Physics 406: Homework 11

1. **Information contained in English text:** Here are the frequencies of the 26 alphabetic letters in the 1.2 million characters of Herman Melville's dreary and frankly odious novel *Moby Dick*:

letter	frequency	percentage	letter	frequency	percentage
A	75982	8.16583	N	64146	6.89381
B	16489	1.77208	O	67654	7.27082
C	22036	2.36822	P	17507	1.88149
D	37387	4.01800	Q	1510	0.16228
E	114225	12.27580	R	50781	5.45746
F	20358	2.18789	S	62704	6.73884
G	20334	2.18531	T	85998	9.24226
H	61366	6.59504	U	25967	2.79069
I	64146	6.89381	V	8429	0.90587
J	1046	0.11241	W	21617	2.32319
K	7888	0.84773	X	1199	0.12886
L	41861	4.49883	Y	16462	1.76918
M	22765	2.44657	Z	630	0.06771

Using these figures, make an estimate of the information content in bits per letter of the English language, as exemplified in this book. Given that it takes about six bits per letter to represent English text, how efficient is the language?

Extra credit: Find a computer. Download a copy of the entire text of *Moby Dick* from, for example, Project Gutenberg (`http://www.ibiblio.org/gutenberg/etext01/moby10b.txt`), which you can lawfully do for free, since the work is long out of copyright. Compress the file using your favorite text compression utility (e.g., WinZip, gzip, or StuffIt). How large a reduction is there in the size of the file? Explain the difference between the figure you find and the answer to the first part of this problem.

2. **The heat-bath algorithm:** An alternative algorithm to the Metropolis Monte Carlo algorithm is the **heat bath** algorithm. The acceptance probability for a move in this algorithm is

$$P(\mu \to \nu) = \frac{e^{-\frac{1}{2}\Delta\varepsilon/\tau}}{e^{-\frac{1}{2}\Delta\varepsilon/\tau} + e^{\frac{1}{2}\Delta\varepsilon/\tau}},$$

where $\Delta\varepsilon = \varepsilon_\nu - \varepsilon_\mu$.

(a) Show that this acceptance probability satisfies detailed balance for the Boltzmann probability distribution.

(b) The heat-bath algorithm is not usually used because it is less efficient than the Metropolis algorithm—for an efficient algorithm we want $P(\mu \to \nu)$ to be as large as possible. Draw a sketch of the acceptance probabilities of the heat-bath and Metropolis algorithms for a range of values of $\Delta\varepsilon$ around zero. For what values of $\Delta\varepsilon$ is the acceptance probability greater for the Metropolis algorithm than for the heat-bath algorithm?

(c) What is the maximum ratio by which $P(\mu \to \nu)$ for the Metropolis algorithm exceeds $P(\mu \to \nu)$ for the heat-bath algorithm, and at what energy difference $\Delta\varepsilon$ does this happen?

3. **Numerical calculation of the free energy and entropy:** The free energy is quite hard to calculate in a numerical simulation. Perhaps the simplest way to do it is the following.

   (a) Write down an expression for the internal energy in terms of the partition function. Rearrange and integrate to get an expression for the partition function. Hence derive an expression for F in terms of an integral over U. Assuming that the ground state energy is zero, make sure your expression gives the correct value for $\tau = 0$. Calculation of the free energy thus involves measuring U at a variety of different temperatures and integrating numerically to get F.

   (b) We could use this result to calculate the entropy from $\sigma = (U - F)/\tau$, but there is another alternative. Write down an expression for the heat capacity of a system in terms of a derivative of the entropy. Hence find σ as an integral over temperature. Suggest a way to measure the heat capacity using measurements of the energy of the system *without* using a derivative. (Hint: we've already seen the answer to this in a previous homework problem.)