Solution 6: MCMC Simulation and Review Problems

Due Monday July 30 at 11:59PM in Instructor's Inbox

Preface: In this assignment, you will use computational software to implement the Markov Chain Monte Carlo (MCMC) simulation algorithm we simulated ourselves in class.

1. MCMC of Spins in a Magnetic Field

Solutions in spin_model_simulation_soln.nb on course website.

2. Optional Bonus: MCMC of Mean-Field Ising Model

Solutions in spin_model_simulation_soln.nb on course website.

Review Problems

(Not to submit)

Preface: This assignment contains a list of review problems to help you prepare for the final exam.

1. Protein Expression and Probability

(a) Given the range of possible protein production, x can go from 0 to ∞ . Therefore, for p(x) to be normalized, we must obtain 1 when we integrate the function over this entire domain:

$$\int_0^\infty dx \, p(x) = 1. \tag{1}$$

From the definition of the probability density we have

$$1 = \int_{0}^{\infty} dx A\left(\frac{x}{b}\right)^{N} e^{-x/b}$$

= $A \int_{0}^{\infty} dx \left(\frac{x}{b}\right)^{N} e^{-x/b}$
= $A \int_{0}^{\infty} du \, b \, u^{N} e^{-u}$
= $A \, b \int_{0}^{\infty} du \, u^{N} e^{-u}$, (2)

where we changed variables with u = x/b in the third line, and factored the *u*-independent constant out of the integral in the final line. By the integral definition of factorial, we have

$$N! = \int_0^\infty du \, u^N e^{-u}.\tag{3}$$

Therefore, the final line of Eq.(2) becomes

$$1 = A b N!, \tag{4}$$

and we can conclude

$$A = \frac{1}{bN!}.$$
(5)

The normalized probability density is therefore

$$p(x) = \frac{1}{bN!} \left(\frac{x}{b}\right)^N e^{-x/b}.$$
(6)

(b) The mean of a random variable defined by the probability density p(x) (which has a nonzero domain for $x \in [0, \infty)$) is

$$\langle x \rangle = \int_0^\infty dx \, x \, p(x). \tag{7}$$

Using Eq.(6) to compute this value, we obtain

$$\begin{aligned} \langle x \rangle &= \int_0^\infty dx \, x \, \frac{1}{bN!} \left(\frac{x}{b}\right)^N e^{-x/b} \\ &= \frac{1}{N!} \int_0^\infty dx \, \left(\frac{x}{b}\right)^{N+1} e^{-x/b} \\ &= \frac{1}{N!} \int_0^\infty du \, b \, u^{N+1} e^{-u} \\ &= b \frac{(N+1)!}{N!}, \end{aligned} \tag{8}$$

where in the third line we performed a change of variables with u = x/b and in the final line we used an equation analogous to Eq.(3). By the definition of factorial, we ultimately find

$$\langle x \rangle = b(N+1). \tag{9}$$

(c) The standard deviation of a random variable is given by

$$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}.$$
 (10)

We have already computed the mean in Eq.(9). Now, to complete the calculation of the standard deviation of this random variable, we need to compute $\langle x^2 \rangle$. Doing so, we have

$$\langle x^{2} \rangle = \int_{0}^{\infty} dx \, x^{2} \, \frac{1}{bN!} \left(\frac{x}{b}\right)^{N} e^{-x/b}$$

$$= \frac{b}{N!} \int_{0}^{\infty} dx \, \left(\frac{x}{b}\right)^{N+2} e^{-x/b}$$

$$= \frac{b^{2}}{N!} \int_{0}^{\infty} du \, u^{N+2} e^{-u}$$

$$= b^{2} \frac{(N+2)!}{N!},$$
(11)

where in the second line we multiplied the integral by b/b, in the third line we performed a change of variables u = x/b, and in the final line we used an equation analogous to Eq.(3). By the definition of the factorial, we find

$$\langle x^2 \rangle = b^2 (N+2)(N+1) = b^2 (N+1)^2 + b^2 (N+1).$$
 (12)

Now, using Eq.(12) and Eq.(9) to compute the standard deviation from Eq.(10), we then obtain

$$\sigma_x = \sqrt{b^2(N+1)^2 + b^2(N+1) - b^2(N+1)^2},$$
(13)

or,

$$\sigma_x = b\sqrt{N+1},\tag{14}$$

2. Adsorption

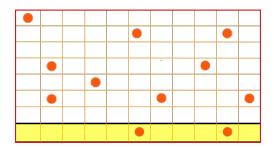


Figure 1: Particular microstate of system. The possible surface states are shown in yellow. The possible bulk states are the remaining states.

