\sum P(a)f(a) - \alpha) + \lambda_2(\sum P(a) - 1)$; we want to find stationary points of this function. The resulting constraints are

$$P^*(a) = Q(a) \cdot 2^{\lambda f(a)} \cdot c2^{c\lambda_2}$$

and

$$P^*(a) = Q(a) \cdot 2^{\lambda f(a)} \cdot c'$$

where

$$c' = \frac{1}{\sum Q(a) 2^{\lambda f(a)}}.$$

λ is such that $\sum P^*(a) f(a) = \alpha$. Thus, we solve for λ in the equation

$$\frac{\sum Q(a) 2^{\lambda f(a)} f(a)}{\sum Q(a) 2^{\lambda f(a)}} = \alpha.$$

Exercise. Solve for λ if $U = \{1,2,3,4\}$, $f = \{0, 1, 1/2, 1/2\}$, $Q = \{1/2, 1/6, 1/6, 1/6\}$.

2 Hypothesis testing

Setup for hypothesis testing. Null hypothesis (H_0): true distribution is P (or, more generally, in Π). Test $T : U^n \to \{0,1\}$. 0 means that H_0 is true, and 1 means that H_0 is false. Two types of errors: type-1 (false positive: incorrectly reject H_0) has probability $\mathbb{P}_{P^n}[T(\mathbf{x}) = 1]$, type-2 (false negative: incorrectly fail to reject H_0) has probability $\mathbb{P}_{Q^n}[T(\mathbf{x}) = 0]$ if the true distribution is Q. Note that the probability of a type-2 error depends on the true distribution Q; this dependence cannot be eliminated.

The way our test will work is $T(\mathbf{x}) = 1 \iff D(P_{\mathbf{x}} \| P) \geq \delta)$.

Then we can compute the probability of a type-1 error as

$$\mathbb{P}_{P^n}[D(P_{\mathbf{x}} \| P) \geq \delta] \leq (n+1)^m 2^{-n\delta} \leq \frac{1}{n+1}$$

if we assign $\delta = \frac{(m+1)\log(n+1)}{n}$.

Then we want to find the probability of a type-2 error $\mathbb{P}_{Q^n}[T(\mathbf{x}) = 0]$. The claim is that

$$\frac{1}{n} \log \mathbb{P}_{Q^n}[T(\mathbf{x}) = 0] \to -D(P \| Q).$$

Exercise. Try proving it.