Introduction to Information Theory, Fall 2019

Practice problem set #1

> You do **not** have to hand in these exercises, they are for your practice only.

1. **Properties of typical sets** These properties will also be discussed in the lecture (so don't worry if you find it difficult), but it will be very helpful to think about them yourself! Let X_i be independent identically distributed random variables with a probability distribution P on a set of outcomes $\mathcal{A}_X$ and let $X^N = (X_1, \ldots, X_N)$, which has $\mathcal{A}_X^N$ as set of outcomes. Recall the definition of a *typical set* from the lecture:

$$T_{N,\varepsilon} = \left\{ x^N \in \mathcal{A}_X^N : |\frac{1}{N} \log \frac{1}{P(x_N)} - H(P)| \leqslant \varepsilon \right\}$$

   Prove the following properties:

   (a) The probability of $x^N \in T_{N,\varepsilon}$ is bounded by

   $$2^{-N(H(P)+\varepsilon)} \leqslant P(x^N) \leqslant 2^{-N(H(P)-\varepsilon)}.$$

   (b) The number of elements in $T_{n,\varepsilon}$ is bounded by

   $$|T_{N,\varepsilon}| \leqslant 2^{N(H(P)+\varepsilon)}.$$

   Hint: use (a)!

   (c) The probability of *not* being in the typical set goes to zero as N goes to infinity, that is

   $$\Pr(X^N \notin T_{N,\varepsilon}) \leqslant \frac{\sigma^2}{N\varepsilon^2} \xrightarrow[N \to \infty]{} 0$$

   where $\sigma^2 = \text{Var}(\log \frac{1}{P(X)})$.

   (d) For any $\delta > 0$ and sufficiently large N, the number of elements in $T_{N,\varepsilon}$ is bounded from below by

   $$|T_{N,\varepsilon}| \geqslant (1-\delta)2^{N(H(P)-\varepsilon)}.$$

   Hint: use 1(a) and 1(c)!

2. **Thinking about lossy compression** We use the same notation as in the previous exercise.

   (a) Explain how you can use typical sets for lossy compression. What is the compression rate?

   (b) Now suppose that the X_i have a Bernoulli distribution with $p \leqslant \frac{1}{2}$. Can you think of a sequence of subsets $S_{N,\varepsilon}$ which is such that $|S_{N,\varepsilon}| < |T_{N,\varepsilon}|$ and $\Pr(X^N \in S_{N,\varepsilon}) \geqslant \Pr(X^N \in T_{N,\varepsilon})$ (you don't have to prove it)? This shows that the typical set is not necessarily the 'optimal' set to use for lossy compression, it's just a convenient choice for computations!

   (c) Recall, or look up, the definition of δ-sufficiency and the definition of the δ-essential bit content. Explain why Shannon's source coding theorem shows that $H(P)$ is the optimal compression rate for a source with distribution P.

3. **Shannon's source coding theorem** Recall from the lecture that Shannon's source coding theorem states that for $0 < \delta < 1$ and $X_1, X_2, \ldots$ independent identically distributed random variables with distribution P

$$\lim_{N \to \infty} \frac{H_\delta(X^N)}{N} = H(P).$$

In this exercise you will be guided through the proof of this theorem. It will be discussed in the lecture, but again, it will be much easier to understand if you have already tried to work out the proof yourself!

(a) Use 1(c) to show that for all $\varepsilon > 0$ and for sufficiently large N,

$$\frac{H_\delta(X^N)}{N} \leqslant H(P) + \varepsilon$$

and conclude that

$$\lim_{N \to \infty} \frac{H_\delta(X^N)}{N} \leqslant H(P).$$

(b) Argue that if

$$\lim_{N \to \infty} \frac{H_\delta(X^N)}{N} \leqslant H(P) - \varepsilon$$

for $\varepsilon > 0$ then there exists a sequence of δ-sufficient sets $S_N \subseteq \mathcal{A}_X^N$ such that

$$|S_N| \leqslant 2^{N(H(P)-\varepsilon)}$$

for sufficiently large N.

(c) Show that

$$\Pr(X^N \in S_N \cap T_{N,\frac{\varepsilon}{2}}) + \Pr(X^N \notin T_{N,\frac{\varepsilon}{2}}) \geqslant 1 - \delta.$$

On the other hand, show by combining 1(c), 1(a) and the properties of S_N that

$$\Pr(X^N \in S_N \cap T_{N,\frac{\varepsilon}{2}}) + \Pr(X^N \notin T_{N,\frac{\varepsilon}{2}}) \xrightarrow[N \to \infty]{} 0.$$

Explain how this proves that

$$\lim_{N \to \infty} \frac{H_\delta(X^N)}{N} \geqslant H(P).$$

and we have now proven the theorem!