Exercise 05

The Boltzmann distribution

Deadline: Please, hand in your solutions by **Thursday, 12 June, 2.15 p.m.**

5.1 Lagrange multiplier (3 P)

The area defined by

$$f(x,y) \quad = \quad e^{-(x^2+y^2)} \tag{1}$$

represents the shape of a mountain. The line

$$x+y \quad = \quad 1 \tag{2}$$

is the projection of a track along the mountain slope onto the x-y plane.

(a) Plot the mountain and the track along the mountain slope. (1 P)

(b) What is the highest point along the track? Solve the problem by using the method of Lagrange multipliers. (2 P)

5.2 Probability theory of alphabet pasta (9 P)

Table 1 shows the relative frequency (in percent) of letters in the English language. You can download a file with this data from the website of the lecture. In a random experiment, N letters are drawn with probabilities given by this distribution. (Imagine sitting in front of a enormous bag of alphabet noodles and picking out noodles, one at a time.) The outcome of the experiment is a sequence L of letters, e.g.

$$L = \{\text{p, r, o, b, a, b, i, l, i, t, y}\}\,.$$

The composition C of a sequence denotes the how often a given letter occurs in the sequence

$$C = \mathsf{p}_1\mathsf{r}_1\mathsf{o}_1\mathsf{b}_2\mathsf{a}_1\mathsf{i}_2\mathsf{l}_1\mathsf{t}_1\mathsf{y}_1\,.$$

Swapping letters changes the sequence, but not the composition. (e.g. $L' = \{\text{p, r, o, b, i, b, a, l, i, t, y}\}$)

(a) What is the most likely sequence of length N=5? (1 P)

(b) For an arbitrary sequence of letters, give general expressions for the number of possible sequences with the same composition, the probability of the sequence $\mathbb{P}(L)$, and the probability of the composition $\mathbb{P}(C)$. (1.5 P)

(c) Approximate $\mathbb{P}(C)$ by taking the logarithm, and applying Stirling's formula

$$\begin{aligned} \log(N!) &\approx \ln\left(\frac{N^N}{e^N}\sqrt{2\pi N}\right) = N\ln N - N + \frac{1}{2}\ln(2\pi N) \\ &\approx N\ln N - N \end{aligned} \tag{3}$$

The resulting quantity is called the entropy $S(C)$. (1.5 P)

(d) Implement a program which generates a sequence of letters according to the distribution in table 1. The program should also calculate the number of possible sequences with the same composition, the probability of the sequence $\mathbb{P}(L)$, the probability of the composition $\mathbb{P}(C)$, the logarithm of $\mathbb{P}(C)$. and the entropy $S(C)$. Report the results for two random sequences. Hand in the program. (5 P)

Letter	Frequency (%)	Letter	Frequency (%)	Letter	Frequency (%)	Letter	Frequency (%)
a	8.167	h	6.094	o	7.507	v	0.978
b	1.492	i	6.966	p	1.929	w	2.360
c	2.782	j	0.153	q	0.095	x	0.150
d	4.253	k	0.772	r	5.987	y	1.974
e	12.702	l	4.025	s	6.327	z	0.074
f	2.228	m	2.406	t	9.056		
g	2.015	n	6.749	u	2.758		

Table 1: Relative frequency of letters in the English language.

5.3 Thermodynamics of a two-state folder (7 P)

Heat denatures proteins, i.e. above a certain temperature proteins lose their three-dimensional fold. This is one of the reasons why a high fever is dangerous. In this exercise, you will study the heat denaturation of proteins using a one-dimensional model potential energy function for a two-state folder

$$E(x) \quad = \quad -10^{4.05}\left(e^{-2\cdot(0.7x-2)^2} + e^{-0.0003\cdot(x-25)^2}\right) \tag{4}$$

A two-state folder is a protein which fold and unfolds via a single transition state. The folded state is represented by a deep and narrow minimum, the unfolded state is represented by a shallow and broad minimum. The Boltzmann distribution of this potential energy function is

$$p(x) \quad = \quad \frac{\exp\left(-\frac{1}{RT}E(x)\right)}{\int \exp\left(-\frac{1}{RT}E(x)\right)dx} \tag{5}$$

(Note that the energy is defined in J/mol and consequently $\beta = \frac{1}{RT}$, where $R = 8.314$ J/(K mol) is the ideal gas constant. In this way numerical instabilities can be avoided.)

(a) Plot the potential and the Boltzmann distribution at body temperature form $x = 0$ to $x = 60$. (1 P)

(b) Determine the transition state. (1 P)

(c) Calculate the probabilities of the folded and the unfolded state from $T = 270$ K to $T = 370$ K (steps of 10 K are sufficient) and plot them as a function of temperature. (2 P)

(d) Explain the changes in the population. (1 P)

The equilibrium constant between folded and unfolded state can be expressed in terms of the state populations

$$K = \frac{p_{\text{unfold}}}{p_{\text{fold}}}\,. \tag{6}$$

Furthermore, the following two equations for the free energy difference ΔG hold

$$\Delta G \quad = \quad -RT\ln K \tag{7}$$

$$\Delta G \quad = \quad \Delta H - T\Delta S \tag{8}$$

(e) Plot $-RT\ln K$ as a function of T and determine the entropy difference between the folded and the unfolded state. (2 P)

(f) Determine the melting temperature of the protein. (The melting temperature is the temperature at which $\Delta G = 0$.) (1 P)