13.2 The source coding theorem

Definition 1 (Classical information source)**.** We model an information source as an infinite sequence of "symbols" $Z_1, Z_2, Z_3, \ldots$, where each symbol Z_i is a random variable taking values in some "alphabet" $\mathcal{A}_Z$ (this is just a set). We write

$$Z^{(n)} := (Z_1, Z_2, \ldots, Z_n)$$

for the string consisting of the first n symbols produced by the source. This is a RV taking values in $\mathcal{A}_Z^n$.

In the simplest kind of model, there are no correlations between the symbols.

Definition 2. Given a distribution p on a set $\mathcal{A}$, p^n denotes the distribution on $\mathcal{A}^n$ with $p^n((x_1, \ldots, x_n)) = \Pi_{i=1}^n p(x_i)$.

Definition 3. A **memoryless source** is one where the Z_j are independently and identically distributed (i.i.d.), meaning that $P_{Z^{(n)}} = P_Z^n$ for some P_Z. This situation is often denoted by $Z_i \overset{iid}{\sim} P_Z$. Here Z is an RV, also distributed according to P_Z and independent of the Z_i, which we use as a representative symbol produced by the source.

Suppose we want to compress the first n symbols of an information source to a bit string of the smallest possible length k given the requirement that the probability of error is no greater than ϵ. This is simply $k = \lceil \log\left(s_\epsilon(P_{Z^{(n)}})\right) \rceil$, where log is to base 2 here and throughout these notes, and where $\lceil x \rceil$ denotes the smallest integer no more than x. The optimal *rate* of compression is

$$\frac{1}{n} \lceil \log\left(s_\epsilon(P_{Z^{(n)}})\right) \rceil$$

bits per source symbol. We'll show that for a memoryless source, the large blocklength limit of the compression rate is given by the *entropy* of a source symbol. First, we will introduce a different way of measuring probability.

Definition 4. The **surprisal** of an event E, in bits, is $\log \frac{1}{\Pr(E)} \in [0, \infty]$.

Clearly a smaller probability means a larger surprisal. Recalling that two events A and B are independent iff

$$\Pr(A \wedge B) = \Pr(A)\Pr(B) \text{ iff } \log \frac{1}{\Pr(A \wedge B)} = \log \frac{1}{\Pr(A)} + \log \frac{1}{\Pr(B)}.$$

Definition 5. The **entropy** $H(X)$ of a random variable X is the expectation of its surprisal

$$H(X) := \mathbb{E} \log \frac{1}{P_X(X)} = S(P_X)$$

where, for any distribution p on a set $\mathcal{A}$, $S(p)$ is the entropy of the distribution p

$$S(p) = \sum_{x \in \mathrm{supp}(p)} p(x) \log \frac{1}{p(x)},$$

where $\mathrm{supp}(p)$ is the *support* of the distribution p, $\mathrm{supp}(p) := \{x \in \mathcal{A} : p(x) > 0\}$.

Definition 6. Given a distribution p on a finite set $\mathcal{A}$, we define the δ-typical set for p to be

$$T_\delta(p) := \left\{ x \in \mathcal{A} : \left| \log \frac{1}{p(x)} - S(p) \right| \leq \delta \right\}. \tag{13.1}$$

Theorem 7 (Weak law of large numbers (WLLN))**.** Given n i.i.d. real-valued random variables Y_i with finite expectation $\mathbb{E}Y_i = \mu$ and variance $\text{var}(Y_i) = \sigma^2$, for all $\delta > 0$

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i - \mu \right| \geq \delta \right) \leq \frac{\sigma^2}{n\delta^2}.$$

When $Z_i \overset{iid}{\sim} P_Z$, $P_{Z^{(n)}} = P_Z^n$ so $H(Z^{(n)}) = S(P_{Z^{(n)}}) = nS(P_Z^n) = nH(Z)$, and

$$\begin{aligned} T_{n\delta}(P_{Z^{(n)}}) &= \left\{ \underline{z} \in \mathcal{A}^n : \left| \log \frac{1}{P_{Z^{(n)}}(\underline{z})} - nH(Z) \right| > n\delta \right\} & (13.2) \\ &= \left\{ \underline{z} \in \mathcal{A}^n : 2^{-(H(Z)+\delta)n} \leq P_{Z^{(n)}}(\underline{z}) \leq 2^{-(H(Z)-\delta)n} \right\}. & (13.3) \end{aligned}$$

Proposition 8. If $Z_i \overset{iid}{\sim} P_Z$ then, for any $\delta > 0$:

1. $\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) \leq \frac{\sigma^2}{n\delta^2}$ where $\sigma^2 = \text{var}(\log \frac{1}{P_Z(Z)})$;
2. $T_{n\delta}(P_{Z^{(n)}}) \leq 2^{(H(Z)+\delta)n}$;
3. For all $\epsilon > 0$, $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \leq H(Z)$.

