ICFP M2 - Statistical physics 2 – Exam

Guilhem Semerjian

March 30th, 2018

The exam is made of two parts. The first one is a series of short independent questions to check your knowledge and understanding of the contents of some of the lectures, the second one is a longer problem with partially independent subparts.

No document, calculator nor phone is allowed.

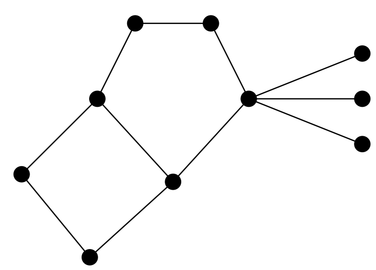
You can write your answers in English or French.

1 Questions on the lectures

1. Consider a random variable X, Gaussian distributed with average 0 and variance $\mathbb{E}[X^2] = b$. We use X to form another random variable, $Z_N = e^{Na+\sqrt{N}X}$, where a is a fixed constant and N mimicks the size of a system. Compute
$$\frac{1}{N} \ln \mathbb{E}[Z_N] \qquad \text{and} \qquad \frac{1}{N} \mathbb{E}[\ln Z_N] \ . \tag{1}$$
Give a name to these two ways of taking the average, and comment on their possible difference.

2. A sequence of 1000 symbols A, B, C produced by a random source with probabilities $p_A = \frac{1}{2}$, $p_B = p_C = \frac{1}{4}$, is compressed into a string of 0 and 1's. How long do you expect the compressed string to be if you use an efficient algorithm ? Propose an encoding of A, B, C into binary symbols that achieve this optimal compression rate.

3. What is the minimal number of colors necessary to color the vertices of the graph drawn below without creating monochromatic edges ?



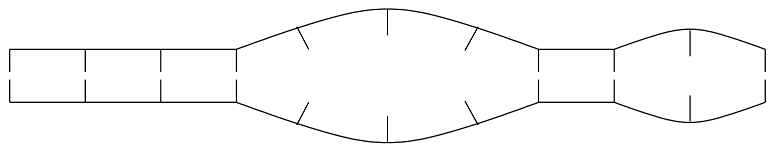
Explain why, and give an example of solution.

4. Consider a positive random variable X with distribution function $F_X(x) = \mathbb{P}[X \le x] = 1 - e^{-x^\beta}$ for $x \ge 0$, with $\beta > 0$ a fixed exponent, and X_1, X_2,. . .independent identically distributed random variables with the same law as X. We denote
$$\widehat{M}_n = \frac{\max(X_1, \dots, X_n) - a_n}{b_n} \tag{2}$$
the rescaled maximum of n such random variables. Write down the distribution function $F_{\widehat{M}_n}$ in terms of F_X. How should one choose a_n and b_n for $\widehat{M}_n$ to converge in distribution as $n \to \infty$? How is called the limit distribution of $\widehat{M}_n$?

2 The denaturation transition of DNA

The DNA is a polymeric molecule made of two complementary strands that, at ambient temperature, are tied together. Upon heating it undergoes a so-called denaturation transition (that can be observed experimentally with UV spectroscopy techniques) : the thermal excitations overcome the binding energy of complementary bases, and some of them detach from each other ; the portions of the molecule which are not tied together can fluctuate in space, which increases its entropy.

We will study this phenomenon with a simplified modelization, due to Poland and Scheraga, in which one describes a microscopic configuration of the molecule by specifying which complementary bases of the molecule are in contact and which are detached. In a molecule of length (total number of bases) L, the configuration is thus specified by k, the number of bases in contact, and $1 \leq \ell_1 < \ell_2 < \cdots < \ell_k \leq L$ the position of these bases along the chain. For instance the configuration on the figure below corresponds to $L = 11$, $k = 7$, $(\ell_1, \dots, \ell_7) = (1, 2, 3, 4, 8, 9, 11)$.



For simplicity we assume that the first and last pair of bases are always attached, i.e. the boundary conditions are such that $k \geq 2$, $\ell_1 = 1$, $\ell_k = L$.

