16 k-to-n Codes

Let $(X_i)_{i\geq 1}$ be a sequence of identically distributed indepent discrete random variables, with probability distribution p, and range X where $|X|$ is finite. Recall that the entropy of p, defined as $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$, measures the average amount of information we have to send to encode the distribution p. Now we will look at a slightly more general situation: Suppose we want to send the outcome of k random variables, allowing a certain error ϵ. We then again ask the question how many bits we need on average, depending on the error.

Definition 16.1 A *k-to-n code* is a pair (f, ϕ),
$f : X^k \to \{0,1\}^n \qquad \phi : \{0,1\}^n \to X^k$.
The function f is called the *encoder*, ϕ is called the *decoder*.
The *error* of the pair (f, ϕ) is defined to be the probability of the event that the decoded codeword is different from the original, i.e. $P^k(\{x \in X^k : \phi(f(x)) \neq x\})$. Let $n(k, \epsilon)$ denote the smallest n such that there is a k-to-n-code of error at most ϵ.

Theorem 16.2 $\lim\limits_{k\to\infty} \frac{n(k,\epsilon)}{k} = H(p)$.

Proof. The existence of the required code (a k-to-n-code with error at most ϵ) is equivalent to the existence of a set $A \subseteq X^k$ with $P^k(A) \geq 1 - \epsilon$ and $|A| \leq 2^n$.
Let

$$s(k, \epsilon) = min\{|A| : A \subseteq X^k, P^k(A) \geq 1 - \epsilon\}.$$

We must show: $\lim\limits_{k\to\infty} \frac{1}{k} \log(s(k, \epsilon)) = H(p)$.
Let

$$B(k, \delta) = \{x \in X^k : \frac{1}{2^{k(H(p)+\delta)}} \leq P^k(x) \leq \frac{1}{2^{k(H(p)-\delta)}}\}$$

Claim: $\forall \delta > 0 \lim\limits_{k\to\infty} P^k(B(k, \delta)) = 1$.
Proof of claim:
Define new random variables $Y_i = \log(\frac{1}{p(X_i)})$ for $i \geq 1$. Then the Y_i's are i.i.d. with $\mathbf{E}[Y_i] = H(p)$. By the weak law of large numbers,

$$\lim_{k\to\infty} \mathrm{P}(|\frac{1}{k}\sum_{i=1}^{k} Y_i - H(p)| \leq \delta) = 1.$$

But

$$
\begin{aligned}
&|\frac{1}{k}\sum_{i=1}^{k} Y_i - H(p)| \leq \delta \\
\Leftrightarrow\ & H(p) - \delta \leq \frac{1}{k}\sum_{i=1}^{k} Y_i \leq H(p) + \delta \\
\Leftrightarrow\ & k(H(p) - \delta) \leq -\sum_{i=1}^{k} \log(X_i) \leq k(H(p) + \delta) \\
\Leftrightarrow\ & \frac{1}{2^{k(H(p)-\delta)}} \geq \prod_{i=1}^{k} p(X_i) \geq \frac{1}{2^{k(H(p)+\delta)}} \\
\Leftrightarrow\ & (X_1, ..., X_k) \in B(k, \delta)
\end{aligned}
$$

This proves the claim. $\square$

Since $P^k(x) \geq \frac{1}{2^{k(H(p)+\delta)}}$ $\forall x \in B(k, \delta)$, we must have that $|B(k, \delta)| \leq 2^{k(H(p)+\delta)}$. Since δ was arbitrary, we have

$$
\limsup_{k\to\infty} \frac{1}{k} \log s(k, \epsilon) \leq H(p) \tag{7}
$$

Conversely let $A \subseteq X^k$ satisfy $P^k(A) \geq 1 - \epsilon$.
By the claim, $P^k(A \cap B(k, \delta)) \geq (1-\epsilon)/2$ for k large enough. But this implies

$$
\begin{aligned}
|A| \geq |A \cap B(k, \delta)| &\geq \sum_{x \in A \cap B(k,\delta)} P^k(x) 2^{k(H(p)-\delta)} \\
&\geq 2^{k(H(p)-\delta)} p^k(A \cap B(k, \delta)) \geq \frac{1}{2}(1 - \epsilon) 2^{k(H(p)-\delta)}
\end{aligned}
$$

