Agenda for the lecture

- Optimal rate fixed-length universal source codes
- Duality between probability assignment and variable-length universal source coding

18.1 Fixed-length universal source codes

For the case of known distribution, finding a good fixed-length source code was seen to be tantamount to finding a large probability subset with a small cardinality. Similarly, finding a good fixed-length universal source code is tantamount to finding a small cardinality set which has large probability under a class of probability distributions. In particular, a universally rate optimal source code of rate $R$ will have large probability under $P^n$ provided $H(P) < R$. An elegant tool for constructing such sets is the so-called method of types which provides a (small) collection of subsets which can be combined to form a large probability subset under any distribution. We first review the basics concepts of types and type sets.
18.1.1 Types and type sets

**Definition 18.1.** The *type* of a sequence \( x \in \mathcal{X} \), denoted \( \mathbb{P}_x \), is a pmf such that

\[
\mathbb{P}_x(a) = \frac{N(a|x)}{n}, \quad \forall a \in \mathcal{X},
\]

where \( N(a|x) = \sum_{i=1}^{n} 1(x_i = a) \) denotes the number of times \( a \) occurs in the sequence \( x = (x_1, ..., x_n) \). The set of all sequences of a given type \( Q \) is called the *type set* or *type class* \( Q \) and is denoted \( T_Q^{(n)} \). The set of all types is denoted by \( T^{(n)} \). In both the notations \( T_Q^{(n)} \) and \( T^{(n)} \), we shall omit the dependence on \( n \) when it is clear from the context.

Note that while each type \( Q \) is a pmf for \( \mathcal{X} \), every pmf does not belong to \( T \). However,

\[
\bigcup_n T^{(n)} = \mathcal{P}(\mathcal{X}).
\]

Note that each sequence of a fixed type \( Q \) has the same probability under \( P^n \) given by

\[
P^n(x) = \prod_{i=1}^{n} P(x_i)
\]

\[
= \prod_{a \in \mathcal{X}} P(a)^{N(a|x)}
\]

\[
= 2^{\sum_{a \in \mathcal{X}} N(a|x) \log P(a)}
\]

\[
= 2^{\sum_{a \in \mathcal{X}} nQ(a) \log P(a)}
\]

\[
= 2^{-n \left[ \sum_{a \in \mathcal{X}} Q(a) \log \frac{Q(a)}{P(a)} + Q(a) \log \frac{1}{Q(a)} \right]}
\]

\[
= 2^{-n[D(Q||P)+H(Q)]}.
\]

Thus, types form a partition of the set of sequences \( \mathcal{X}^n \) such that probabilities are constant in each part. The key property of this partition is that while there are exponentially many (in \( n \)) sequences in \( \mathcal{X}^n \), there are only polynomially many types.
Lemma 18.2. For a finite alphabet \( \mathcal{X} \),

\[
|T^{(n)}| \leq (n + 1)^{|\mathcal{X}| - 1}.
\]

\textit{Proof.} For a fixed \( n \), a type is described by a vector of \( |\mathcal{X}| - 1 \) numbers with coordinate denoting the number of occurrences of an element \( x \in \mathcal{X} \) (since all total number of occurrences is \( n \), there are only \( |\mathcal{X}| - 1 \) degrees of freedom). Each coordinate can take values between 0 and \( n \). Therefore, there are no more than \( (n + 1)^{|\mathcal{X}| - 1} \) such vectors. \( \Box \)

In fact, each type class except those corresponding to constant sequences has exponentially many sequences.

Lemma 18.3. For every \( Q \in T^{(n)} \)

\[
\frac{1}{(n+1)^{|\mathcal{X}| - 1}} \cdot 2^{nH(Q)} \leq |T_Q| \leq 2^{nH(Q)}.
\]

\textit{Proof.} For a type \( Q \), using (1) we get

\[
Q^n(T_Q) = |T_Q|2^{-nH(Q)}.
\]

