Mid-term Exam: Statistical Mechanics 2018/19, ICFP Master (first year)

Werner Krauth, Victor Dagard
(Dated: 12 November 2018)

Introduction

General information:

- Starting time: 12 November 2018, 10:45 AM.
- Finishing time: 12 November 2018, 12:45 PM.
- Closed-books exam (no books, scripts, calculators, smart phones, computers, etc.).
- Use only paper provided by ENS.
- Do not hesitate to ask questions.
- Do not forget to write your name onto the cover sheet.
- Please transfer your answers from the green scratch paper (brouillon) to the white exam paper.
- Please leave the scratch paper at your desk.
- Do not forget to sign the register ("feuille d'émargement").

I. EXPONENTIAL AND BERNOULLI DISTRIBUTIONS

In this problem, we study inequalities for probability distributions. We will illustrate the approximate nature of the Chebychev inequality, and also the importance of the Hoeffding inequality, in direct relation to the first part of our lecture course.

A. Useful formulas

Before starting, we recall the following (μ_X: mean value, Var_X: Variance, σ_X: standard deviation):

Chebychev The Chebychev inequality considers a random variable X. It states:

$$P(|X-\mu_X| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2} \tag{1}$$

Hoeffding Hoeffding's inequality considers random variables $X_i, i = 1, \ldots, n$ with zero expectation and $a_i \leq X_i < b_i$. For every $t > 0$, Hoeffding's inequality states:

$$P(\sum_{i=1}^{n} X_i \geq \epsilon) \leq \exp(-t\epsilon) \prod_{i=1}^{n} \exp\left[t^2(b_i - a_i)^2/8\right]. \tag{2}$$

Exponential distribution The distribution Exponential(β) is defined through the probability density $\pi(x) = 1/\beta \exp(-x/\beta)$ (for $x \geq 0$).

Bernoulli distribution The distribution Bernoulli(θ) takes on the value 1 with probability θ and the value 0 with probability $1-\theta$.

B. Problems

1. Consider a random variable $X \sim$ Exponential(β). What is its mean value μ_X, its variance Var_X, and its standard deviation σ_X?

   - *We first note that the distribution is normalized, because of*

$$\frac{1}{\beta}\int_0^\infty dx\ \exp(-x/\beta) = 1$$

   - *The mean value of this distibution is $\mu_X = \beta$, because of*

$$\frac{1}{\beta}\int_0^\infty dx\ x\exp(-x/\beta) = \beta$$

   - *The variance of this distribution is $Var_X = \beta^2$ because $Var_X = \langle x^2 \rangle - \mu_X^2$ and*

$$\frac{1}{\beta}\int_0^\infty dx\ x^2\exp(-x/\beta) = 2\beta^2$$

   - *The standard distribution of this distribution is $\sigma_X = \sqrt{Var_X} = \beta$.*

2. What is the tail probability $P(|X-\mu_X| \geq k\sigma_X)$ (with $k \geq 1$) for $X \sim$ Exponential(β)? Compare this tail probability to the bound you obtain from the Chebychev inequality.

   The tail probability is $\exp(-k)$. Chebychev gives $1/k^2$. The two functins are the same for $k = 2W(1/2) = 0.703467..$, where W is the Product Log (the Lambert W function) (but our derivation does not apply to this case). For all $k \geq 1$, the Chebychev bound is not tight.

3. Find a simplified expression of Hoeffding's inequality, for the special case where the random variables X_i are i.i.d Bernoulli distributed, and for the sample mean $\overline{X}_n = 1/n \sum_{i=1}^{n} X_i$, that is, formulate Hoeffding's inequality for $P(|\overline{X}_n - \theta| \geq \epsilon)$ (for any $\epsilon > 0$) for Bernoulli-distributed random variables. Hint: Remember that Hoeffding's inequality, in its original formulation, is for random variables of zero mean.

   One finds $P(|\overline{X}_n - \theta| \geq \epsilon) \leq 2\mathrm{e}^{-2n\epsilon^2}$

4. For the Bernoulli variables $X_1, \dots, X_n$, as above, compare the bound from Hoeffding's inequality to the bound one obtains from Chebychev's inequality. Under which condition, precisely, is Hoeffding's inequality stricter than Chebychev's inequality (hint: don't say "large n", be more precise!).