(a) To determine the number of microstates the system has for a particular value of k we need to determine the number of arrangements we can have in the system for a particular value of k. Let us call this quantity $\Omega_{N,M}(k)$ to denote the fact that it is a function of k with both N and M as constant parameters.

We note that k denotes the number of filled sites on the surface. If there are N possible sites and k of them are filled with identical atoms, then there are $\binom{N}{k}$ ways we can select a particular collection of filled surface sites. However, after choosing the surface sites, we must also choose the bulk sites. There are M total bulk sites and if k atoms are filling the surfaces sites then N - k atoms must be filling the bulk sites. Thus, there are $\binom{M}{N-k}$ ways we can select a particular collection of bulk sites to hold the identical atoms. The total number of microstates in the system is a product of the number of ways to fill the surface sites and the number of ways to fill the bulk sites for a particular value of k. Thus we have

$$\Omega_{N,M}(k) = \binom{N}{k} \binom{M}{N-k}.$$
(15)

(b) The free energy of the system is the energy minus the product of temperature and the entropy. Since we are concerned with the free energy of a macrostate, we compute the entropy according to the Boltzmann definition of entropy:

$$S_{N,M}(k) = k_B \ln \Omega_{N,M}(k). \tag{16}$$

Given that the energy of the system is equal to $-\varepsilon$ times the number of atoms on the surface, then, by our definition of k, we have

$$E_{N,M}(k) = -\varepsilon k, \tag{17}$$

as the energy of the system. Thus we can conclude the free energy is

$$F_{N,M}(k) = E_{N,M}(k) - TS_{N,M}(k) = -\varepsilon k - k_B T \ln\left[\binom{N}{k}\binom{M}{N-k}\right].$$
(18)

(c) Next, we want to approximate Eq.(18) using Stirling's approximation. Stirling's approximation applies to factorials and there are only factorials in the last term of Eq.(18) so we will focus our attention there. The form of Stirling's approximation we will use is

$$\ln N! \simeq N \ln N - N,\tag{19}$$

where we have neglected all the terms which are not proportional to N. Applying this approxi-

mation to the final term in Eq.(18), we obtain

$$\ln \left[\binom{N}{k} \binom{M}{N-k} \right] = \ln N! - \ln k! - \ln(N-k)! + \ln M! - \ln(N-k)! - \ln(M-N+k)! \\ \simeq N \ln N - k \ln k - (N-k) \ln(N-k) + M \ln M - (N-k) \ln(N-k) \\ - (M-N+k) \ln(M-N+k).$$
(20)

The free energy then has the (somewhat complicated) form

$$F_{N,M}(k) \simeq -\varepsilon k - k_B T \Big[N \ln N - k \ln k - (N-k) \ln(N-k) + M \ln M - (N-k) \ln(N-k) - (M-N+k) \ln(M-N+k) \Big].$$
(21)

(d) We want to determine the condition that k must satisfy in order for the system to be in thermal equilibrium. Let $k = \overline{k}$ be the value of k at thermal equilibrium. By the relationship between the free energy and thermal equilibrium, we have

$$\frac{\partial}{\partial k}F_{N,M}(k)\Big|_{k=\overline{k}} = 0, \quad \frac{\partial^2}{\partial k^2}F_{N,M}(k)\Big|_{k=\overline{k}} > 0.$$
(22)

The problem asks us to derive **one** condition that k must satisfy in order to be at thermal equilibrium. We will use the first condition in Eq.(22) for this purpose. From Eq.(21), we have

$$0 = \frac{\partial}{\partial k} F_{N,M}(k) \Big|_{k=\overline{k}}$$

= $-\varepsilon - k_B T \Big[-\ln \overline{k} - 1 + \ln(N - \overline{k}) + 1 + \ln(N - \overline{k}) + 1 - \ln(M - N + \overline{k}) - 1 \Big],$
= $-\varepsilon - k_B T \ln \frac{(N - \overline{k})^2}{\overline{k}(M - N + \overline{k})}$ (23)

which—upon adding ε , dividing by $-k_BT$, and taking the exponential of both sides—can be written as

$$e^{-\varepsilon/k_B T} = \frac{(N-\overline{k})^2}{\overline{k}(M-N+\overline{k})}.$$
(24)

If \overline{k} is the value of k at which the system is in thermal equilibrium at temperature T, then \overline{k} must satisfy Eq.(24).