Proof. For the first claim, we have

$$\begin{aligned} \Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) &= \Pr\left(\left| \frac{1}{n} \log \frac{1}{P_Z^n(Z^{(n)})} - H(Z) \right| > \delta \right) & (13.4) \\ &= \Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} Y_i - H(Z) \right| > \delta \right) & (13.5) \end{aligned}$$

where Y_i is the random variable $Y_i = \log \frac{1}{P_Z(Z_i)}$. Because the Z_i are i.i.d. the Y_i are also i.i.d. and, for all i, $\mathbb{E}Y_i = H(Z)$. Therefore, the WLLN tells us that

$$\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) \leq \frac{\sigma^2}{n\delta^2}.$$

For the second claim, using (13.3), we have

$$1 \geq \Pr(Z^{(n)} \in T_{n\delta}(P_{Z^{(n)}})) = \sum_{\underline{z} \in T_{n\delta}(P_{Z^{(n)}})} P_{Z^{(n)}}(\underline{z}) \geq |T_{n\delta}(P_{Z^{(n)}})| 2^{-(H(Z)+\delta)n}. \tag{13.6}$$

The first claim tells us that, for any $\delta > 0$ and $\epsilon > 0$, there is some n_0 (which depends on ϵ and δ) such that for all $n \geq n_0$, $\Pr\left(Z^{(n)} \in T_{n\delta}(P_{Z^{(n)}})\right) \geq 1 - \epsilon$. Given this and using the second claim,

$$\frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \leq \frac{1}{n} \log |T_{n\delta}(P_{Z^{(n)}})| \leq H(Z) + \delta.$$

Therefore, for all $\delta > 0$, $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \leq H(Z)$. $\square$

Note that things are quite different when $\epsilon = 0$: $s_0(P_{Z^{(n)}}) = |\mathrm{supp}(P_{Z^{(n)}})|$.

Proposition 9. If $Z_i \overset{iid}{\sim} P_Z$ then, for any $\epsilon \in [0,1)$, and $\delta > 0$:

1. $s_\epsilon(P_{Z^{(n)}}) \geq (1 - \epsilon - \frac{\sigma^2}{n\delta^2})2^{(H(Z)-\delta)n}$ where $\sigma^2 = \mathrm{var}(\log \frac{1}{P_Z(Z)})$.

2. $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \geq H(Z)$.

Proof. For any RV X and sets $A, B \subseteq \mathcal{A}_X$ we have

$$\Pr(X \in A \cap B) \geq \Pr(X \in A) - \Pr(X \notin B). \tag{13.7}$$

Suppose $A \subseteq \mathcal{A}_Z^n$ is a set such that $\Pr(Z^{(n)} \in A) \geq 1 - \epsilon$. From the first claim in Proposition 8 we know that, for any $\delta > 0$, $\Pr(Z^{(n)} \notin T_{n\delta}(P_{Z^{(n)}})) \leq \frac{\sigma^2}{n\delta^2}$. Using (13.7) and (13.3) we find that

$$1 - \epsilon - \frac{\sigma^2}{n\delta^2} \leq \Pr(Z^{(n)} \in A \cap T_{n\delta}(P_{Z^{(n)}})) \leq |A| 2^{-(H(Z)-\delta)n},$$

and the first claim follows.

Because $\epsilon < 1$, for any $\delta > 0$, there exists n_0 (which, again, will depend on ϵ and δ) such that for all $n \geq n_0$, $\frac{\sigma^2}{n\delta^2} \leq (1-\epsilon)/2$, and by the first claim

$$s_\epsilon(P_{Z^{(n)}}) \geq \frac{1-\epsilon}{2} 2^{(H(Z)-\delta)n}$$

or, equivalently,

$$\frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \geq \frac{1}{n} \log\left(\frac{1-\epsilon}{2}\right) + H(Z) - \delta.$$

It follows that, for all $0 \leq \epsilon < 1$ and $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}}) \geq H(Z) - \delta$$

which gives us the second claim. □

If $\epsilon = 1$, then there is no limit on how much we can compress the source. Putting Propositions 8 and 9 together, and noting that $\lim_{n\to\infty} \frac{1}{n} \lceil \log s_\epsilon(P_{Z^{(n)}}) \rceil$ is equal to $\lim_{n\to\infty} \frac{1}{n} \log s_\epsilon(P_{Z^{(n)}})$, we have