Therefore, \( |T_Q| \leq 2^{nH(Q)} \). For the lower bound, note that

\[
1 = \sum_{P \in T} Q^n(T_P) \leq |T| \max_{P \in T} Q^n(T_P) \leq (n + 1)^{|\mathcal{X}| - 1} \max_{P \in T} Q^n(T_P),
\]

where the last step uses Lemma 18.2. Therefore, the lower bound will follow from (2) upon showing that

\[
\max_{P \in T} Q^n(T_P) = Q^n(T_Q).
\]
To see that, note

\[
\frac{Q^n(T_P)}{Q^n(T_Q)} = \frac{|T_P| \prod_{a \in \mathcal{X}} Q(a)^{nP(a)}}{|T_Q| \prod_{a \in \mathcal{X}} Q(a)^{nQ(a)}}
\]

\[
= \frac{|T_P|}{|T_Q|} \prod_{a \in \mathcal{X}} Q(a)^{nP(a) - nQ(a)}
\]

\[
= \prod_{a \in \mathcal{X}} (nQ(a))! \prod_{a \in \mathcal{X}} Q(a)^{nP(a) - nQ(a)},
\]

where in the last step we have used a simple combinatorial fact that

\[
|T_P| = \frac{n!}{\prod_{a \in \mathcal{X}} (nP(a))!}.
\]

Noting that \(k!/l! \leq k^{k-l}\) and applying this bound with \(k = nQ(a)\) and \(l =nP(a)\), we get

\[
\frac{Q^n(T_P)}{Q^n(T_Q)} \leq \prod_{a \in \mathcal{X}} (nQ(a))^{nQ(a) - nP(a)} \prod_{a \in \mathcal{X}} Q(a)^{nP(a) - nQ(a)}
\]

\[
= \prod_{a \in \mathcal{X}} n^{nQ(a) - nP(a)}
\]

\[
= n^n \left( \sum_{a \in \mathcal{X}} P(a) - \sum_{a \in \mathcal{X}} Q(a) \right)
\]

\[
= 1.
\]

Therefore, the exponent of the cardinality of \(|T_Q|\) is roughly \(nH(Q)\), i.e., for a large \(n\) we can represent the sequences of type \(Q\) using \(nH(Q)\) bits.
18.1.2 An optimal rate fixed-length universal code

To describe our fixed-length source code of rate $R$, we simply describe a subset of $X^n$ with cardinality no more than $2^{nR}$. Specifically, consider the set

$$A = \bigcup_{Q \in \mathcal{T} : H(Q) < R - \eta_n} T_Q,$$

where $\eta_n = (|X| - 1) \frac{\log(n+1)}{n}$. Note that by Lemma 18.2 and Lemma 18.3

$$|A| = \sum_{Q \in \mathcal{T} : H(Q) < R - \eta_n} |T_Q| \\
\leq \sum_{Q \in \mathcal{T} : H(Q) < R - \eta_n} 2^{n(R - \eta_n)} \\
\leq (n + 1)^{|X| - 1} \cdot 2^{-n\eta_n} \cdot 2^{nR} \\
= 2^{nR}.$$

Consider any pmf $P$ such that $H(P) < R$. For such a $P$,

$$P^n(A^c) = \sum_{Q \in \mathcal{T} : H(Q) \geq R - \eta_n} P^n(T_Q) \\
= \sum_{Q \in \mathcal{T} : H(Q) \geq R - \eta_n} |T_Q|2^{-n(D(Q\|P) + H(Q))} \\
\leq \sum_{Q \in \mathcal{T} : H(Q) \geq R - \eta_n} 2^{nH(Q)}2^{-n(D(Q\|P) + H(Q))} \\
= \sum_{Q \in \mathcal{T} : H(Q) \geq R - \eta_n} 2^{-nD(Q\|P)} \\
\leq (n + 1)^{|X| - 1} \max_{Q \in \mathcal{T} : H(Q) \geq R - \eta_n} 2^{-nD(Q\|P)} \\
= 2^{-n \left[ \min_{Q \in \mathcal{T} : H(Q) \geq R - \eta_n} D(Q\|P) - \eta_n \right]}.$$
Note that since $\eta_n \to 0$ as $n \to \infty$, for every $\delta > 0$

$$P^n(A) \leq 2^{-n\left[\min_{Q \in T : H(Q) \geq R - \delta} D(Q\|P) - \eta_n\right]},$$

for all $n$ sufficiently large. Choosing $\delta < R - H(P)$,

$$\min_{Q \in T : H(Q) \geq R - \delta} D(Q\|P) > 0.$$

Therefore, $P^n(A^c)$ goes to 0 as $n$ goes to infinity for every $P$ with $H(P) < R$. Thus, the described code is universally rate optimal. Note that our bound shows that probability of error for the code goes to 0 exponentially rapidly in $n$. In fact, it can be shown that this exponent is the largest possible for any code of rate $R$; our universal code achieves this optimal exponent without even the knowledge of the distribution!