Two bases in contact at position ℓ contribute to the energy of the molecule by an additive term ε_ℓ ; we assume $\varepsilon_\ell \leq 0$, because the pairing interaction is attractive. The entropic effect of the detached loop between two contacts at positions ℓ_i and ℓ_{i+1} is modelized by a factor $A(\ell_{i+1} - \ell_i)$, with $A(\ell)$ a positive function of the length ℓ of the detached segment. The probability of a configuration is thus

$$\frac{1}{Z_L(\beta, \underline{\varepsilon})} A(\ell_2 - \ell_1) A(\ell_3 - \ell_2) \dots A(\ell_k - \ell_{k-1}) e^{-\beta(\varepsilon_{\ell_1} + \varepsilon_{\ell_2} + \cdots + \varepsilon_{\ell_k})} , \tag{3}$$

where the partition function normalizing the probability is

$$Z_L(\beta, \underline{\varepsilon}) = \sum_{k=2}^{L} \sum_{1 < \ell_2 < \cdots < \ell_{k-1} < L} A(\ell_2 - \ell_1) A(\ell_3 - \ell_2) \dots A(\ell_k - \ell_{k-1}) e^{-\beta(\varepsilon_{\ell_1} + \varepsilon_{\ell_2} + \cdots + \varepsilon_{\ell_k})} \tag{4}$$

with the boundary conditions $\ell_1 = 1$, $\ell_k = L$, and $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_L)$ denoting the binding energies along the chain.

We shall assume that $A(\ell) > 0$ for all $\ell \geq 1$, that it has the following behavior at large distances :

$$A(\ell) \sim A_0 \frac{\Omega^\ell}{\ell^c} , \tag{5}$$

with $A_0 > 0$, $\Omega > 1$ and $c > 0$ some constants, and that at the minimal distance $0 < A(1) < \Omega$.

2.1 The homogeneous case

In a real DNA molecule the binding energies of the complementary bases A-T and C-G are not equal, hence $\underline{\varepsilon}$ is a quenched disorder that depends on the genomic sequence. We will first study an homogeneous molecule with $\varepsilon_\ell = \varepsilon$ all along the chain, and denote $Z_L^{\rm h}(\beta, \varepsilon)$ the partition function in this case. We recall that $\varepsilon < 0$ due to the attractive interaction between the two strands. Our goal is to compute and characterize the free-energy density

$$f^{\rm h}(\beta, \varepsilon) = -\frac{1}{\beta} \lim_{L \to \infty} \frac{1}{L} \ln Z_L^{\rm h}(\beta, \varepsilon) . \tag{6}$$

1. Before computing $f^{\rm h}(\beta, \varepsilon)$ we shall derive simple bounds on it to gain some intuition on the physics of this model.
   (a) Compute the contribution to $Z_L^{\rm h}(\beta, \varepsilon)$ of the configuration where all the bases are in contact. Deduce from it an upperbound to $f^{\rm h}(\beta, \varepsilon)$.
   (b) Do the same with the configuration where all the bases (except those in $\ell = 1$ and $\ell = L$) are detached.
   (c) Draw these two bounds as a function of the temperature $T = 1/\beta$. Comment on the low and high temperature limits.

2. Give an interpretation of $\frac{\partial}{\partial\varepsilon} \ln Z_L^{\rm h}(\beta,\varepsilon)$, and justify that

$$\theta(\beta,\varepsilon) = \frac{\partial}{\partial\varepsilon} f^{\rm h}(\beta,\varepsilon) \tag{7}$$

plays the role of an order parameter for the denaturation transition.

3. To compute the free-energy at all temperatures it is easier to use the grand canonical formalism in which the length of the molecules is allowed to fluctuate, with a conjugated control parameter $x \geq 0$ called fugacity. The grand partition function $\mathcal{Z}^{\rm h}(x,\beta,\varepsilon)$ is defined as

$$\mathcal{Z}^{\rm h}(x,\beta,\varepsilon) = \sum_{L=2}^{\infty} Z_L^{\rm h}(\beta,\varepsilon)\, x^L \ , \tag{8}$$

and the critical fugacity $x_*(\beta,\varepsilon)$ as

$$x_*(\beta,\varepsilon) = \sup\{\, x \,:\, \mathcal{Z}^{\rm h}(x,\beta,\varepsilon) < \infty\} \ . \tag{9}$$

   (a) Express the free-energy density $f^{\rm h}(\beta,\varepsilon)$ in terms of the critical fugacity $x_*(\beta,\varepsilon)$.

