1 The Method of Types

Fix a finite universe \( U \) with \( |U| = m \), and let \( x = (x_1, x_2, \ldots, x_n) \) be a sequence with each element drawn i.i.d. from some distribution \( Q \) over \( U \).

**Definition 1.1** The type \( P_x \) of \( x \), also called the empirical distribution of \( x \), is a distribution \( \hat{P} \) on \( U \). Here \( \hat{P} \) is defined by
\[
\forall a \in U : \hat{P}(a) = \frac{|\{i : x_i = a\}|}{n}.
\]

The number of possible types on \( U^n \) is \((n + 1)^m \leq (n + 1)^m\). The type class of a type \( P \) is
\[
\mathcal{T}_P^n := \{x \in U^n : P_x = P\}.
\]

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

**Proposition 1.2** For any type \( P \) on \( U^n \), we have
\[
\frac{2^n H(P)}{(n + 1)^m} \leq |\mathcal{T}_P^n| \leq 2^n H(P).
\]

**Proof:** For each \( a_i \in U \), let \( P(a_i) = k_i/n \). Then \( |\mathcal{T}_P^n| = n!/(k_1!k_2! \ldots k_m!) \). So for the upper bound:

\[
n^n = (k_1 + k_2 + \cdots + k_m)^n = \sum_{j_1 + \cdots + j_m = n} \frac{n!}{j_1! \ldots j_m!} (k_1^{j_1} \ldots k_m^{j_m}) \\
\geq \frac{n!}{k_1! \ldots k_m!} (k_1^{k_1} \ldots k_m^{k_m}) \\
n^n \geq |\mathcal{T}_P^n| \cdot (k_1^{k_1} \ldots k_m^{k_m})
\]

\[
|\mathcal{T}_P^n| \leq \frac{n^{k_1 + k_2 + \cdots + k_m}}{k_1^{k_1} \ldots k_m^{k_m}} = \left( \frac{n}{k_1^{k_1}} \right) \ldots \left( \frac{n}{k_m^{k_m}} \right)^{k_m} = 2^{k_1 \log(n/k_1) + \cdots + k_m \log(n/k_m)} = 2^{n(P(a_1) \log(1/P(a_1)) + \cdots + P(a_m) \log(1/P(a_m)))} = 2^n H(P).
\]
For the lower bound:

\[
n^n = (k_1 + k_2 + \cdots + k_m)^n \\
= \sum_{j_1 + \cdots + j_m = n} \frac{n!}{j_1! \cdots j_m!} (k_1^{j_1} \cdots k_m^{j_m}) \\
\leq \left( \frac{n + m - 1}{m - 1} \right)^n \max_{j_1 + \cdots + j_m = n} \frac{n!}{j_1! \cdots j_m!} (k_1^{j_1} \cdots k_m^{j_m}) \\
= \left( \frac{n + m - 1}{m - 1} \right)^n \frac{n!}{k_1! \cdots k_m!} (k_1^{k_1} \cdots k_m^{k_m}) \\
\leq (n + 1)^n \frac{n!}{k_1! \cdots k_m!} (k_1^{k_1} \cdots k_m^{k_m}) \\
\leq \frac{1}{(n+1)^m} \frac{n^{k_1+k_2+\cdots+k_m}}{k_1^{k_1} \cdots k_m^{k_m}} \leq \frac{n!}{k_1! \cdots k_m!} \frac{2^n H(P)}{(n+1)^m} \leq |T^n_P|.
\]

(Here (1) is left as an exercise. Hint: if \( j_r > k_r \) for some \( r \), then \( j_s < k_s \) for some \( s \).)

**Proposition 1.3** Sequences of the same type are assigned the same probability by any product distribution \( Q^n \).

**Proof**: Let \( Q^n(X_1, \ldots, X_n) = \prod_{i=1}^n Q(X_i) \) be the product distribution on \( U^n \), obtained from some distribution \( Q \). Then we have:

\[
Q^n(x) = \prod_{a \in U} (Q(a))^{|\{i : x_i = 1\}|} = \prod_{a \in U} (Q(a))^{n_{P_x}(a)}.
\]

So if \( P_x = P_y \), then \( Q^n(x) = Q^n(y) \).

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

**Theorem 1.4** For any product distribution \( Q^n \) and type \( P \) on \( U^n \), we have

\[
\frac{2^{-n D(P\|Q)}}{(n+1)^m} \leq \frac{\text{Prob}(T^n_P)}{Q^n} \leq 2^{-n D(P\|Q)}.
\]
Proof: Let \( x \) be of type \( P_x = P \). For the upper bound:

\[
Q^n(x) = \frac{\prod_{a \in U} (Q(a))^{nP(a)}}{\prod_{a \in U} (P(a))^{nP(a)}} = \prod_{a \in U} \left( \frac{Q(a)}{P(a)} \right)^{nP(a)} = 2^n \sum_{a \in U} P(a) \log \left( \frac{Q(a)}{P(a)} \right) = 2^{-nD(P\|Q)}
\]

\[
Q^n(x) = P^n(x) 2^{-nD(P\|Q)}
\]

\[
\sum_{y \in T^n_P} Q^n(y) = \sum_{y \in T^n_P} P^n(y) 2^{-nD(P\|Q)}
\]

\[
\text{Prob}_{Q^n}(T^n_P) \leq 2^{-nD(P\|Q)}.
\]

For the lower bound:

\[
\text{Prob}_{Q^n}(T^n_P) = |T^n_P| \cdot P^n(x) \cdot 2^{-nD(P\|Q)}
\]

\[
= |T^n_P| \cdot \left( \frac{k_1}{n} \right)^{k_1} \cdots \left( \frac{k_m}{n} \right)^{k_m} 2^{-nD(P\|Q)}
\]

\[
= |T^n_P| \cdot 2^{-nH(P)} \cdot 2^{-nD(P\|Q)}
\]

\[
\geq \frac{2^{nH(P)}}{(n+1)^m} \cdot 2^{-nH(P)} \cdot 2^{-nD(P\|Q)}
\]

\[
\geq \frac{2^{-nD(P\|Q)}}{(n+1)^m},
\]

using Proposition 1.2.

