Exercises for Tuesday, second hour

Mathias Winther Madsen
mathias.winther@gmail.com

January 14, 2014

Source coding (Cover and Thomas, Exercise 3.7) An information source produces produces pixels $X_1, X_2, X_3, \ldots$ with $\Pr(X_i = \texttt{WHITE}) = 0.995$ and $\Pr(X_i = \texttt{BLACK}) = 0.005$.

You decide to brute-force encode outputs from this source, 100 pixels at a time, by means of a table of equally long codewords. You include all sequences with three or fewer black pixels in the table and accept there will be an error in the remaining cases.

1. Compute or estimate the number of codewords you will need for this encoding scheme.
2. How many bits do you need in order to encode that many sequences, and how does that compare to the theoretical minimum?
3. What are your options for improving this performance, theoretically and practically?
4. Find out how likely this encoding scheme is to encounter an untabulated sequence, either by using the Markov bound or by summing the binomial probabilities directly.

Probability threshold sets (Cover and Thomas, Exercise 3.5) Let $X_1, X_2, X_3, \ldots$ be a series of independent random variables drawn from a distribution X with entropy $H(X)$. Let further $C_n(\tau)$ be the set of all high-probability sequences $(x_1, x_2, \ldots, x_n)$ for which

$$\Pr(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n) \geq 2^{-n\tau}.$$

1. What's the highest number of elements such a set can contain?
2. Sketch a graph of $\Pr C_n(\tau)$ as a function of τ for a large value of n, and for an extremely large value of n.

Random volumes (Cover and Thomas, Exercise 3.5) A sloppy company produces n-dimensional boxes with random sidelengths and volume

$$V_n \;=\; X_1 \times X_2 \times X_3 \times \ldots \times X_n.$$

The sidelengths are drawn independently from a uniform distribution on the unit interval.

A more stringent company wants to copy this product, but use n-dimensional cubes with standardized dimensions $W \times W \times W \times \cdots \times W$. They consider two different methods for choosing the constant W:

$$W = \mathsf{E}[V_n]^{1/n} = \left(\int (x_1 x_2 x_3 \cdots x_n)\, dB \right)^{1/n}$$

or

$$W = \mathsf{E}[V_n^{1/n}] = \int \left(x_1 x_2 x_3 \cdots x_n \right)^{1/n} dB.$$

In both cases, the integral is over the n-dimensional unit cube $B = [0,1]^n$.

1. Find an expression for both $\mathsf{E}[V_n]^{1/n}$ and $\mathsf{E}[V_n^{1/n}]$ in terms of n.
2. Find or estimate the limits $\lim_{n\to\infty} \mathsf{E}[V_n]^{1/n}$ and $\lim_{n\to\infty} \mathsf{E}[V_n^{1/n}]$.
3. Which of the two methods is the "correct" option?