3. Nucleotide zipper

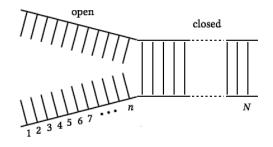


Figure 2: "Zipper" of nucleotides

- (a) By the problem prompt, we are modeling a strand of DNA (double-stranded DNA, to be specific) as a zipper with links between base-pairs. From Fig. 2, we see that there is only one way for a single zipper link to be open and if link k is open then all links up to k are also open. Therefore, if there are N total links then there are N possible microstates in the system. Also, each time we have an open link there is an energy ε associated with it, so a configuration (i.e., microstate) with k open links has an energy εk.
- (b) From (a), we can construct the partition function of this system. We have

$$Z_{\text{zipper}} = \sum_{k=0}^{N} e^{-\varepsilon k/k_B T} = \sum_{k=0}^{N} \left(e^{-\varepsilon/k_B T} \right)^k.$$
(25)

Using the geometric series identity

$$\sum_{k=0}^{N} x^{k} = \frac{1 - x^{N+1}}{1 - x} \quad (\text{ for } |x| < 1)$$
(26)

and noting that $e^{-\varepsilon/k_BT}$ is less than 1 for $\varepsilon > 0$, we then have

$$Z_{\text{zipper}} = \frac{1 - e^{-(N+1)\varepsilon/k_B T}}{1 - e^{-\varepsilon/k_B T}}.$$
(27)

(c) If we take $N \gg 1$, we can approximate Eq.(27) as simply

$$Z_{\text{zipper}} \simeq \frac{1}{1 - e^{-\varepsilon/k_B T}},\tag{28}$$

since $e^{-N\varepsilon/k_BT} \ll 1$. The average number of open links can be computed from this partition function. First we use Eq.(25) and the definition of this average to obtain

$$\langle k \rangle = \frac{1}{Z_{\text{zipper}}} \sum_{k=0}^{N} k \, e^{-\beta \varepsilon k}$$
$$= -\frac{1}{Z_{\text{zipper}}} \frac{\partial}{\partial(\beta \varepsilon)} Z_{\text{zipper}}$$
$$= -\frac{\partial}{\partial(\beta \varepsilon)} \ln Z_{\text{zipper}}, \tag{29}$$

where we used $1/k_BT = \beta$ in the first line. Now, applying Eq.(28) we obtain

$$\langle k \rangle \simeq \frac{\partial}{\partial(\beta\varepsilon)} \ln\left(1 - e^{-\beta\varepsilon}\right) = \frac{e^{-\beta\varepsilon}}{1 - e^{-\beta\varepsilon}},$$
(30)

or, upon multiplying by $e^{\beta\varepsilon}/e^{\beta\varepsilon}$,

$$\langle k \rangle \simeq \frac{1}{e^{\varepsilon/k_B T} - 1}.$$
(31)

4. Entropy of an Ideal Gas

(a) We want to show that the Gibbs entropy is the negative of the temperature derivative of the free energy. The free energy of a system in terms of the partition function is

$$F = -k_B T \ln Z = -k_B T \ln \left[\sum_{\{i\}} e^{-E_i/k_B T} \right],$$
(32)

where E_i is the energy of the microstate *i* and $\sum_{\{i\}}$ is a summation over all microstates. Computing the temperature derivative of this free energy, we find

$$\frac{\partial F}{\partial T} = -k_B \ln \left[\sum_{\{i\}} e^{-E_i/k_B T} \right] - k_B T \frac{\sum_{\{i\}} e^{-E_i/k_B T} \cdot \left(\frac{E_i}{k_B T^2}\right)}{\sum_{\{i\}} e^{-E_i/k_B T}}$$