18.2 Variable-length codes: Minmax regret and minmax redundancy

We now move to variable-length source codes. As discussed in the previous lecture, we seek uniquely decodable codes which achieve the minmax regret

$$r_n^* = \min_{\mathcal{C}} \max_P r_n(\mathcal{C}, P^n),$$

where $r_n(\mathcal{C}, P^n)$ equals the average length of the code $\mathcal{C}$ under $P^n$ minus $nH(P)$. This expression, while operationally relevant, is not very revealing. But luckily it can be replaced for all practical purposes with a simpler expression given by

$$R_n^* = \min_{Q^{(n)} \in \mathcal{P}(X^n)} \max_P D(P^n\|Q^{(n)}).$$
We show that $R^*_n \leq r^*_n \leq R^*_n + 1$. In showing this relation, we establish a simple duality between assigning probabilities to a sequence of symbols and designing uniquely decodable codes for it.

First, given a uniquely decodable code which assigns a codeword of length $l(x)$ to a sequence $x \in \mathcal{X}^n$, we use $l(x)$ to define a pmf $Q^{(n)}$ on $\mathcal{X}^n$. Specifically, let

$$Q^{(n)}(x) = \frac{2^{-l(x)}}{\sum_{x' \in \mathcal{X}^n} 2^{-l(x')}} \quad x \in \mathcal{X}^n.$$  

Then,

$$r_n(\mathcal{C}, P^n) = \sum_{x \in \mathcal{X}^n} P^n(x)l(x) - nH(P)$$

$$= \sum_{x \in \mathcal{X}^n} P^n(x) \log \frac{1}{2^{-l(x)}} - nH(P)$$

$$= \sum_{x \in \mathcal{X}^n} P^n(x) \log \frac{P^n(x)}{2^{-l(x)}}$$

$$= \sum_{x \in \mathcal{X}^n} P^n(x) \log \frac{P^n(x)}{Q^{(n)}(x)} - \log \sum_{x} 2^{-l(x)}.$$  

The last term on the right-side is positive by Kraft’s inequality (recall that we are restricting to uniquely decodable codes). Therefore, $r_n(\mathcal{C}, P^n) \geq D(P^n \| Q^{(n)})$, for every $P^n$, which in turn yields

$$\max_{P^n} r_n(\mathcal{C}, P^n) \geq \max_{P^n} D(P^n \| Q^{(n)})$$

and finally

$$\max_{P^n} r_n(\mathcal{C}, P^n) \geq \min_{Q^{(n)}} \max_{P^n} D(P^n \| Q^{(n)}).$$

Thus, $r^*_n \geq R^*_n$.

For the other direction, consider a pmf $Q$ on $\mathcal{X}^n$. Note that the lengths $l(x) = \lfloor - \log Q(x) \rfloor$ satisfy Kraft’s inequality. Therefore, there exists a prefix-free code with these
lengths. Furthermore, the average length of this code under $P^n$ is given by

$$\sum_{x \in \mathcal{X}^n} P^n(x)l(x) = \sum_{x \in \mathcal{X}^n} P^n(x)\left[-\log Q(x)\right]$$

$$= \sum_{x \in \mathcal{X}^n} P^n(x) \log \frac{1}{Q(x)} + 1$$

$$= \sum_{x \in \mathcal{X}^n} P^n(x) \log \frac{P^n(x)}{Q(x)} + H(P^n) + 1$$

$$= D(P^n \| Q) + nH(P) + 1.$$ 

Therefore, the regret $r_n(C, P^n)$ for this code under $P^n$ satisfies

$$r_n(C, P^n) \leq D(P^n \| Q) + 1, \forall P^n.$$

Proceeding as before we obtain $R_n^* \leq r_n^* + 1$.

Therefore, we shall work with the redundancy $R_n^*$ and seek pmfs $Q^{(n)}$ that attain it, keeping in mind the duality between code design and probability assignment established above. Note that we left the code design vague in the discussion above – we simply chose codeword lengths that satisfied Kraft’s inequality. One good choice of code with codewords with Shannon lengths (of $Q$) is arithmetic code. It only remains to identify a good choice of probability assignment $Q^{(n)}$. 
