Information Theory — Tutorial 3

Iain Murray

October 15, 2010

1. **The meaning of unbiased:** Dr. Fran Quentin has devised an unbiased estimator, $\hat{\sigma}^2$, for the variance, σ^2, of a process. The head of the lab, Prof. Natalie Ural is actually interested in the precision $\tau = 1/\sigma^2$, while the programmer on the project only cares about the log variance, $\texttt{Lvar} = \log \sigma^2$. At a heated lab meeting Dr. Quentin is accused of incompetence: experience with her 'unbiased' estimator suggests that it tends to underestimate the variance. Why might this have happened? Is it possible that another lab member could simultaneously think that $\hat{\sigma}^2$ tends to be too big?

2. **Arithmetic coding:** Read chapter 6 of MacKay. Answer Ex 6.1: show that arithmetic coding encodes within 2 bits of the information content of a message.

3. **Non-binary encoding:** NANOCORP have just developed a new material for high-density digital storage in their patent-pending *NanoDrive*. Each region of the material can take on one of three stable physical states: $\{\texttt{r}, \texttt{d}, \texttt{n}\}$ or 'raised', 'depressed', 'normal'. The drive is amazingly reliable as long as about 50% of the regions remain in the normal state, otherwise the drive's material degrades over time.

   (a) In theory, how many bits of information can be stored on average in N *NanoDrive* regions?

   (b) Describe how arithmetic coding could be used to store files stably and efficiently on the *NanoDrive*. How many extra regions might be required to store a file compared to the theoretical capacity you derived in the previous part?

   (c) Can you think of a better way to encode files for the drive?

4. **Inference and prediction:** Solve MacKay Ex 3.1 (p47) *and* work out the probability distribution over the next outcome given the sequence that you have observed. If you needed to look at the answer or otherwise needed help, also do Ex 3.2.