It may be that \( \text{Supp}(Q) \subseteq \text{Supp}(P) \), 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 \( +\infty \), \( 2^{-nD(P\|Q)} = \text{Prob}_{Q^n}(T^n_P) = 0 \). \( \blacksquare \)

2 Chernoff bounds

Take \( U = \{0,1\} \), and let \( x = (x_1, \ldots, x_n) \) be a sequence drawn from \( U^n \) according to \( Q^n \), where

\[
Q = \begin{cases} 
0 : \text{ with probability } \frac{1}{2} \\
1 : \text{ with probability } \frac{1}{2}.
\end{cases}
\]

We expect there to be around \( n/2 \) occurrences of 1 in \( X \); 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

\[
P = \begin{cases} 
0 : \text{ with probability } \frac{1}{2} - \varepsilon \\
1 : \text{ with probability } \frac{1}{2} + \varepsilon,
\end{cases}
\]
then we have
\[
\text{Prob}_{Q^n}(X_1 + \cdots + X_n = \frac{n}{2} + \epsilon n) = \text{Prob}_{Q^n}(T^n_P)
\]
\[
\leq 2^{-nD(P\|Q)}
\]
\[
= 2^{-n\epsilon^2},
\]
by Theorem 1.4, for a constant c. This gives one answer to our question, but we may want to know how likely we are to see any sufficiently large deviation.

**Theorem 2.1 (Chernoff bound)** For \( X = (X_1, \ldots, X_n) \sim Q^n U^n \) with \( Q \) the uniform distribution on \( U = \{0, 1\} \), we have
\[
\text{Prob}_{Q^n}(x : \sum_{i=1}^n x_i \geq \frac{n}{2} + \epsilon n) \leq (n + 1)^2 \cdot 2^{-nD(P^*\|Q)},
\]
where
\[
P^* = \begin{cases} 
0 & \text{with probability } \frac{1}{2} - \epsilon \\
1 & \text{with probability } \frac{1}{2} + \epsilon.
\end{cases}
\]

**Proof:**
Let \( |U| = m \). By Theorem 1.4, for any type \( P \) on \( U \), we have \( Q^n(T^n_P) \leq 2^{-nD(P\|Q)} \). For any \( \delta \):
\[
\text{Prob}_{Q^n}(x : D(P\|Q) \geq \delta) \leq \sum_{P : D(P\|Q) \geq \delta} \text{Prob}_{Q^n}(T^n_P)
\]
\[
\leq \sum_{P : D(P\|Q) \geq \delta} 2^{-nD(P\|Q)}
\]
\[
\leq \sum_{P} 2^{-n\delta}
\]
\[
\leq (n + 1)^m \cdot 2^{-n\delta}.
\]
Note that the \( (n + 1)^m \) term was obtained by counting all types on \( U^n \), not just the ones with \( D(P\|Q) \geq \delta \), so this might be improved somewhat. For the case where \( U = \{0, 1\} \), if \( P_X(1) \geq \frac{1}{2} + \epsilon \) then \( D(P_X\|Q) \geq D(P^*\|Q) := \delta \). Hence,
\[
\text{Prob}_{Q^n}(x : \sum_{i=1}^n x_i \geq \frac{n}{2} + \epsilon n) = \text{Prob}_{Q^n}(x : P_X(1) \geq \frac{1}{2} + \epsilon)
\]
\[
\leq \text{Prob}_{Q^n}(x : D(P_X\|Q) \geq \delta)
\]
\[
\leq (n + 1)^{|U|} \cdot 2^{-n\delta}
\]
\[
\leq (n + 1)^2 \cdot 2^{-nD(P^*\|Q)}.
\]
3 Sanov’s theorem (preview)

We obtained the bound
\[-D(P\|Q) - \frac{\log(n + 1)^m}{n} \leq \frac{\log(\text{Prob}_Q^n(\mathbf{X} \in T_P^n))}{n} \leq -D(P\|Q).\]

With \(m\) held constant, \(\frac{1}{n} \log(\text{Prob}_Q^n(\mathbf{x} \in T_P^n)) \to -D(P\|Q)\) as \(n \to \infty\).

**Theorem 3.1 (Sanov’s theorem)** Let \(\Pi\) be a set of distributions which is equal to the closure of its interior. Then as \(n \to \infty\),
\[
\frac{1}{n} \log \left( \text{Prob}_Q^n(\mathbf{x} \in T_P^n) \right) \to -D(P^*\|Q),
\]
where
\[
P^* = \arg\min_{P \in \Pi} D(P\|Q).
\]

We will prove this theorem in the next lecture.