   (b) Compute $\mathcal{Z}^{\rm h}(x,\beta,\varepsilon)$ in terms of the generating function associated to $A(\ell)$, i.e.

$$\widehat{A}(x) = \sum_{\ell=1}^{\infty} A(\ell)\, x^\ell \ , \tag{10}$$

   and conclude that

$$f^{\rm h}(\beta,\varepsilon) = \frac{1}{\beta} \ln\left(\sup\{\, x \,:\, e^{-\beta\varepsilon}\widehat{A}(x) < 1\}\right) \ . \tag{11}$$

4. Study carefully the function $\widehat{A}(x)$ for $x \geq 0$, recalling the assumptions on $A(\ell)$ made in equation (5). You should in particular consider the following questions :

   (a) What is the value of $\widehat{A}(0)$?

   (b) How does $\widehat{A}(x)$ behaves for $x \to 0$?

   (c) Is the function $\widehat{A}(x)$ monotonous for $x \geq 0$?

   (d) What is the radius of convergence R of the series (10), such that $\widehat{A}(x)$ is finite for $0 \leq x < R$, infinite for $x > R$?

   (e) Is $\widehat{A}(R)$ finite ? (The answer to this question depends on c)

   (f) Is $\widehat{A}'(R)$ finite ? (idem)

   Your answers to the last two questions should lead you to distinguish three regimes, $c < c_1$, $c_1 < c < c_2$, and $c > c_2$; specify the values of c_1 and c_2. Draw three sketches of $\widehat{A}(x)$, one for each of these regimes.

5. Consider first the case $c < c_1$, and show that there is no phase transition in this case, in the sense that $f^{\rm h}(\beta,\varepsilon)$ has no singularity as a function of the temperature.

6. We shall now consider the case $c > c_1$, and make the additional assumption $\widehat{A}(R) < 1$ (view $\widehat{A}(R)$ as a known number, do not try to compute it). Show the existence of a phase transition at a temperature $T_{\rm c}$ that you will specify, and give the value of $f^{\rm h}(\beta,\varepsilon)$ in the high temperature phase. What is the value of the order parameter $\theta(\beta,\varepsilon)$ in this phase ? What are the typical configurations of the molecule ?

7. We still assume $c > c_1$ and $\widehat{A}(R) < 1$, and consider now the low temperature phase $T < T_{\rm c}$, and in particular the limit $T \to T_{\rm c}^-$.

   (a) Express $f^{\rm h}(\beta,\varepsilon)$ in the low temperature phase in terms of the function B reciprocal of $\widehat{A}$, i.e. such that $B(\widehat{A}(x)) = \widehat{A}(B(x)) = x$.

   (b) Draw the shape of B in the case $c > c_2$, deduce the qualitative behavior of the free-energy for $T \to T_{\rm c}^-$, and conclude that the transition is first order in this case.

   (c) Consider finally the situation $c_1 < c < c_2$; proove (or admit) that

$$\widehat{A}(x) \sim \widehat{A}(R) - C\,(R-x)^{c-1} \tag{12}$$

   in the limit $x \to R^-$, where C is a strictly positive constant. Deduce the behavior of B close to $\widehat{A}(R)$, and conclude that the transition is now of second order, with a singular behavior of $f^{\rm h}$ of the form $(T_{\rm c} - T)^{2-\alpha}$ in the limit $T \to T_{\rm c}^-$, with an exponent

$$\alpha = \frac{2c-3}{c-1} \ . \tag{13}$$

8. The conclusions of the previous questions seem at first sight contradictory with the well-known statement according to which there cannot be a finite temperature phase transition in one dimension. Why is this situation only apparently paradoxical ?

2.2 The disordered case

We suppose now that the binding energies ε_ℓ along the chain are independent identically distributed negative random variables, and denote $\mathbb{E}$ the expectation with respect to this disorder. We will also assume $c_1 < c < c_2$, in such a way that the homogeneous model has a second order phase transition.