$$= -k_B \ln Z - \frac{1}{T} \frac{\sum_{\{i\}} e^{-E_i/k_B T} E_i}{Z}$$

$$= -k_B \left[\ln Z \cdot 1 + \frac{1}{k_B T} \frac{1}{Z} \sum_{\{i\}} e^{-E_i/k_B T} E_i \right]$$

$$= -k_B \left[\ln Z \cdot \frac{1}{Z} \sum_{\{i\}} e^{-E_i/k_B T} + \frac{1}{k_B T} \frac{1}{Z} \sum_{\{i\}} e^{-E_i/k_B T} E_i \right]$$

$$= -k_B \left[\sum_{\{i\}} \frac{e^{-E_i/k_B T}}{Z} \left(\ln Z + \frac{E_i}{k_B T} \right) \right]$$
(33)

where in the fourth line we used the partition function definition $Z = \sum_{\{i\}} e^{-E_i/k_B T}$ and in the final line we used the identity $\ln e^a = a$. Using this identity again, the identity $-\ln x = \ln x^{-1}$, and the identity $\ln x + \ln y = \ln xy$ in the final line, we then obtain

$$\frac{\partial F}{\partial T} = k_B \sum_{\{i\}} \frac{e^{-\beta E_i}}{Z} \ln\left(\frac{e^{-\beta E_i}}{Z}\right),\tag{34}$$

which, if we recall that $p_i = e^{-\beta E_i}/Z$, leaves us with

$$\frac{\partial F}{\partial T} = k_B \sum_{\{i\}} p_i \ln p_i = -S,\tag{35}$$

which is the desired result

(b) We want to compute the entropy for an ideal gas from Eq.(35). For the ideal gas, the partition function is

$$Z_{\text{ideal}} = \frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}.$$
(36)

Therefore the free energy of the system is

$$F_{\text{ideal}} = -k_B T \ln\left[\frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}\right].$$
(37)

Computing the entropy from Eq.(35), we obtain

$$S_{\text{ideal}} = -\frac{\partial}{\partial T} \left\{ -k_B T \ln\left[\frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}\right] \right\}$$
$$= k_B \ln\left[\frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h}\right)^{3N/2}\right] + k_B T \frac{\partial}{\partial T} \ln\left[\frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}\right]$$
$$= k_B \ln\left[\frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}\right] + k_B T \frac{\partial}{\partial T} \frac{3N}{2} \ln T$$
$$= k_B \ln\left[\frac{V^N}{N!}\right] + k_B \frac{3N}{2} \ln\left(\frac{2\pi m k_B T}{h^2}\right) + \frac{3N k_B}{2}.$$
(38)

Now, we will replace $\ln N!$ with the result from Stirling's approximation. From the approximation, we have

$$\ln \frac{1}{N!} = -(N \ln N - N) = \ln \frac{1}{N^N} + N,$$
(39)

where we neglected terms which were not proportional to N. Thus, Eq.(38) becomes

$$S_{\text{ideal}} \simeq k_B \ln\left[\frac{V^N}{N^N} + N\right] + k_B \frac{3N}{2} \ln\left(\frac{2\pi m k_B T}{h^2}\right) + \frac{3Nk_B}{2},$$
$$= Nk_B \ln\left[\frac{V}{N}\right] + Nk_B + k_B \frac{3N}{2} \ln\left(\frac{2\pi m k_B T}{h^2}\right) + \frac{3Nk_B}{2},$$
(40)

or

$$S_{\text{ideal}} \simeq Nk_B \left[\ln \left(\frac{V}{N} \right)^N + \frac{5}{2} - \frac{3}{2} \ln \left(\frac{h^2}{2\pi m k_B T} \right) \right],\tag{41}$$

which is the desired result.

