Assignment 4

36-462, Spring 2009

Due 13 February 2009

1. BASIC PROPERTIES OF ENTROPY
   (a) Prove that the entropy of a random variable is non-negative.
   (b) If the number of values X can take on is m, prove that $H[X] \leq \log_2 m$, and the maximum is attained just when every possibility is equally likely. (*Hint*: One approach is define $p_i = \frac{1}{m} + \delta_i$, and use $\sum_i p_i = 1$.)
2. MARKOV CHAIN ESTIMATION (continued).
   (a) Write a program (`markov.mle.m`) to estimate the transition matrix of a Markov chain of order m.
   (b) Write a program (`markov.mle.bic`) which uses BIC to select a m between 1 and $m_{\max}$. Your program should output the estimated m as well as the transition matrix.
3. CHAIN ESTIMATION (applied). The file `turb.dat` on the syllabus page contains a time series of velocity measurements from a wind-tunnel, taken about thirty thousand times a second. (Thanks to Dr. Dowman Varn for sharing this data.) The units are arbitrary (based on the precision of an analog-to-digital converter), but the mean was $1.7\mathrm{m/s}$.
   (a) Discretize the measurements so that the symbol "0" represents velocities below the mean and "1" represents velocities at or above the mean. Use your `markov.mle.bic` program to fit the data. What order of chain is selected? (Hint: see `help(cut)` in R.)
   (b) Repeat the previous step, discretizing at the median.
   (c) Repeat the previous step, but use the symbol "0" to represent an *decrease* in velocity, and "1" the velocity increasing or staying the same. (Hint: see `help(diff)` in R.)
   (d) How might you chose between these three discretizations?

   Note: The `turb.dat` file has a million measurements. If this is too long for your code, feel free to truncate it, but (a) indicate that you are doing so, and (b) include results from multiple truncations of the same size, to show that your results are not flukes.