   We note that Hoeffding inequality is for the sum of random variables, not for the mean value. However, the $1/n$ *is innocuous:*

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \frac{1}{n}\epsilon\right) \leq \exp\left(-t\frac{1}{n}\epsilon\right)\prod_{i=1}^{n}\exp\left[t^2(b_i - a_i)^2/8\right]. \tag{3}$$

   leads for $\epsilon/n \to \epsilon$ *and Bernoulli variables, for which* $b_i - a_i = 1$, *to the probability:*

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \epsilon\right) \leq \exp(-t\epsilon)\exp\left[nt^2/8\right]. \tag{4}$$

   The rhs is $\exp\left(-t\epsilon + nt^2/8\right)$. *We find the best value of* t *by deriving. This gives* $(n/4)t = \epsilon$ *and for the r.h.s. exponentiall the term* $-2\epsilon^2/n$. *Therefore, we find that the tail probability from Hoeffding's inequality is* $\exp\left(-2n\epsilon^2\right)$, *that is, exponential in* $n\epsilon^2$. *Chebychev gives* $\theta(1-\theta)/(n\epsilon^2)$. *We need* $n \sim 1/\epsilon^2$ *in both cases, but for* $n \gg 1/\epsilon^2$, *Hoeffding is much sharper.*

5. We have now seen two examples, where the Chebychev inequality gave a much worse bound for the tail probability than either the exact tail probability itself, or than sharper inequalities. Does this mean that the Chebychev inequality can be replaced by a better inequality without making additional assumptions? If not, give a counter-example.

   No, it is easy to show that the Chebychev inequality is sharp, if no additional assumption is made. An example is the double δ*-peak distribution:*

$$\pi(x) = \frac{1}{2}\delta(-1) + \frac{1}{2}\delta(1) \tag{5}$$

   whose variance equal 1. *Chebychev's inequality yields*

$$P(|X| > \epsilon) \leq \frac{1}{\epsilon^2} \tag{6}$$

   In the limit $\epsilon \to 1^-$, *where the above probability is* 1, *this becomes sharp.*

II. ICE-TYPE MODELS IN STATISTICAL MECHANICS, TRANSFER MATRIX

We will consider crystals with hydrogen bonding. The most familiar example is ice, where the oxygen atoms form a lattice of coordination number four (i.e each oxygen has 4 oxygen neighbors), and between each adjacent pair of atoms is an hydrogen ion. Each ion is located near one or other end of the bond in which it lies. Slater (1941) proposed (on the basis of local electric neutrality) that the ions should satisfy the ice rule:

Ice rule: *Two of the four hydrogens surrounding each oxygen are close to the oxygen, and two are removed from it.*

This means that the partition function is given by

$$Z = \sum \exp(-\mathcal{E}/k_B T) \tag{7}$$

where the sum is now over all arrangements of the hydrogen ions that are allowed by the ice rule, and $\mathcal{E}$ is the energy of such arrangement.

The hydrogen–oxygen bonds between atoms form electric dipoles (see Fig. 1a), and they can conveniently be represented by arrows placed on the bonds pointing toward the end occupied by the oxygen, as in Fig. 1b. Slater's ice rule is then equivalent to stating that at each site (or vertex) of the lattice there are two arrows in, and two arrows out. There are just six such ways of arranging the arrows i.e there are six types of vertices, see Fig. 2.

Yet another way of representing the hydrogen–oxygen dipoles is to draw a line on an edge if the corresponding arrow points down or to the left, otherwise to leave the edge empty, see Fig 1c.

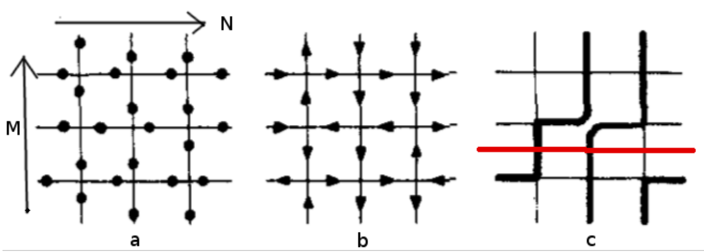

FIG. 1: H_2O Ice configuration with one "crystal" per lattice site ("Oxygen" ions are on the vertices of the lattice). *a*: "Hydrogens" bond towards the closest Oxygen *b*: Representation in terms of arrows. *c*: Representation in terms of non-crossing lines. A "row" of edges is shown in red.