5. Partition Function as an Integral

We want to approximate the integral

$$Z(T) = \int_{-\infty}^{\infty} d\rho \, \exp\left[-\beta N f(\rho, T)\right],\tag{42}$$

where

$$f(\rho, T) = -a(T - T_c)\rho^2 + c\rho^4.$$
(43)

By Laplace's method, we have

$$Z(T) = \int_{-\infty}^{\infty} d\rho \, \exp\left[-\beta N f(\rho, T)\right] \simeq \sqrt{\frac{2\pi}{\beta N f''(\overline{\rho}, T)}} e^{-\beta N f(\overline{\rho}, T)},\tag{44}$$

where $\overline{\rho}$ is the value of ρ where $f(\rho, T)$ is at a local minimum. To find the value at which $f(\rho, T)$ is at a local minimum, we employ the standard algorithm: setting the first derivative to zero and selecting the solution that yields a positive second derivative. Completing the first task, we have

$$0 = \frac{\partial}{\partial \rho} f(\rho, T) \Big|_{\rho = \overline{\rho}}$$

= $-2a(T - T_c)\overline{\rho} + 4c\overline{\rho}^3$
= $2\overline{\rho} \left[-a(T - T_c) + 2c\overline{\rho}^2 \right],$ (45)

which yields the three solutions

$$\overline{\rho}_0 = 0, \quad \text{and} \quad \overline{\rho}_{\pm} = \pm \sqrt{\frac{a}{2c}(T - T_c)}.$$
 (46)

We determine which of these solutions leads to a local minimum of the free energy by computing the second derivative at each of these values. Doing so, we find

$$\frac{\partial^2}{\partial \rho^2} f(\rho, T) \Big|_{\rho = \overline{\rho}_0} = -2a(T - T_c) + 12c\overline{\rho}_0^2$$
$$= -2a(T - T_c), \tag{47}$$

and

$$\frac{\partial^2}{\partial \rho^2} f(\rho, T) \Big|_{\rho = \overline{\rho}_{\pm}} = -2a(T - T_c) + 12c\overline{\rho}_{\pm}^2$$
$$= -2a(T - T_c) + 12c\frac{a}{2c}(T - T_c)$$
$$= 4a(T - T_c). \tag{48}$$

If we take a > 0, we see that the $\overline{\rho}_{\pm}$ solutions in Eq.(46) correspond to local minima when $T > T_c$ and the $\overline{\rho}_0$ solution corresponds to a local minimum when $T < T_c$. Conversely, if we take a < 0, the $\overline{\rho}_{\pm}$ solutions correspond to local minima when $T < T_c$ and the $\overline{\rho}_0$ solution corresponds to a local minimum when $T > T_c$.

We will assume a < 0 (or, equivalently a = -|a|) in what follows because this assumption is consistent with the high temperature (i.e., $T > T_c$) behavior of the system having $\overline{\rho} = 0_0$ which is akin to the behavior of the order parameter of the mean-field Ising model. Evaluating the function Eq.(43) at the computed values Eq.(46), we find

$$f(\rho,T)\Big|_{\rho=\overline{\rho}_0} = -a(T-T_c)\overline{\rho}_0^2 + c\overline{\rho}_0^4$$
$$= 0$$
(49)

and

$$f(\rho, T)\Big|_{\rho = \overline{\rho}_{\pm}} = -a(T - T_c)\overline{\rho}_{\pm}^2 + c\overline{\rho}_{\pm}^4$$

$$= -a(T - T_c)\frac{a}{2c}(T - T_c) + c\frac{a^2}{4c^2}(T - T_c)^2$$

$$= -\frac{a^2}{2c}(T - T_c)^2.$$
 (50)

Therefore, assembling the pieces in Eq.(44), we find, for a = -|a| and $T > T_c$,

$$Z(T) \simeq \sqrt{\frac{2\pi}{N\beta(2|a|(T-T_c))}} \exp\left[-\beta N \cdot 0\right]$$
$$= \sqrt{\frac{k_B \pi}{N|a|(1-T_c/T)}},$$
(51)

where we used $\beta = 1/k_BT$ in the final line. And, for a = -|a| and $T < T_c$, we find

$$Z(T) \simeq \sqrt{\frac{2\pi}{N\beta(4|a|(T_c - T))}} \exp\left[-\beta N \cdot \left(-\frac{a^2}{2c}(T - T_c)^2\right)\right] = \sqrt{\frac{k_B\pi}{2N|a|(T_c/T - 1)}} \exp\left[\frac{a^2 N}{2ck_BT}(T - T_c)^2\right],$$
(52)

where we again used $\beta = 1/k_B T$. More summarily, we find

$$Z(T) \simeq \begin{cases} \sqrt{\frac{k_B \pi}{N|a|(1 - T_c/T)}} & \text{for } T > T_c, \\ \sqrt{\frac{k_B \pi}{2N|a|(T_c/T - 1)}} \exp\left[\frac{a^2 N}{2ck_B T}(T - T_c)^2\right] & \text{for } T < T_c. \end{cases}$$
(53)

We can then infer that Laplace's method is a suitable approximation provided we adhere to the temperature range and function association in Eq.(53) and we stay away from $T = T_c$.