Theorem 10 (Source coding theorem)**.** Given a memoryless source, $Z_i \overset{iid}{\sim} P_Z$,

$$\forall \epsilon \in (0,1) : \lim_{n\to\infty} \frac{1}{n} \lceil \log s_\epsilon(P_{Z^{(n)}}) \rceil = H(Z) = S(P_Z).$$

Note that we could express this theorem without explicit reference to random variables by saying that, for any distribution p on a finite set,

$$\forall \epsilon \in (0,1) : \lim_{n\to\infty} \frac{1}{n} \lceil \log s_\epsilon(p^n) \rceil = S(p). \tag{13.8}$$

13.3 Schumacher's quantum source coding theorem

Suppose we have a composite system $\mathsf{Q}^\mathsf{n} := \mathsf{Q}_1 \dots \mathsf{Q}_\mathsf{n}$ where each Q_i has the same dimension d_Q. We know that the state of Q^n is the product state $\rho^{\otimes n} = \rho_{\mathsf{Q}_1} \otimes \dots \otimes \rho_{\mathsf{Q}_\mathsf{n}}$, but we bear in mind that Q^n may be correlated with some reference system R and we want our compression procedure to preserve these correlations.

We want to know the minimum number k such that there is an encoding operation $\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}^\mathsf{n}}$ where K is a system of k qubits, i.e. $\dim(\mathsf{K}) = 2^k$, and decoding operation $\mathcal{D}^{\mathsf{Q}^\mathsf{n}\leftarrow\mathsf{K}}$, such that the operation $\mathcal{D}^{\mathsf{Q}^\mathsf{n}\leftarrow\mathsf{K}}\mathcal{E}^{\mathsf{K}\leftarrow\mathsf{Q}^\mathsf{n}}$ transmits any state $\rho_{\mathsf{Q}^\mathsf{n}\mathsf{R}}$ of the composite system which satisfies $\mathrm{Tr}_\mathsf{R}\rho_{\mathsf{Q}^\mathsf{n}\mathsf{R}} = \rho^{\otimes n}$ with squared fidelity $1-\epsilon$. Recalling the definitions from the previous handout, we know that k is simply

$$k = \lceil \log\left(s_\epsilon(\rho^{\otimes n})\right) \rceil.$$

We have just described the quantum analog of encoding n symbols produced by a memoryless source. Again, we ask for the large n limit of the compression *rate* $\lim_{n\to\infty} \frac{1}{n}\lceil \log\left(s_\epsilon(\rho^{\otimes n})\right) \rceil$. Now this rate is measured in "qubits per source system".

Definition 11. The log of a positive operator M is defined on the subspace $\mathrm{supp}(M)$ and, if M has an eigendecomposition $M = \sum_j \lambda_j |\alpha_j\rangle\langle\alpha_j|$, then $\log(M)$ is given by

$$\log(M) = \sum_{j:\lambda_j>0} \log(\lambda_j)|\alpha_j\rangle\langle\alpha_j|.$$

Definition 12. The **von Neumann entropy** of a density operator ρ is

$$S(\rho) = -\mathrm{Tr}\rho\log(\rho) \tag{13.9}$$

where we are treating both ρ and $\log(\rho)$ as operators on the subspace $\mathrm{supp}(\rho)$.

If ρ has an eigendecomposition $\rho = \sum_k p(k)|\alpha_k\rangle\langle\alpha_k|$ then

$$S(\rho) = -\sum_{i:p(i)>0}\sum_{j:p(j)>0} \mathrm{Tr} p(i)|i\rangle\langle i|\log(p(j))|j\rangle\langle j| = -\sum_{i\in\mathrm{supp}(p)} p(i)\log(p(i)) = S(p). \tag{13.10}$$

Furthermore, $\rho^{\otimes n}$ has an eigendecomposition

$$\rho^{\otimes n} = \sum_{k_1}\dots\sum_{k_n} p(k_1)\cdots p(k_n)|\alpha_{k_1}\rangle\langle\alpha_{k_1}| \otimes \dots \otimes |\alpha_{k_n}\rangle\langle\alpha_{k_n}|,$$

so the theorem proved in section 13.1 (in the previous handout) tells us that, for all strictly positive integers n,

$$s_\epsilon(p^n) \le s_\epsilon(\rho^{\otimes n}) \le s_{\epsilon/2}(p^n).$$

Using this fact, the classical source coding theorem (13.8), and $S(p) = S(\rho)$ we have

Theorem 13. For any ρ and all $\epsilon \in (0,1)$, $\displaystyle\lim_{n\to\infty}\frac{1}{n}\lceil \log s_\epsilon(\rho^{\otimes n})\rceil = S(\rho)$.