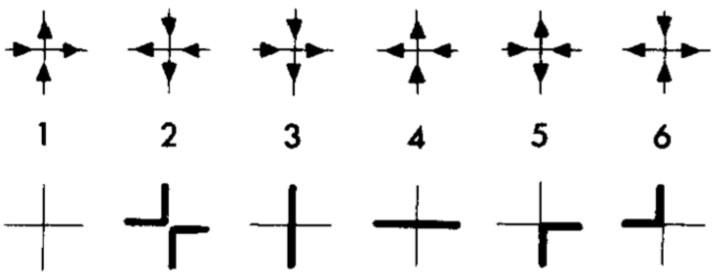

FIG. 2: The six vertex types (*upper*) and their line representations (*lower*).

1. Give a very simple (no calculation necessary) expression for the energy $\mathcal{E}$ of a configuration as a function of ϵ_i (the energy of a vertex of type i) and of n_i (the number of vertices i in the configuration).

   $\mathcal{E} = \sum_i n_i \epsilon_i$

2. From now on, we assume that $\epsilon_1 = \epsilon_2$, $\epsilon_3 = \epsilon_4$ and $\epsilon_5 = \epsilon_6$. Assume a lattice with M rows and N columns, and impose periodic boundary conditions. Consider a row of N vertical edges (between two adjacent rows of sites, see the red line in Fig. 1c). There are M such rows: label them $r = 1, 2, ..., M$ sequentially upwards. Let ϕ_r

denote the 'state' of row r: i.e. the arrangement of lines on the N vertical edges. Since each edge may or may not be occupied by a line, ϕ_r has 2^N possible values.

Show that one can write the partition function as $Z = \mathrm{Tr}\ \left(V^M\right)$. State the (very basic) condition that a pair of ϕ and ϕ' may give a non-zero $V(\phi, \phi')$ (no calculation necessary). State the (formal, non-explicit) expression of $V(\phi, \phi')$ in terms of the ϵ_i and their multiplicities between the two neighboring rows (note that the lines ϕ and ϕ' may correspond to one or to more than one configuration of vertices, or to none at all). Give the expression of Z as a function of $\Lambda_{\max}$, the largest eigenvalue of V, in the thermodynamic limit.

$$Z = \sum_{\phi_i}\sum_{\phi_2}...\sum_{\phi_M} V(\phi_1,\phi_2)V(\phi_2,\phi_3)...V(\phi_{M-1},\phi_M)V(\phi_M,\phi_1) = \mathtt{Tr}[V^M]$$

where V in the 2^N by 2^N transfer matrix. The summation is over all the 2^N possible value of each ϕ.

$$V(\phi,\phi') = \sum \exp(-\beta(m_1\epsilon_1 + ... + m_6\epsilon_6))$$

where ϕ is the arrangement of lines on one row of vertical edges, and ϕ' is the arrangement on the row above. The summation is over all allowed arrangements of lines on the intervening horizontal edges. These arrangements must satisfy the ice rule at each vertex: If there is no such arrangement, then $V(\phi,\phi')$ is zero.

3. Explain why if we have n lines on the first row then we have n lines in all other rows also. What can you say about $V(\phi, \phi')$ if ϕ and ϕ' have different numbers of lines? What does this imply for the general structure of the transfer matrix V?

If one starts by following a path upwards, or to the right, then one will always be traveling in one or other of these two directions, never down or to the left. The cyclic boundary conditions ensure that a path never ends. Suppose there are n such paths from the bottom of the lattice to the top. Each path will go through a row of vertical edges once and only once. It follows that if there are n lines on the bottom row of vertical edges, then there are n lines on every row. It follow that from question 2 that $V(\phi,\phi')$ is zero unless ϕ and ϕ' contain the same number of lines.

Therefore the matrix V breaks up into $N + 1$ diagonal blocks, one between the state with no lines, another between states with one line, and so on up to the state with N lines.

4. If we have n lines, we identify a state by specifying the positions $\{x_1, ..., x_n\}$ of the lines, ordered so that $1 \le x_1 < ... < x_n \le N$. Let $X = \{x_1, \dots, x_N\}$ be such a specification, and let $g(X)$ be the corresponding element of the eigenvector g of V. Then the eigenvalues equation can be written as:

$$\Lambda g(X) = \sum_Y V(X, Y) g(Y) \tag{8}$$

where $V(X, Y)$ is the element of V between states X and Y.