6. Transition probabilities and equilibrium probabilities

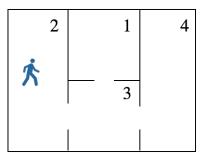


Figure 3: Four connected rooms. The person has a probability 1/(M + 1) of remaining in place and a probability of 1/(M + 1) of moving to any one of the adjacent rooms.

(a) We want to fill in the elements of the table according to the stipulation that a person has a probability 1/(M + 1) of remaining in place and a probability of 1/(M + 1) of moving to any one of

the adjacent rooms. Given Fig. 3 we see that room 1 is connected to M = 1 room; room 2 is connected to M = 1 room; room 3 is connected to M = 3 rooms; room 4 is connected to M = 1 room. Therefore, our transition probabilities are

$$\pi_{1\to1} = \frac{1}{2} \qquad \pi_{1\to2} = 0 \qquad \pi_{1\to3} = \frac{1}{2} \qquad \pi_{1\to4} = 0$$

$$\pi_{2\to1} = 0 \qquad \pi_{2\to2} = \frac{1}{2} \qquad \pi_{2\to3} = \frac{1}{2} \qquad \pi_{2\to4} = 0$$

$$\pi_{3\to1} = \frac{1}{4} \qquad \pi_{3\to2} = \frac{1}{4} \qquad \pi_{3\to3} = \frac{1}{4} \qquad \pi_{3\to4} = \frac{1}{4}$$

$$\pi_{4\to1} = 0 \qquad \pi_{4\to2} = 0 \qquad \pi_{4\to3} = \frac{1}{2} \qquad \pi_{4\to4} = \frac{1}{2}$$
(54)

(b) We can use the transition probabilities to determine the equilibrium probabilities p_j^{eq} by using the detailed balance equation:

$$\frac{p_j^{\text{eq}}}{p_i^{\text{eq}}} = \frac{\pi_{i \to j}}{\pi_{j \to i}}.$$
(55)

Choosing unique ratios for Eq.(55), we find

$$\frac{p_1^{\text{eq}}}{p_3^{\text{eq}}} = \frac{1/4}{1/2}$$
$$= \frac{1}{2} \longrightarrow p_1^{\text{eq}} = \frac{1}{2}p_3^{\text{eq}}$$
(56)

$$\frac{p_2^{\text{eq}}}{p_3^{\text{eq}}} = \frac{1/4}{1/2}
= \frac{1}{2} \longrightarrow p_2^{\text{eq}} = \frac{1}{2} p_3^{\text{eq}}
\frac{p_3^{\text{eq}}}{p_4^{\text{eq}}} = \frac{1/2}{1/4}$$
(57)

$$\begin{array}{l} {}^{\mathrm{eq}}_{4} & 1/4 \\ = 2 & \longrightarrow & p_{4}^{\mathrm{eq}} = \frac{1}{2} p_{3}^{\mathrm{eq}}. \end{array}$$

$$(58)$$

With probability normalization and these results we have

$$1 = p_1^{\text{eq}} + p_2^{\text{eq}} + p_3^{\text{eq}} + p_4^{\text{eq}} = \left(\frac{1}{2} + \frac{1}{2} + 1 + \frac{1}{2}\right) p_3^{\text{eq}} = \frac{5}{2}p_3^{\text{eq}},\tag{59}$$

which implies

$$p_3^{\rm eq} = \frac{2}{5}.$$
 (60)

From Eq.(56), Eq.(57), and Eq.(58) we then obtain

$$p_1^{\rm eq} = p_2^{\rm eq} = p_4^{\rm eq} = \frac{1}{5}.$$
(61)