Thus, we conclude that

$$
\liminf_{k\to\infty} \frac{1}{k} \log s(k, \epsilon) \geq \liminf_{k\to\infty} \frac{1}{k} \log(\frac{1}{2}(1 - \epsilon) 2^{k(H(p)-\delta)}) \underset{\delta\to 0}{\longrightarrow} H(p) \tag{8}
$$

Inequalities (7) and (8) imply the theorem. $\square$

17 Graph Entropy

Suppose now we have a distinguishability relation on X given by a graph G (i.e., $u \sim v$ iff u and v are indistinguishable). We want to map elements of X^k onto codewords such that distinguishable elements are mapped to distinct codewords (where $(u_1, ..., u_k), (v_1, ..., v_k)$ are called distinguishable iff there exists an index i s.t. $u_i \sim v_i$). Let $\epsilon > 0$. The *rate of an encoding* is $log \frac{M}{k}$, where M is the number of codewords needed. Define R(G, p, k, ϵ) to be the minimum rate of encoding of a set $U \subseteq X^k$ with $p^k(U) \geq 1 - \epsilon$. That is, we are interested in encoding "most" of X^k "correctly" (i.e. distinguishable elements of U are mapped onto distinct elements), but we don't worry about the rest of X^k. The value of interest is

$$\liminf_{\epsilon \to 0^+} \liminf_{k \to \infty} R(G, p, k, \epsilon).$$

This turns out to be the *graph entropy*, introduced by Körner. However, we will use a different definition of graph entropy.

Definition 17.1 Let G be a graph, and p be a probability distribution on $V(G)$. Define the *graph entropy* of G to be:

$$H(G, p) := \min_{x \in STAB(G)} \sum_{i=1}^{n} p_i \log \frac{1}{x_i}$$

Note that the above minimum is actually attained, because we are minimizing a convex function over a subset of the nonnegative orthant, and the function goes to positive infinity at the boundary, which implies that the minimum is attained in the interior.

Observation 17.2 $H(K_n, p) = H(p)$.

Proof. Let $q = (q_1, ..., q_n)$ with $\sum_{i=1}^{n} = 1$. By concavity of the logarithm, we have

$$\sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \leq \log(\sum_{i=1}^{n} p_i \frac{q_i}{p_i})$$

But the right hand side equals 0, and so the minimum of the difference $H(K_n, p) - \sum_{i=1}^{n} p_i \log \frac{1}{q_i}$ is attained at $p_i = q_i$ for all i. □

Lemma 17.3 *Let F, G graphs on the same vertex set. Then for every (discrete) probability distribution p,*

$$H(F \cup G, p) \leq H(F, p) + H(G, p)$$

Proof. Let $H(F,p) = \sum_{i=1}^{n} p_i \log \frac{1}{a_i}$, $H(G,p) = \sum_{i=1}^{n} p_i \log \frac{1}{b_i}$, where $a \in STAB(F)$, $b \in STAB(G)$ are minimizers. Let $a \cdot b = (a_1 b_1, ..., a_n b_n)$. Then we have $a \cdot b \in STAB(F \cup G)$: If $a = \mathbf{1}_{S_1}$, $b = \mathbf{1}_{S_2}$ for S_1 stable in F, S_2 stable in G, then $a \cdot b = \mathbf{1}_{S_1 \cap S_2}$.
But this implies

$$H(F,p) + H(G,p) = \sum_{i=1}^{n} p_i (\log \frac{1}{a_i} + \log \frac{1}{b_i}) \geq H(F \cup G, p).$$

□

Corollary 17.4 $H(p) \leq H(G,p) + H(\overline{G}, p)$ *for every (discrete) probability distribution p.*

Proof. We have $H(p) = H(K_n, p)$. □