Show that $V(X,Y) = \sum a^{m_1+m_2} b^{m_3+m_4} c^{m_5+m_6}$ where the summation is over the allowed arrangements of lines on the intervening row of horizontal edges; and $m_1, ..., m_6$ are the numbers of intervening vertices of types $1, ...6$. Give the expression of a, b and c.

We define: $\omega_i = \exp(-\beta\epsilon_i) \quad i \in \{1,\dots,6\}$. If we choose $a = \omega_1 = \omega_2$, $b = \omega_3 = \omega_4$ and $c = \omega_5 = \omega_6$, then from **??** *we directly have:*

$$V(X,Y) = \sum a^{m_1+m_2} b^{m_3+m_4} c^{m_5+m_6}$$

where X and Y replace ϕ and ϕ'. Again the sum is over the allowed arrangements of lines on the intervening row of horizontal edges; and $m_1,\dots,m_6$ are the numbers of intervening vertices of types $1,\dots,6$.

5. We consider $n = 0$. What are the possible arrangements? Find the expression of $V(X, Y)$ in the $n = 0$ sector.

If $n = 0$, then there are no vertical lines in the two successive rows. There are two possible arrangements of lines on the intervening horizontal row of edges: either all the edges are empty, or they all contain a line. In the first instance, all vertices are of type 1; in the second they are all of type 4. Therefore the $n = 0$ block of V is a one-by-one matrix, with value

$$\Lambda = a^N + b^N$$

6. Consider the $n = 1$ sector. Show that, assuming $g(x) = z^x$, the eigenvalue equation can be written as

$$\Lambda z^x = a^N L(z) z^x - \frac{a^{x-1} b^{N-x} c^2}{(a - bz)} z^{N+1} + b^N M(z) z^x + \frac{a^{x-1} b^{N-x} c^2}{(a - bz)} z \tag{9}$$

Find the expression of $L(z)$ and $M(z)$. The second and fourth term of the RHS are boundary terms. Find the expressions of the eigenvectors and eigenvalues of the $n = 1$ block of V (note that $g(X)$ is now $g(x)$ and use the ansatz $g(x) = z^x$ with z a complex number). Which symmetry could have predicted such a result?

If $n = 1$, we can write $g(X)$ as $g(x)$, where x is the position of the vertical line in the row. This x can take the values $1, ..., N$, so this block of V is an N by N matrix, with elements $V(x, y)$. If x is less than y, then all horizontal edges between x and y must contain a line, and all others must be empty. If it is greater than y, then the reverse is true. If $x = y$, then either all horizontal edges are empty, or they are all full. Counting $m_1, \ldots, m_6$ for the various cases, the eigenvalue equation becomes

$$\begin{aligned} \Lambda g(x) =& a^{N-1} b g(x) + \sum_{y=x+1}^{N} a^{N+x-y-1} c^2 b^{y-x-1} g(y) \\ &+ a b^{N-1} g(x) + \sum_{y=1}^{x-1} a^{x-y-1} c^2 b^{N+y-x-1} g(y) \end{aligned} \tag{10}$$

*We look for a solution of the form $g(x) = z^x$ where z is a complex number. Substituting this form for $g(x)$ into eq. (**??**) and summing some elementary geometric series, the equation becomes:*

$$\Lambda z^x = a^N L(z) z^x - \frac{a^{x-1} b^{N-x} c^2}{(a - bz)} z^{N+1} + b^N M(z) z^x + \frac{a^{x-1} b^{N-x} c^2}{(a - bz)} z$$

with $L(z) = \frac{ab+(c^2-b^2)z}{a^2-abz}$ and $L(z) = \frac{a^2-c^2-abz}{ab-b^2z}$.
The second and fourth term differ only by a factor z^N, so their sum can be made to cancel by choosing $z^N = 1$.
The remaining first and third terms on the RHS are 'wanted terms', in that they have the same form as the LHS (constant times z^x). Thus the equation is now satisfied if $\Lambda = a^N L(z) + b^N M(z)$. There are N solutions to $z^N = 1$ for the complex number z, these give the N expected eigenvectors of this block of the matrix V. The corresponding eigenvalues are given by $\Lambda = a^N L(z) + b^N M(z)$. The solutions $g(x) = z^x$ and $z^N = 1$ could have been predicted on translation-invariance grounds.