The properties of a typical molecule are described by the free-energy

$$f_{\rm q}(\beta) = -\frac{1}{\beta} \lim_{L\to\infty} \frac{1}{L} \mathbb{E}[\ln Z_L(\beta, \underline{\varepsilon})] \ ; \tag{14}$$

we will denote $T_{\rm c,q}$ the temperature at which this function exhibits a phase transition (non-analyticity), if it does so. The computation of $f_{\rm q}(\beta)$ is in general impossible, but one can obtain some informations on it, as we shall see.

1. Show that $\mathbb{E}[Z_L(\beta, \underline{\varepsilon})] = Z_L^{\rm h}(\beta, \varepsilon_{\rm eff})$, for a well-chosen effective binding energy $\varepsilon_{\rm eff}$ (that implicitly depends on β).
2. Deduce from this result a lowerbound on $f_{\rm q}$ in terms of the free energy $f^{\rm h}$ of the homogeneous model. We shall denote $T_{\rm c,a}$ the temperature at which this bound has a singularity. Give an expression linking $T_{\rm c,a}$ and the constant $\widehat{A}(R)$.
3. Reconsider your answer to the question (1b) of Section 2.1, and show that the same upperbound is valid for $f_{\rm q}$.
4. Conclude that in the high temperature phase $f_{\rm q}$ has the same value as $f^{\rm h}$; why is this a posteriori physically obvious ?
5. Show that $T_{\rm c,q} \le T_{\rm c,a}$.

2.3 Harris criterion

The study of the homogeneous version of the model has shown that for $c_1 < c < c_2$ the system undergoes a second-order phase transition at finite temperature, with an exponent α that can be positive or negative depending on $c < 3/2$ or $c > 3/2$ (see equation (13)). According to the Harris criterion seen in the lectures, an infinitesimal disorder is relevant (modifies the properties of the system with respect to the homogeneous case) when $\alpha > 0$, irrelevant if $\alpha < 0$. We shall revisit this criterion with a slightly different method here.

To describe an infinitesimal disorder we write $e^{-\beta\varepsilon_\ell} = e^{-\beta\varepsilon}(1+u_\ell)$, where ε is deterministic and the u_ℓ are independent identically distributed random variables with zero average and very small variance.

1. Write explicitly $Z_L(\beta, \underline{\varepsilon})$ in terms of the variables u_ℓ.
2. As the variables u are very small we shall only keep the terms where they appear linearly and write

$$Z_L(\beta, \underline{\varepsilon}) = Z_L^{\rm h}(\beta, \varepsilon) \left(1 + \sum_{\ell=1}^{L} u_\ell \, p_\ell(L, \beta, \varepsilon)\right) + O(u^2) \ , \tag{15}$$

   where $p_\ell(L, \beta, \varepsilon)$ is independent of the disorder. Give the explicit form of this quantity, and interpret it as a probability of an event in the homogeneous model.
3. Take the logarithm of this expression, expand it for small disorder, and take the average to obtain, neglecting higher order terms :

$$-\frac{1}{\beta L} \mathbb{E} \ln[Z_L(\beta, \underline{\varepsilon})] = -\frac{1}{\beta L} \ln Z_L^{\rm h}(\beta, \varepsilon) + \frac{1}{2\beta} \mathbb{E}[u^2] \left(\frac{1}{L} \sum_{\ell=1}^{L} (p_\ell(L, \beta, \varepsilon))^2 \right) \ . \tag{16}$$

4. Argue that the last parenthesis tends to $\theta(\beta, \varepsilon)^2$ in the large size limit, which yields

$$f_{\rm q}(\beta) = f^{\rm h}(\beta, \varepsilon) + \frac{1}{2\beta} \mathbb{E}[u^2] \theta(\beta, \varepsilon)^2 \ . \tag{17}$$

   Discuss the scaling of the two terms in the right hand side in the limit $T \to T_{\rm c}^-$, and recover in this way the Harris criterion.

To learn more about this problem you can consult the paper B. Derrida, M. Retaux, *The depinning transition in presence of disorder : a toy model*, J. Stat. Phys **156**, 268 (2014), and references therein.