PhD Entrance Examination
Physics and Chemistry of Biological Systems
March 2020

IMPORTANT GUIDELINES

- Solve **three** of the following problems.
- No extra credit is given for attempts to solve more than three problems.
- Do not write your name on the problem sheet nor use any mark that can identify you, as this would invalidate your exam.
- Write out solutions clearly and concisely. State each approximation used. Diagrams welcome.
- Number page, problem, and question clearly.
- All essays/solutions should be written in English.
- Data Science problems are exercises 8, 9 and 10.

Exercise 1

Consider a system with two states, A and B. The probability p_A of observing the system in A satisfies the differential equation

$$\frac{dp_A}{dt} = -k_{AB}\, p_A + k_{BA}\, p_B$$

where p_B is the probability of observing the system in B and k_{AB} and k_{BA} are time-independent rate constants. (1): Compute the equilibrium probability of observing the system in A. (2): compute the probability of observing the system in A at time t, given that the system is in A at time 0 (namely $p_A\,(t = 0) = 1$).

Exercise 2

A protein can be in $N + 1$ different states, one of which is folded. The folded state haves energy equal to ϵ_F, and each of the unfolded states have energy $\epsilon_U > \epsilon_F$. Assuming $N = 1000$, $\epsilon_U - \epsilon_F = 20$ kJ/mol, and given the Boltzmann constant $k_B = 0.00831$ kJ/mol/K, find the probability that the protein is folded at temperature T=275K. Also find the melting temperature of the protein, defined as the temperature at which the probability to find the protein in its folded state is 50%.

Exercise 3

Consider a system of N free particles, the energy of each particle can assume only two values $E_0 = 0$ and $E_1 = E > 0$ with occupation numbers n_0 and n_1. The total energy of the system is $\mathcal{U}$.

By employing only the *microcanonical* ensemble, find:

1. The expression for the entropy $\mathcal{S}$ of the system.
2. The most probable values for n_0 and n_1.
3. *Optional.* The temperature T of the system as a function of $\mathcal{U}$. Show that T can be negative and briefly explain why.

Exercise 4

Consider a system of N non-interacting free particle in *one dimension* in thermal equilibrium at temperature T. Neglecting prefactors, the energy of each particle is given by the expression:

$$E = |\vec{p}|^{1/2}$$

where $\vec{p}$ is the momentum of the particle. Find:

1. The internal energy U and specific heat C_V at constant volume of the system.
2. The density of states $\omega = \omega(E)$ of the system.

Exercise 5

Consider a system of N spins in an external field H and in canonical equilibrium at temperature T. The Hamiltonian of the system is $\mathcal{H} = -H \sum_{i=1}^{N} S_i$ where $S_i = \{-1,\ 0,\ +1\}$.

1. Compute the expectation value of the magnetization per spin, $m = \langle \frac{\sum_i S_i}{N} \rangle$.
2. Compute $\langle S_i\, S_j \rangle$ where i and j are different sites.
3. Optional: Consider now two specific spin realizations picked independently from the canonical ensemble, $\{S_1^a, ..., S_N^a\}$ and $\{S_1^b, ..., S_N^b\}$. What is the expected value of $S_i^a\, S_i^b$ and of $S_i^a\, S_{j\neq i}^b$?

Exercise 6

A protein is studied by thermal denaturation at constant pressure.
A preliminary experiment records fluorescence spectra at five different temperatures with the following intensity readings:

Temperature	Intensity (arbitrary units)
20 C	50.0
40 C	49.5
50 C	44.0
60 C	30.0
80 C	25.0

Assume that at 20 C and 80 C the protein is essentially folded and unfolded, respectively.
Assume also that the recorded signal (y) is the sum of the signals (y_f and y_u not dependent on temperature) proportional to the folded and unfolded populations (p_f and p_u), i.e: $y = p_f y_f + p_u y_u$.

Estimate the enthalpy, entropy and Gibbs free energy of folding at 25 C, and the melting temperature.

Exercise 7

1. Consider two methane molecules interacting with each other. What types of forces exist between the two molecules? Make a sketch of the interaction potential between the two methane molecules as a function of the distance r between them.
2. Now consider two ions, a sodium (Na+) and chloride ion (Cl-) interacting with each other. What types of forces exist between these two ions? Make a sketch of the interaction potential between Na+ and Cl- as a function of the distance r and overlap it with the sketch in the previous question.

3. If the two ions Na+ and Cl- now interact with each other in a liquid like water, discuss how the interaction between the ions changes if at all. Contrast this to the interaction of two methane molecules in water - what is particular to this interaction.

Exercise 8 - Data Science exercise

Let $\mathbf{x} \sim \mathcal{N}(0, I_D)$ be a D-dimensional Gaussian random variable with zero mean and unitary variance. Calculate the probability distribution of the squared norm of $\mathbf{x}$, $\|\mathbf{x}\|^2 = \sum_{i=1}^{D} x_i^2$. By comparing the mean of the distribution with its standard deviation, comment on the typical length of highly distributed normal random variables.

You can use the following formulae

$$x \sim \mathcal{N}(m, \sigma) \leftrightarrow p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \tag{1}$$

$$y \sim \mathcal{G}(a, b) \leftrightarrow p(y) = \frac{1}{\Gamma(a)b^a} y^{a-1} \exp\left[-\frac{y}{b}\right] \tag{2}$$

for the probability density of the Normal and Gamma respectively. Useful properties of the Gamma distribution that you may use without proof:

- If $y \sim \mathcal{G}(a, b)$, then $\langle y \rangle = ab$ and $\mathrm{var}(y) = ab^2$;
- If $y_i \sim \mathcal{G}(a, b)$ for $i = 1, \ldots, D$, then $\sum_{i=1}^{D} \sim \mathcal{G}(Da, b)$

Exercise 9 - Data Science exercise

Let $n_i|\mu \sim \mathcal{P}(\mu)$ be independent Poisson random variables with rate μ, and let $\mu \sim \mathcal{G}(a, b)$ be Gamma distributed a priori. Show that the posterior distribution $p(\mu|n_i)$ is also a Gamma and compute its parameters. What happens to the posterior variance when the number of observations becomes very large?

You can use the following formulae

$$y \sim \mathcal{G}(a, b) \leftrightarrow p(y) = \frac{1}{\Gamma(a)b^a} y^{a-1} \exp\left[-\frac{y}{b}\right] \tag{3}$$

$$n \sim \mathcal{P}(\mu) \leftrightarrow p(n) = \frac{\mu^n}{n!} \exp[-\mu] \tag{4}$$

for the probability density or distribution of the Gamma and Poisson distributions respectively. Useful properties of the Gamma distribution that you may use without proof:

- If $y \sim \mathcal{G}(a, b)$, then $\langle y \rangle = ab$ and $\mathrm{var}(y) = ab^2$;
- If $y_i \sim \mathcal{G}(a, b)$ for $i = 1, \ldots, D$, then $\sum_{i=1}^{D} \sim \mathcal{G}(Da, b)$

Exercise 10 - Data Science exercise

Let $z_i \sim \mathcal{G}(a, b)$ be independent, identically Gamma distributed variables. Compute the maximum likelihood estimate of the mean of the Gamma distribution.

You can use the following formulae about the Gamma distribution

$$y \sim \mathcal{G}(a, b) \leftrightarrow p(y) = \tfrac{1}{\Gamma(a) b^a} y^{a-1} \exp\left[-\tfrac{y}{b}\right] \tag{5}$$

$$y \sim \mathcal{G}(a, b) \langle y \rangle = ab \tag{6}$$