

Lecture 4, Friday w2, 2014-09-26

This lecture was *fast*. I expect most people have to go over the Source Coding Theorem themselves more slowly—more than once—before they get it. I know I did. The next few lectures are slower. And there's only one lecture next week. Make some time to work through the material so far.

Main lecture points:

- Information content, $\log(1/p_i) = -\log(p_i)$, as a random variable.
- Entropy: expected information content.
- Binary entropy: special case for a Bernoulli. Function of one number, p .
- Extended ensemble: block of N i.i.d. draws, each with entropy H .
- **Source coding theorem:**
 - Info. content of block is usually near the mean (law of large numbers).
 - *Typical set*: contains all the corresponding blocks we usually see.
 - Size of this set is close to 2^{NH} as $N \rightarrow \infty$.
 - Therefore can code block with NH bits, or H bits/symbol.
 - And, as we showed, no less.

Check your progress

Things you should be able to do, and questions to think about after after reviewing the lecture:

- Understand $p(x)$, $\log p(x)$ are random variables. Q1, Tutorial sheet 1.
- Be able to compute the entropy of a discrete distribution. What do you do if the probability of an outcome is zero? Why?
- The source coding theorem is for large blocks of $N \rightarrow \infty$ symbols. If we were only compressing one or a few symbols, could we do better than H bits/symbol on average? Might we have to do worse?
- Understand the proof of the source coding theorem well enough that you could explain it to another MSc student not taking this course. So they could actually follow each step and understand the implications.

Recommended reading

We have finished the ‘**week 2**’ slides, except detail I skipped for the binary entropy function, and the numerical note on $\log(\text{sum}(\exp(\cdot)))$. I will return to those things I skipped. Ask on NB if anything else is unclear, or too compressed.

MacKay Chapter 4 gives more details on the Source Coding Theorem, so I recommend reading the rest of Chapter 4. MacKay’s notation is slightly different. For example, I indexed the typical set with m , the number of standard deviations away from the mean. MacKay uses β , which absorbs the standard deviation of $h(x)$. Also, I didn’t use the symbols H_δ or S_δ . IT exam questions won’t require you to recall the detail of notation from my proof or MacKay’s. However, you do need to understand the ideas behind the proof. You may have to explain parts of the argument, or answer questions relating to it.